AWARD NUMBER: W81XWH-16-1-0599

TITLE: Dissecting the Mechanisms of Drug Resistance in BRCA1/2-Mutant Breast Cancers

PRINCIPAL INVESTIGATOR: Dr. Andre Nussenzweig

CONTRACTING ORGANIZATION: The Geneva Foundation, Tacoma, WA

REPORT DATE: January 2021

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 222024302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
January 2021	Final Report	30Sep2016-29Sep2020

4. TITLE AND SUBTITLE

Dissecting the Mechanisms of Drug Resistance in BRCA1/2-Mutant Breast Cancers

5a. CONTRACT NUMBER W81XWH-16-1-0599
5b. GRANT NUMBER BC151331 and BC151331P1
5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Dr. Andre Nussenzweig, andre_nussenzweig@nih.gov

Dr. Alan D. D'Andrea , Alan_Dandrea@dfci.harvard.edu

5d. PROJECT NUMBER
5e. TASK NUMBER
5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

The Geneva Foundation
917 Pacific Ave. # 600
Tacoma, WA 98402

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Poly(ADP-ribose) polymerase (PARP) inhibition provides a promising therapeutic modality for targeting homologous recombination (HR) deficient tumors such as BRCA1 and BRCA2-mutated triple negative breast cancers (TNBCs). Although PARP inhibitors have shown activity in the BRCA-associated TNBCs, several of these tumors develop de novo as well as acquired PARP inhibitor (PARPi) resistance. Besides attenuation in intracellular uptake of drugs, the only known mechanism that drives chemotherapy resistance of BRCA1/2-deficient cancers is through the restoration of HR. Recent studies from our laboratories (Nussenzweig and D'Andrea) indicate that deregulation of pathways that promote extensive degradation of nascent DNA strands and alternative end-joining (Alt-EJ) can render BRCA1/2-deficient cells resistant to PARPi in a HR-independent manner. The objective of our project is to collaboratively test the hypothesis that complex processes involving Alt-EJ or replication fork stability promote survival and drives resistance to chemotherapy. A detailed assessment of the critical mediators that regulate the balance between HR, Alt-EJ and replication fork degradation should identify novel means to overcome acquired chemoresistance in BRCA1/2-mutated breast cancers. During the first year of the DOD funding, we have made progress in identifying the proteins which contribute to the replication fork stability and we have identified new mechanisms of chemoresistance in BRCA2-deficient tumors.

15. SUBJECT TERMS

Breast cancer, BRCA1, BRCA2, PARP inhibitors, chemotherapy, resistance, HR, replication fork stability, haploinsufficiency, RNF168, 53BP1, PTIP, Shieldin, end resection

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
Unclassified	Unclassified	Unclassified	Unclassified	256	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

Page

1. **Introduction…….1**
2. **Keywords……..1**
3. **Accomplishments……..…….. 2**
4. **Impact……...……...…….10**
5. **Changes/Problems...…….....10**
6. **Products…….…….…..…….11**
7. **Participants & Other Collaborating Organizations…… 15**
8. **Special Reporting Requirements……20**
9. **Appendices……20**

1. INTRODUCTION:

The inactivation of the tumor suppressor genes *BRCA1* and *BRCA2* by mutations or epigenetic silencing is a critical event in breast and ovarian carcinogenesis. *BRCA1* and *BRCA2* encode proteins that are essential for accurate double strand break (DSB) repair by homologous recombination (HR). BRCA1 functions early during DSB resection, BRCA2 functions later in HR by catalyzing RAD51 nucleo-filaments at processed DSBs. Accordingly, HR deficient breast and ovarian tumors are highly sensitive to poly(ADP-ribose) polymerase (PARP) inhibitors, since PARP inhibitors exhibit synthetic lethality in tumors with defective HR DNA repair. PARP inhibitors are currently in development for BRCA- or otherwise HR repair-deficient cancers, with FDA approval of Olaparib, Rucaparib and Niraparib. Of these drugs, Olaparib has been the most widely studied thus far, and it has been approved by the FDA as a monotherapy for treatment of ovarian cancer patients with germline *BRCA1* or *BRCA2* mutations. Nonetheless, *de novo* and acquired PARP inhibitor (PARPi) resistance, is a pressing clinical problem in patients with BRCA-deficient cancers treated with PARP inhibitors. Therefore, identification of the mechanisms underlying PARPi resistance is crucial for improving treatment and predicting tumor responses. Besides attenuation in intracellular uptake and increased efflux of drugs, the other known mechanism of PARPi resistance in BRCA-deficient tumors include restoration of HR due to somatic reversion of *BRCA1/BRCA2* or loss of other genes such as 53BP1, RIF1 or REV7. Recently, it was shown that BRCA1 and BRCA2 protect stalled replication forks from Mre11-mediated degradation, independent of their roles in HR. Accordingly, restoration of either HR capacity or replication fork stability is also associated with PARPi resistance in BRCA-deficient tumors. Indeed, a recent study from the Nussenzweig laboratory indicated that loss of *PTIP* protects replication forks from degradation in both *BRCA1*- and *BRCA2*- deficient cells and confers PARPi resistance. The D'Andrea laboratory has recently identified a novel DNA repair pathway, the so-called PARP/POLQ end-joining pathway, which, when upregulated, provides the HR-deficient breast tumor cell with an alternative mechanism of DNA repair. Collectively, recent studies from both laboratories indicate that deregulation of pathways that promote extensive degradation of nascent DNA strands and alternative end-joining (Alt-EJ) can render BRCA1/2-deficient cells resistant to PARPi in a HR independent manner. *We had therefore hypothesized that replication fork protection and PARP mediated Alt-EJ are novel and potentially interlinked mechanisms by which BRCA1/2-deficient breast cancers acquire resistance to chemotherapy.* Accordingly, the objective of our project is to provide a more detailed assessment of the factors that contribute to replication fork protection and Alt-EJ. This could lead to therapeutic approaches to overcome acquired resistance by targeting new vulnerabilities in both *BRCA1/2*-mutant and *BRCA1/2*–wildtype breast cancer.

2. KEYWORDS:

Breast cancer, BRCA1, BRCA2, PARP inhibitors, chemotherapy, resistance, HR, replication fork stability, haploinsufficiency, RNF168, 53BP1, PTIP, Shieldin, end resection, Thap1, REV7, TRIP13, Novobiocin

3. ACCOMPLISHMENTS:

The major goal of the project is to identify the molecular mechanisms of PARPi resistance in BRAC1/BRCA2 mutated breast tumors in order to improve therapeutic options for breast cancer patients.

The following specific aims were proposed:

Specific Aim 1. Understand how PTIP-MLL3/4 and PARP1 confers chemoresistance and replication fork (RF) degradation in BRCA1/2-deficient cells.
Specific Aim 2. Determine the interactions of BRCA2, FANCD2, and POLQ in replication fork (RF) stability and Alt-EJ
Specific Aim 3. Assess mechanisms of PARPi resistance in mouse models and patient derived xenografts

- What was accomplished under these goals?

We have described major activities, specific objectives, significant results or key outcomes, conclusions and other achievements related to each specific aim in the following sections. Both Drs. Nussenzweig and D'Andrea have noted tasks for which they were responsible (**Site 1, Nussenzweig: NCI, NIH; Site 2, D'Andrea: DFCI**).

In our Year 1 annual report (2017), we better defined the relationships between PARylation, Brca2-deficiency and genomic instability that lead to cancer development and additionally identified a novel role of EZH2 (a histone methyltransferase) at replisomes in BRCA2-deficient tumors.

In our Year 2 annual report (2018), we made significant progress in understanding the mechanisms by which proteins such as PTIP and PARP1 contribute to replication fork stability and chemoresistance in BRCA2-deficient tumors. Using a mass spectrometry approach to identify proteins associated with stalled replicating forks, we identified numerous candidate proteins that could contribute to replication fork protection or destabilization. Additionally, we made progress in identifying proteins which contribute to replication fork stability and identified a new mechanism of chemoresistance in BRCA2-deficient tumors. We determined that loss of function of EZH2 in BRCA2-deficient cancer cells promotes PARPi resistance. Importantly, our results identify EZH2 expression as a biomarker of BRCA2-deficient tumor response to chemotherapy.

In our Year 3 annual report (2019), we demonstrated the functional importance of PTIP-53BP1 interactions in regulating homologous recombination (HR), and in doing so identified a new mechanism by which 53BP1 can suppress HR directly by blocking RAD51 loading post-resection. We also determined that restoration of replication fork (RF) protection and HR represent two independent means by which Brca1/2-deficient cells can become chemoresistance. We additionally confirmed a role of the POLQ protein in Alt-EJ and in replication fork stability.

In our final report, Drs. Nussenzweig and D'Andrea have made significant advances for major tasks as they relate to Specific Aims 1 and 2. Using a CRISPR-based genetic screen, Dr. Nussenzweig has identified a novel protein that confers chemosensitivity in Brca1-deficient cells (**Specific Aim 1**) and using an *in vitro* screen, Dr. D'Andrea has discovered a small molecule inhibitor of Polymerase Q which can selectively kill HR-deficient cancer cells (**Specific Aim 2**).

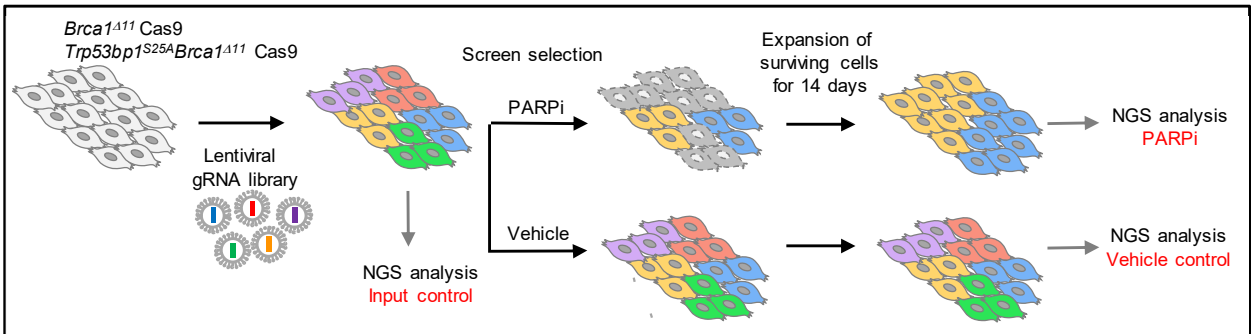
Specific Aim 1:

Major Task 2: Determine the mechanisms by which PARP1 modulates chemosensitivity in BRCA-deficient cells.

To more systematically explore the genetic basis of PARPi resistance, we conducted whole-genome CRISPR-Cas9 synthetic-viability screens in BRCA1-mutant cells. *Brca1*^{D11} and *Brca1*^{D11}*Trp53bp1*^{S25A} mouse embryonic fibroblasts (MEFs) were transduced with lentiviral libraries of single-guide RNAs (sgRNAs), and the resultant pools of edited cells were exposed to near-lethal doses of the clinically used PARPi olaparib for two weeks (**Figure 1, Top**). Gene enrichment or depletion was calculated for each cell line with positive and negative β-scores indicating positive and negative selection of a given gene deletion event, respectively.

Identification of THAP1 as a modifier of chemosensitivity in BRCA1-deficient cells

We identified multiple genes whose deletion have previously been shown to confer PARPi resistance in BRCA1-deficient cells, including Trp53bp1, Shld1, Shld2, Ctc1, Stn1 and Dynll1



Gene	-log$_{10}$(MAGeCK score)	Vehicle vs. PARPi		
		P value	FDR	Rank
Trp53bp1	10.237	2.52E-07	0.00248	1
Shld2	8.408	2.52E-07	0.00248	2
Parp1	6.695	7.55E-07	0.00495	3
Stn1	6.077	2.77E-06	0.01361	4
Shld1	5.362	1.61E-05	0.06076	5
Thap1	5.223	2.34E-05	0.06076	6
Ctc1	5.190	2.44E-05	0.06076	7
Atp6v1b2	5.154	2.74E-05	0.06076	8
Smu1	5.124	3.05E-05	0.06076	9
Hist1h2ac	5.085	3.30E-05	0.06076	10
Zfp706	5.065	3.40E-05	0.06076	11
Dynll1	4.438	0.00014369	0.17791	14

Figure 1: Identification of THAP1 as a modifier of chemosensitivity in BRCA1-deficient cells. (Top) Schematic of CRISPR-based screen for genes whose deletion confer resistance to the clinical PARP inhibitor (PARPi) olaparib in BRCA1-mutant MEFs (Brca1D11 and Brca1D11Trp53bp1S25A).
(Bot) Analysis of guide RNA enrichment in PARPi survivors by MAGeCK. Previously identified resistance genes are shown in blue and Thap1 is shown red.

(**Figure 1, Bot**), thereby confirming the validity of our approach. In addition, we identified Thap1 as a candidate gene whose deletion strongly imparts PARPi resistance in both *Brca1*^{D11}

and *Brca1*^{D11}*Trp53bp1*^{S25A} MEFs (**Figure 1, Bot**). Thap1 encodes the murine homolog of Thanatos-associated protein 1 (THAP1), a transcription factor that is causatively linked to adult-onset torsion dystonia type 6 (DYT6) in humans. To verify the results of the primary screens, we conducted competitive growth assays using sgRNAs targeting either Thap1 or Trp53bp1 and found that each led to the outgrowth of two independently-derived *Brca1*^{D11} MEF cell lines treated with PARPi (data not shown). By contrast, Thap1 deletion did not affect the PARPi sensitivity of BRCA1-proficient cells as measured by cell viability (data not shown).

These results support THAP1 as a bona fide modifier of chemosensitivity in mouse and human cell lines with BRCA1 deficiency.

Loss of THAP1 confers chemoresistance in BRCA1-deficient human tumors

To test whether dysregulated THAP1 has an impact on tumorigenesis, we extracted human tumor expression data from The Cancer Genome Atlas (TCGA) database. We found that low expression of THAP1 was significantly correlated with shorter progression-free survival (PFS) of patients with BRCA1-mutated serous ovarian carcinoma (**Figure 2**). Our results suggest that THAP1 loss alleviates PARPi and cisplatin hypersensitivity in BRCA1-deficient cells and tumors, and this effect is not species-, cell lineage- or BRCA1 mutation-specific.

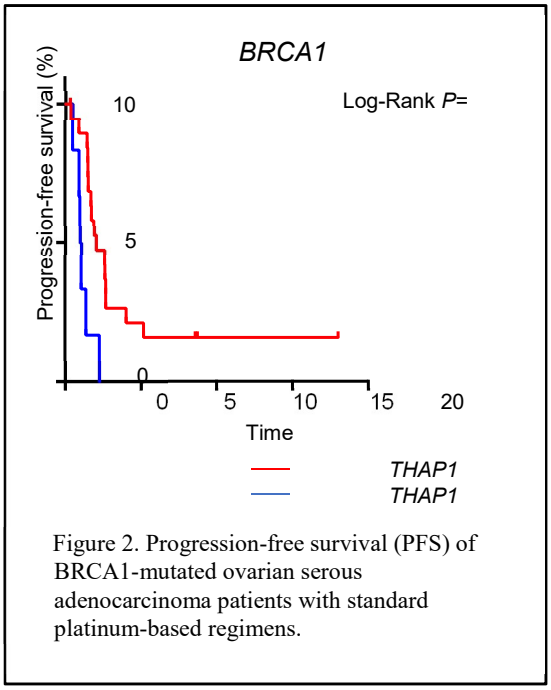


Figure 2. Progression-free survival (PFS) of BRCA1-mutated ovarian serous adenocarcinoma patients with standard platinum-based regimens.

THAP1 is a direct transcriptional regulator of Shld1 expression

Mutations in THAP1 have been identified as the cause of DYT6 (dystonia), which is associated with abnormal muscle contractions without overt neuropathological lesions. Given that most of the dystonia-causing mutations in THAP1 disrupt its sequence-specific DNA binding activity, we hypothesized that THAP1 may regulate the chemosensitivity of BRCA-deficient cells directly through transcriptional mechanisms. To investigate this possibility, we performed nascent RNA-sequencing (RNA-seq) to systematically probe for genes that are differentially expressed in Thap1-/- and Thap1-/-Brca1D11 MEFs, as compared to wild-type and Brca1D11 parental cells, respectively. 98 genes (57 downregulated and 41 upregulated) were common to both Thap1-/- and Thap1-/- Brca1D11 cells. To determine which of these genes are direct targets of THAP1, we examined published THAP1 ChIP-sequencing (ChIP-seq) dataset in mESCs which identified 2,758 unique THAP1 binding sites and nine differentially expressed genes in Thap1-deficient MEFs were also bound by THAP1 (**Figure 3, Top Left**). Among these genes we identified Shld1, the penultimate

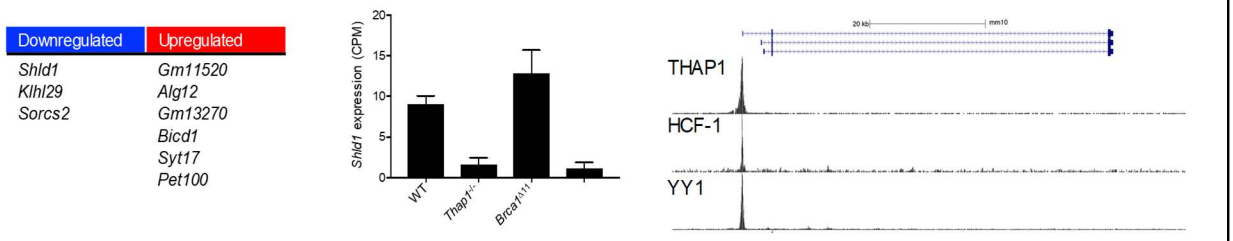


Figure 3. (Top Left): List of differentially expressed THAP1-bound genes that were common to both Thap1-/- and Thap1-/-Brca1D11 MEFs. (Top Middle) Levels of Shld1 gene expression in WT, Thap1-/-, Brca1D11 and Thap1-/-Brca1D11 MEFs, as detected by nascent RNA-seq. (Top Right) Binding of THAP1at the Shld1 gene locus.

component of the Shieldin complex and a known effector of 53BP1. The Shieldin complex, consisting of SHLD1, SHLD2, SHLD3 and REV7, were identified as critical effectors of 53BP1-mediated end protection downstream of RIF1. In addition, Shieldin has been shown to directly suppress RAD51 nucleofilament formation on ssDNA and notably, depletion of any Shieldin complex component restores HR in BRCA1-deficient cells, and reduced expression of Shieldin is associated with the acquisition of PARPi resistance in PDX models of BRCA1-deficient breast cancers. The Shld1 promoter contains a consensus THAP1 binding sequence (THABS) motif where a strong THAP1 ChIP-seq signal was detected (**Figure 3, Top Right**) and our RNA-seq datasets confirmed that Shld1 expression was significantly decreased in Thap1-/- and Thap1-/-BRCA1D11/D11 MEFs (**Figure 3, Top Middle**).

THAP1 inhibits DNA end resection and RAD51 nucleofilament formation

We found that deletion of THAP1 rescued both irradiation-induced RPA foci and RAD51 nucleofilament formation in Brca1D11 MEFs to wild-type levels (**Figure 4, Top Left and Right**). Thus, loss of

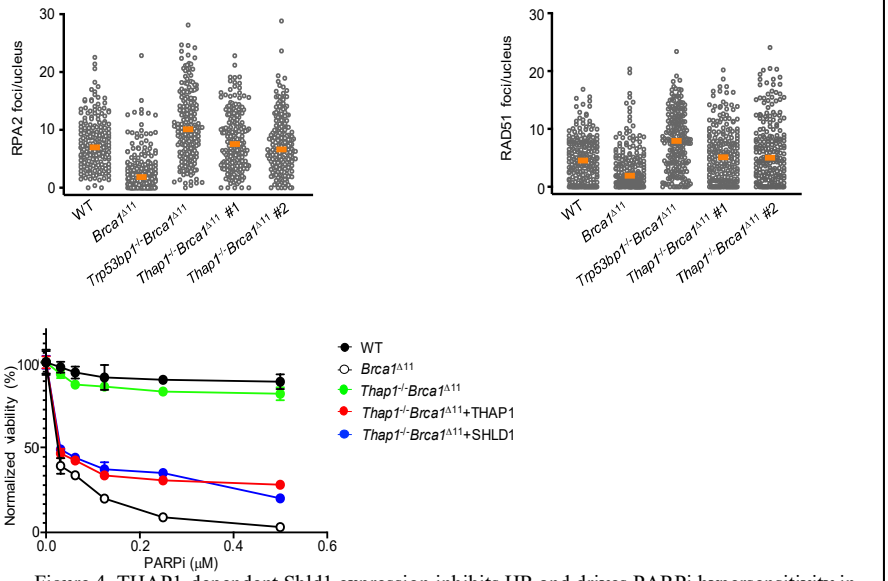


Figure 4. THAP1-dependent Shld1 expression inhibits HR and drives PARPi hypersensitivity in BRCA1-deficient cells. (Top, Left and Right) Quantification of RPA2 and RAD51 foci in individual EdU-positive (S-phase) nuclei of WT, Brca1D11, Trp53bp1-/-Brca1D11 and two individual clones of Thap1-/-Brca1D11 MEFs. Cells were irradiated with 10 Gy and analyzed 4 h post-IR. (Bot) Viability of WT, Brca1D11, Thap1-/-Brca1D11 and Thap1-/-Brca1D11 MEFs complemented with either THAP1 (Thap1-/-Brca1D11+THAP1) or SHLD1 (Thap1-/-Brca1D11+SHLD1) cDNA, as measured by CellTiter-Glo seven days after PARPi treatment.

THAP1 promotes end-resection and HR proficiency in BRCA1-deficient cells, which could explain why doubly deficient cells acquire resistance to PARPi and cisplatin.
To provide definitive proof that the observed PARPi resistance in Thap1-/-Brca1D11 MEFs was due to decreased Shld1 expression, we overexpressed SHLD1 in Thap1-/-Brca1D11 MEFs. Overexpression of SHLD1 in Thap1-/-Brca1D11 MEFs restored their hypersensitivity to PARPi. Addback of THAP1 in Thap1-/-Brca1D11 MEFs also restored PARPi hypersensitivity (**Figure 4, Bot**). Moreover, SHLD1 or THAP1 overexpression in Thap1-/-Brca1D11 MEFs increased chromosomal aberrations to levels comparable to or higher than in Brca1D11 parental cells (data not shown). Thus, THAP1-dependent Shld1 transcription is essential for maintaining the functionality of Shieldin.

Specific Aim 2:

Major Task 6: To identify and characterize proteins which cooperate with FANCD2 and POLQ in replication fork stability and Alt-EJ. *In this section, Dr. D'Andrea has cited the six publications from our laboratory which were supported by this grant from the Department of Defense.*

Subtask 1: To identify and characterize proteins which cooperate with FANCD2 and POLQ in replication fork stability and Alt-EJ using iPOND/mass spectrometry (time line:6-18 months) **(Site 2, DFCI, Dr. D'Andrea)**.

We have not performed the experiments proposed for this subtask.

Subtask 2: Validation of FANCD2 and POLQ interacting proteins identified by mass spec using coimmunoprecipitation, shRNA knockdown, changes in POLQ nuclear foci and Alt-EJ template assays (proposed timeline: 18-30 months) **(Site 2, DFCI, Dr. D'Andrea)**.

We have not performed the experiments proposed for this subtask.

Rather than focus on Subtask 1 and Subtask 2, we focused on the following highly related projects which turned out to be much more productive. These projects led to novel insights into the proteins **REV7 and TRIP13** and their role in breast cancer. During the last year, we have further evaluated the role of POLQ in alternative end joining (Alt-EJ). We published a review article on Alt-EJ (*Patterson-Fortin J, D'Andrea AD. Exploiting the Microhomology-Mediated End-Joining Pathway in Cancer Therapy. Cancer Res. 2020 Jul 10:canres.1672.2020. doi: 10.1158/0008-5472.CAN-20-1672*). We have demonstrated that inhibition of POLQ blocks Alt-EJ and selectively kills HR-deficient tumor cells. Specifically, we devised an in vitro screen to identify a small molecule inhibitor of Polymerase Q. The POLQ protein has an N-terminal ATPase domain and a C-terminal Polymerase domain. Since the N-terminal ATPase domain has strong ATPase activity in vitro, we were able to perform a high throughput screen. We identified Novobiocin (NVB) as a potent and specific inhibitor of the POLQ-ATPase. We have performed several mouse xenograft and GEMM models to confirm that NVB kills BRCA1-deficient tumor cells, as well as other HR-deficient tumor cells.

As part of Subtask 2, we have identified a new mechanism by which BRCA1-deficient PARP inhibitor sensitive tumor cells became PARP inhibitor resistant. We have shown that BRCA1-deficient tumor cells can amplify the gene TRIP13 as a mechanism for restoring HR repair. Specifically, the TRIP13 gene encodes an ATPase which can inactivate REV7, leading to an opening and inactivation of the SHLD complex. The relevance of the SHLD complex results in increased dsb end resection, increased RAD51 loading, and increased HR, leading to acquired PARPi resistance. This work has recently been published in two papers (*Clairmont CS*, Sarangi P*, Ponnienselvan K, Galli LD, Csete I, Moreau L, Adelmant G, Chowdhury D, Marto JA, D'Andrea AD. TRIP13 Regulates DNA Repair Pathway Choice through REV7 Conformational Change. Nature Cell Biol, 2020.) (Sarangi P, Clairmont CS, D'Andrea AD. Disassembly of the Shieldin Complex by TRIP13. Cell Cycle. 2020 Jul;19(13):1565-1575. doi: 10.1080/15384101.2020.1758435. Epub 2020*)

TRIP13 overexpression promotes HDR and PARP inhibitor resistance in BRCA1-deficient breast cancer cells

Since our last progress report, we proceeded to directly assay the effect of TRIP13 in BRCA1-deficient breast cancers using patient-derived cell lines. We assessed TRIP13 protein levels in a panel of normal and patient-derived breast cancer cell lines and observed wide variation (**Figure 5A**). We focused on the SUM149PT cell line, as it is a widely used BRCA1-deficient model

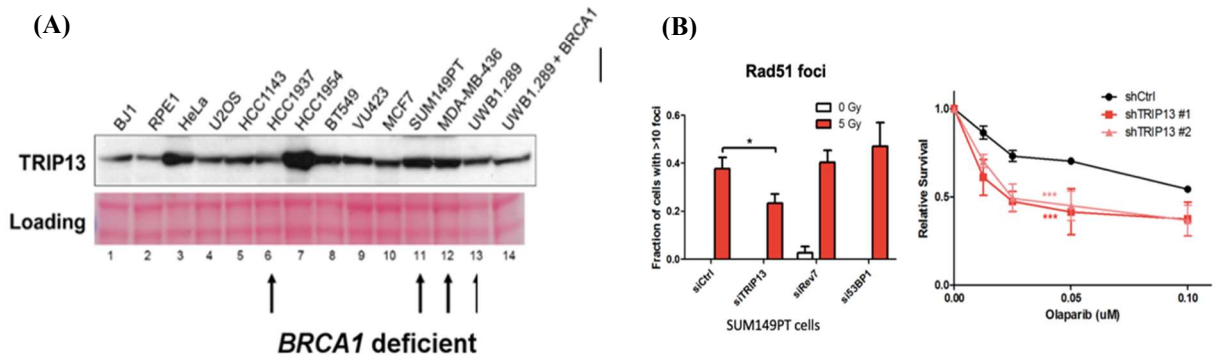


Figure 5. BRCA1 deficient breast and ovarian cancer cell lines have enhanced TRIP13 Expression.
(A) The indicated breast and ovarian cancer cell lines have upregulated the expression of the TRIP13 protein, secondary to TRIP13 gene amplification. Interestingly, the SUM149 cell line has biallelic BRCA1 mutations but is still resistant to PARP inhibitor (B) Knockdown of TRIP13 decreases HR proficiency and causes Olaparib sensitivity of a breast cancer cell line

with high TRIP13 expression. Despite their BRCA1 deficiency, SUM149PT cells were partially HDR-proficient as monitored by RAD51 focus formation and Olaparib resistance. We therefore asked whether TRIP13 promotes HDR in these cells. Indeed, stable knockdown of TRIP13 significantly impaired both RAD51 focus formation and Olaparib resistance (**Figure 5B**). Interestingly, knockdown of REV7 or 53BP1 did not appreciably increase RAD51 focus formation in SUM149PT cells, unlike the case in other BRCA1-deficient contexts. This is likely because the 53BP1-REV7-Shieldin axis in these cells is already inactivated by their high TRIP13 expression. This work was also described in our recent review article which was supported by

this DOD grant. (*Sarangi P, Clairmont CS, D'Andrea AD. Disassembly of the Shieldin Complex by TRIP13. Cell Cycle. 2020 Jul;19(13):1565-1575. doi: 10.1080/15384101.2020.1758435. Epub 2020*). Moreover, using CRISPR screens, we were able to identify novel sgRNAs, and hence gene products, which further regulate DNA repair and

homologous recombination activity in breast cancer cells (1. *Li F, Kozono D, Deraska P, Branigan T, Dunn C, Zheng X, Parmar K, Nguyen H, DeCaprio J, Shapiro GI, Chowdhury D, D'Andrea AD. CHK1 Inhibitor Blocks Phosphorylation of FAM122A and Promotes Replication Stress. Mol Cell, 2020;* 2. *Cai, M., C. Dunn, W. Chen, B. Kochupurakkal, H. Nguyen, L.A. Moreau, G. I. Shapiro, D. Kozono and A. D. D'Andrea. "Cooperation of the ATM and Fanconi Anemia/BRCA Pathways in Double Strand Break End Resection." Cell Rep. 2020 Feb 18;30(7):2402-2415.e5.;* 3. *Li, N., J. Wang, S. S. Wallace, J. Cheng, J. Zhou and A. D. D'Andrea "Cooperation of the NEIL3 and Fanconi Anemia/BRCA Pathways in Interstrand Crosslink Repair." Nucleic Acids Res. 2020 Jan 25. pii: gkaa038. doi: 10.1093/nar/gkaa038.*)
Major Task 8: Evaluate POLQ and PARP inhibitor resistance in murine xenograft and GEMM models **(Site 2, DFCI, Dr. D'Andrea)**

As described above in Major Task 6, we have identified **Novobiocin (NVB) as a novel inhibitor of POLθ** (*reviewed in Patterson-Fortin and D'Andrea, Cancer Research, 2020*) and evaluated whether POLθ inhibition could represent a novel therapeutic option for PARPi-resistant tumors, we have tested the efficacy in vivo of NVB in the patient-derived xenograft (PDX) models. The PDX model, DF83, was generated from an HR-deficient (loss of RAD51C expression), human high-grade serous ovarian carcinoma (**Figure 6A**). NSG mice bearing the HR-deficient DF83

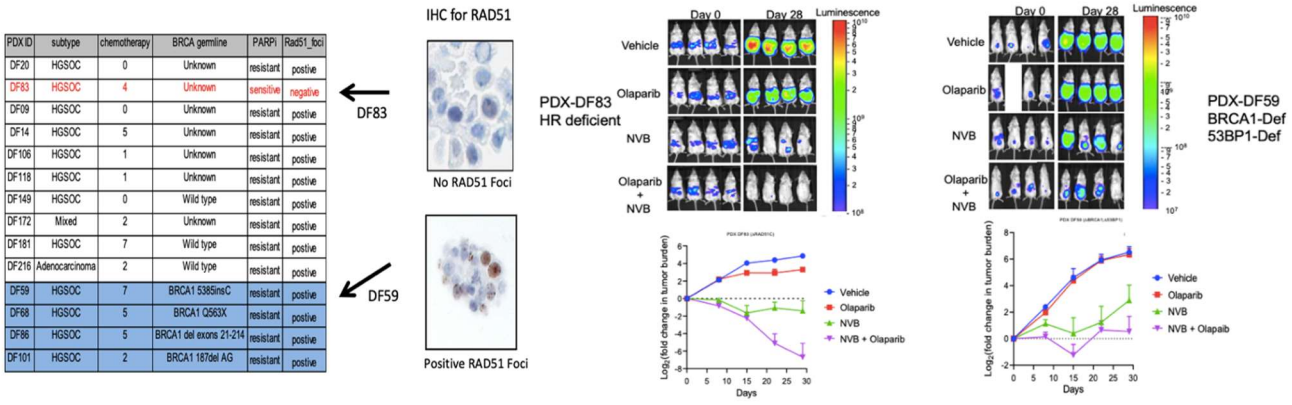


Figure 6. Ovarian Cancer PDX models. (A) The DF83 model has the absence of RAD51 foci and is PARP inhibitor sensitive. The DF59 model has the presence of RAD51 foci and is PARP inhibitor resistant (B). Polymerase Theta Inhibition with Novobiocin overcomes PARP Inhibitor Resistance in vivo. This work was published in Zhou et al, 2020.

ovarian cancer PDX model were treated with olaparib, NVB, or a combination of the two, and tumor growth was monitored by bioluminescence imaging (BLI). In order to determine the combined effect with NVB, olaparib was delivered at a submaximal dose (50 mg/kg daily), and we observed only a small degree of tumor growth inhibition. Treatment of DF83 with NVB

alone (75 mg/kg, twice a day) led to tumor regression, while tumors in vehicle-treated mice showed exponential growth (**Figure 6B**). Strikingly, when the DF83 tumor-bearing mice were given NVB and olaparib in combination, we observed complete tumor regression, and few tumor cells were detectable via BLI on day 28 in this group. These in vivo results suggest NVB may be useful in combination with PARPi to achieve better efficacy and prevent drug resistance in HR-deficient tumors. No significant toxicity to normal mouse tissue was observed from the drug combination.

Next, we evaluated the efficacy of NVB alone and in combination with olaparib in the PARPi-resistant PDX model DF59, which was isolated from a patient with a germline BRCA1 mutation and acquired PARPi resistance. This model harbors biallelic mutations in the TP53BP1 locus, providing a mechanism of the acquired PARPi resistance. In the DF59 PDX model, there was no response to olaparib monotherapy, but NVB monotherapy substantially reduced tumor growth, indicating a lack of cross-resistance between PARP inhibition and NVB. Importantly, the combination of NVB and olaparib further inhibited DF59 tumor growth, with tumor regression in the first two weeks (**Figure 6B**). In contrast to DF83 and DF59 models, the BRCA1 WT, HR proficient PDX model DF149 was resistant to both drugs and to the combination (data not shown). These data demonstrate that NVB can inhibit the growth of a BRCA1-mutated, PARPi-resistant tumors in vivo and that the combination of NVB and a PARPi is particularly effective in treating at least some PARPi-resistant tumors. These new data are posted in our recent Bioarchives paper (*Zhou J, Gelot C, Pantelidou C, Li A, Yücel H, Davis RE, Farkkila A, Shapiro GI, Blagg BSJ, Ceccaldi R, D'Andrea AD. Polymerase Theta Inhibition Overcomes PARP Inhibitor Resistance. bioRxiv 2020. https://doi.org/10.1101/2020.05.23.111658*). This paper also demonstrates the mechanism of NVB activity-namely, the increase in DSB end resection which leads to cell death.

Summary of the key research accomplishments: Our studies have revealed that the evolutionarily conserved transcription factor THAP1 promotes transcription of the Shld1 gene in mammalian cells. Our observation that loss of THAP1 restores HR and chemoresistance in both Brca1D11 (resection-deficient) and Brca1D11Trp53bp1S25A (RAD51 loading-deficient) MEFs is consistent with a critical function of Shieldin pre- and post-resection. We have also identified Novobiocin (NVB) as a novel small molecule inhibitor of POLθ and determined that POLθ inhibition represents a novel therapeutic option for PARPi-resistant tumors.

What opportunities for training and professional development did the project provide?

Nothing to report

- How were the results disseminated to communities of interest?

Results were shared with the scientific community via informal discussions, posters and presentations at scientific meetings and through publications in peer-reviewed journals

- What do you plan to do during the next reporting period to accomplish the goals and objectives?

Nothing to report

4. IMPACT:

Nothing to Report

What was the impact on the development of the principal discipline(s) of the project?

We have shown that chemosensitivity in *Brca1/2*-deficient cells can result from either loss of RF protection or loss of HR (and perhaps both in some cases). Conversely, restoration of RF stability or HR can independently confer chemoresistance. These findings have a high impact on our ability to identify patient populations who may/may not benefit from chemotherapy and may open new avenues for targeted intervention. We have additionally shown that mutations in different domains of Brca1 can have distinct effects on HR. This suggested that all *Brca1* mutant tumors may not respond to chemotherapy in the same manner.

What was the impact on other disciplines?

Nothing to Report

What was the impact on technology transfer?

Nothing to Report

What was the impact on society beyond science and technology?

Nothing to Report

5. CHANGES/PROBLEMS:

Changes in approach and reasons for change
Nothing to Report

Actual or anticipated problems or delays and actions or plans to resolve them

Nothing to Report

Changes that had a significant impact on expenditures

Nothing to Report

Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents:

Nothing to Report

Significant changes in use or care of human subjects:
Not applicable

Significant changes in use or care of vertebrate animals:

Nothing to Report

Significant changes in use of biohazards and/or select agents:

Nothing to Report

6. PRODUCTS

Publication: Journal publications.

DOD BCRP Publications October 2019-Present (Nussenzweig) citing DOD support in the acknowledgments

1. Shinoda, K, Zong D, Callen E, Wu W, Dumitrache L.C, Belinky F, Wong N, Ishikawa M, Konstantinopoulos P.A, Ehrlich M, McKinnon P.J, Nussenzweig A. "The Dystonia Gene THAP1 Controls Physiological DNA repair and Chemosensitivity in BRCA-deficient Cells." **Molecular Cell** (in revision)

2. van Wietmarschen N, Sridharan S, Nathan WJ, Tubbs A, Chan EM, Callen E, Wu W, Belinky F, Tripathi V, Wong N, Foster K, Noorbakhsh J, Garimella K, Cruz-Migoni A, Sommers JA, Huang Y, Borah AA, Smith JT, Kalfon J, Kesten N, Fugger K, Walker RL, Dolzhenko E, Eberle MA, Hayward BE, Usdin K, Freudenreich CH, Brosh RM Jr, West SC, McHugh PJ, Meltzer PS, Bass AJ, Nussenzweig A. "Repeat expansions confer WRN dependence in microsatellite-unstable cancers." **Nature**. 2020 Oct;586(7828):292-298. doi: 10.1038/s41586-020-2769-8. Epub 2020 Sep 30

3. Sciascia N, Wu W, Zong D, Sun Y, Wong N, John S, Wangsa D, Ried T, Bunting SF, Pommier Y, Nussenzweig A. "Suppressing proteasome mediated processing of topoisomerase II DNA-protein complexes preserves genome integrity." **Elife**. 2020 Feb 14;9:e53447. doi: 10.7554/eLife.53447.

4. Paiano J, Wu W, Yamada S, Sciascia N, Callen E, Paola Cotrim A, Deshpande RA, Maman Y, Day A, Paull TT, Nussenzweig A. "ATM and PRDM9 regulate SPO11-bound recombination

intermediates during meiosis." **Nat Commun**. 2020 Feb 12;11(1):857. doi: 10.1038/s41467-020-14654-w.

5. Shinoda K, Maman Y, Canela A, Schatz DG, Livak F, Nussenzweig A. "Intra-Vκ Cluster Recombination Shapes the Ig Kappa Locus Repertoire." **Cell Rep**. 2019 Dec 24;29(13):4471-4481.e6. doi: 10.1016/j.celrep.2019.11.088.

DOD BCRP Publications January 2020-Present (D'Andrea) citing DOD support in the acknowledgments

1. Li N, J. Wang J, Wallace S.S, Cheng J, Zhou J, D'Andrea AD. "Cooperation of the NEIL3 and Fanconi Anemia/BRCA Pathways in Interstrand Crosslink Repair." **Nucleic Acids Res**. 2020 Jan 25. pii: gkaa038. doi: 10.1093/nar/gkaa038.

2. Cai, M, Dunn C, Chen W, Kochupurakkal B, Nguyen H, Moreau L.A, Shapiro G.I, Kozono D, and D'Andrea AD. "Cooperation of the ATM and Fanconi Anemia/BRCA Pathways in Double Strand Break End Resection." **Cell Rep**. 2020 Feb 18;30(7):2402-2415.e5.

3. Zhou J, Gelot C, Pantelidou C, Li A, Yücel H, Davis RE, Farkkila A, Shapiro GI, Blagg BSJ, Ceccaldi R, D'Andrea AD. Polymerase Theta Inhibition Overcomes PARP Inhibitor Resistance. **bioRxiv** 2020. https://doi.org/10.1101/2020.05.23.111658.

4. Sarangi P, Clairmont CS, D'Andrea AD. Disassembly of the Shieldin Complex by TRIP13. **Cell Cycle**. 2020 Jul;19(13):1565-1575. doi: 10.1080/15384101.2020.1758435. Epub 2020 May 18.

5. Patterson-Fortin J, D'Andrea AD. Exploiting the Microhomology-Mediated End-Joining Pathway in Cancer Therapy. **Cancer Res**. 2020 Jul 10:canres.1672.2020. doi: 10.1158/0008-5472.CAN-20-1672.

6. Li F, Kozono D, Deraska P, Branigan T, Dunn C, Zheng X, Parmar K, Nguyen H, DeCaprio J, Shapiro GI, Chowdhury D, D'Andrea AD. CHK1 Inhibitor Blocks Phosphorylation of FAM122A and Promotes Replication Stress. **Mol Cell**. 2020 Oct 21:S1097-2765(20)30691-2. doi: 10.1016/j.molcel.2020.10.008.

Books or other non-periodical, one-time publications.

Nothing to Report

Other publications, conference papers, and presentations.

Dr. Andre Nussenzweig (Site 1, NCI, NIH)

2019

- Invited Speaker, 50th. Annual Meeting of the Environmental Mutagenesis and Genomics Society (EMGS), Washington D.C., September 2019.
- Keynote Speaker, ASBMB-BSC Symposium on the Interplay between Epigenetic Regulation and Genome Integrity, Suzhou, China, October 2019.
- Invited Speaker, Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences, Shanghai, China, October 2019.
- Invited Speaker, Department of Cancer Genetics and Epigenetics, Beckman Research Institute, City of Hope, Duarte, CA, October 2019.
- Invited speaker, 9th NCI Symposium on Chromosome Biology, National Institutes of Health, Bethesda, MD, November 2019.
- Invited speaker, National Institutes of Health Wednesday Afternoon Lecture Series (WALS), Bethesda, MD. Postponed.
- Invited speaker, Pruitt Memorial Lecture, Roswell Park Comprehensive Cancer Center, Buffalo, New York. Postponed.
- Invited speaker, Mechanisms of Recombination Conference, Cambridge, UK. Postponed.
- Invited speaker, 85th. Cold Spring Harbor Symposium on Quantitative Biology "Genome Stability and Integrity, Cold Spring Harbor, New York. Postponed.
- Keynote speaker, Nucleotide Excision Repair and Crosslink Repair: Molecules to Mankind, Smolenice, Slovakia. Postponed.
- Invited speaker, Institute of Molecular Genetics of the Czech Academy of Sciences, Prague, Czech Republic. Postponed.
- Keynote speaker, Gordon Research Conference on Genomic Integrity, DNA Repair and Human Diseases, Kowloon, Hong Kong. Postponed.
- Invited Speaker, Instituto de Medicina Molecular (iMM), Lisbon, Portugal, Postponed.

2020

- Invited Speaker, DFHCC Connect: Science Virtual Seminar Series, Boston, MA, July 2020.
- Invited Speaker, Genomic Stability and DNA Repair Keystone Virtual Symposia, Breckenridge, Colorado, September 2020.
- Invited Speaker, Virtual Seminar Department of Cell Biology and Division of Biology, Albert Einstein College of Medicine, New York, September 2020.
- Invited Speaker, Virtual Cancer Center Research Seminars, O'Neal Comprehensive Cancer, University of Alabama at Birmingham (UAB), November 2020.
- Invited speaker, EMBO Virtual Workshop Chromatin Dynamics and Nuclear Organization in Genome Maintenance, Leiden, The Netherlands, December 2020.

Dr. Alan D'Andrea (Site 2, DFCI)

2019

- Keystone Symposium: DNA Replication and Genome Instability: From Mechanism to Disease, Snowbird, Utah
- Stand Up To Cancer Scientific Summit, Santa Monica, CA
- DNA Damage Response Therapeutics Summit, Boston, MA
- Gordon Research Conference on Mammalian DNA Repair, Ventura, CA
- ESMO Targeted Anticancer Therapies Conference, Paris, France
- DF/HCC Breast and Ovarian Cancer Symposium, Boston, MA
- NCI Board of Scientific Counselors Meeting, Bethesda, MD
- Sierra Oncology External Advisory Board Meeting, Atlanta, GA
- AACR Annual Meeting, Atlanta, GA
- Susan F. Smith Center Executive Council Breakfast, Boston, MA
- Mt. Sinai Tisch Cancer Institute External Advisory Board Meeting, New York, NY
- Merck KGA DDR Summit, Boston, MA
- Perspectives in Science Seminar, Fred Hutchinson Cancer Research Center, Seattle, Washington
- Cyteir Scientific Advisory Board Meeting, Lexington, MA
- SU2C Pancreatic Cancer Mini-Summit, Philadelphia, PA
- Aspen Cancer Conference, Aspen, CO
- SU2C Merck-Catalyst Review Meeting, Philadelphia, PA
- AACR Advances in Ovarian Cancer Research Conference, Atlanta, GA
- FARF Annual Scientific Symposium, Chicago, IL
- Ideaya Scientific Advisory Board Meeting, San Francisco, CA
- Invited Seminar, Biomedical Research Institute, National Autonomous University of Mexico (UNAM), Mexico City, Mexico
- BCRF Grantees Meeting, New York, NY
- DFCI Joint Visiting Committee Meeting, Boston, MA
- BRCA Answers from Cancer Interactome Structures (BACIS) Scientific Advisory Board Meeting, MD Anderson Cancer Center, Houston, TX
- Susan F. Smith Center for Women's Cancers Executive Council Beyond Boston Luncheon, Wellesley, MA
- isDDRHD Meeting, Shenzhen, China
- Society for Neuro-Oncology (SNO) Meeting, Phoenix, AZ

2020

- Vanderbilt Cancer Center ESAB Meeting, Nashville, TN
- Wienman Symposium, Honolulu, HI
- SU2C Scientific Summit, Santa Monica, CA
- PARP and DDR Inhibitors Summit, Boston, MA
- FARF SAB Meeting, Orlando, FL
- Gray Foundation Summit, New York, NY

- Keynote Address, 2020 University of Massachusetts Medical School MD/PhD Retreat, Worcester, MA
- AACR Virtual Annual Meeting
- SU2C Pancreatic Cancer Collective Symposium
- Cell Press LabLinks Symposium: Targeting the DNA Damage Response in Cancer
- Ideaya Scientific Advisory Board Meeting
- MD Anderson Cancer Center Ovarian SPORE Virtual EAB Meeting
- Rivkin Ovarian Cancer Research Symposium
- Keynote Address, Keystone Symposium: Genomic Stability and DNA Repair

- **Website(s) or other Internet site(s)**

Nothing to Report

- **Technologies or techniques**

Nothing to Report

- **Inventions, patent applications, and/or licenses**

Nothing to Report

- **Other Products**

Nothing to Report

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS (Site 1, NCI, NIH)

Name:	Andre Nussenzweig
Project Role:	Principal Investigator
Researcher Identifier (e.g. ORCID ID):	N/A
Nearest person month worked:	3
Contribution to Project:	Dr. Nussenzweig was responsible for the project management and collaboration with Dr. D'Andrea.
Funding Support:	Dr. Nussenzweig is supported by the NIH Intramural Research Program

Name:	Elsa Callen
Project Role:	Staff Scientist
Researcher Identifier (e.g. ORCID ID):	N/A
Nearest person month worked:	3
Contribution to Project:	Dr. Callen has been instrumental in developing the replication fork protection assay and in assessing fork stability in a variety of genetic contexts
Funding Support:	Dr. Callen is supported by the NIH Intramural Research Program

Name:	Dali Zong
Project Role:	Scientist
Researcher Identifier (e.g. ORCID ID):	N/A
Nearest person month worked:	6
Contribution to Project:	Dr. Zong has been a major contributor to the laboratory's research efforts in understanding the mechanisms of DNA repair pathway choice. He has been responsible for conducting fundamental studies on the molecular mechanisms of DNA damage/repair and has developed a number of sophisticated genetic, biological, biochemical and especially imaging assays to further these studies.
Funding Support:	Dr. Zong is supported by the NIH Intramural Research Program

Name:	Kenta Shinoda
Project Role:	Post-doctoral fellow
Researcher Identifier (e.g. ORCID ID):	N/A
Nearest person month worked:	6

Contribution to Project:	Dr. Shinoda has performed work in the area of PARP1 biology and how affecting co-factor (NAD+) levels influences enzymatic activity and genome stability in BRCA-deficient cells
Funding Support:	Dr. Shinoda is supported by the NIH Intramural Research Program

PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS (Site 2, DFCI)

Name:	Alan D'Andrea
Project Role:	Principal Investigator
Researcher Identifier (e.g. ORCID ID):	0000-0001-6168-6294
Nearest person month worked:	No change.
Contribution to Project:	No change.
Funding Support:	Dr. D'Andrea is supported by the following grants: DoD *BC151331P1*, 5R01HL52725-24, BCRF, P50CA168504, SU2C-AACR-DT16-15, SU2C-AACR-DT16-15/Supplement, PCF, 5P01HL048546, Fanconi Anemia Research Fund, Inc, Merck & Co., Inc., Sierra Oncology, Inc., Lilly, SU2C/Lustgarten Foundation, Bridge Project, 1P01CA228696-01A1, GI SPORE, SU2C Phillip A. Sharp Innovation Award, Basser Initiative, Breast Cancer SPORE

Name:	Jessica Filiatrault
Project Role:	Technician
Researcher Identifier (e.g. ORCID ID):	N/A
Nearest person month worked:	4 calendar months

Contribution to Project:	She assisted Dr. Zhou in performing cell-based and biochemical assays.
Funding Support:	Ms. Filiatrault is supported by this DoD grant *BC151331P1* and Ovarian Tumor Living Biobank grant.

Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?

Dr. Nussenzweig has added Dr. Kenta Shinoda back to this project for conducting CRISPR-related screening experiments (described above in Specific Aim 1, Major Task 2).

Dr. D'Andrea has the following changed in his other support:

Ongoing Research Support

1. 5R01HL52725-25 (PI: D'Andrea) 1/1/20-12/31/23
NIH/NHLBI
Molecular Pathogenesis of Fanconi Anemia
This study investigates the molecular pathogenesis of Fanconi Anemia using the recently cloned human Fanconi Anemia Complementation Group C (FACC) cDNA and a newly developed antiserum against the FACC polypeptide.

2. Breast Cancer Research Foundation (D'Andrea) 10/1/11-9/30/21
Extending the use of PARP Inhibitors for Triple Negative Breast Cancer Therapy
Goals/Aims: To determine whether the combination of Velcade plus ABT-888 has enhanced cytotoxic activity for TNBC cell lines in vitro, compared to each agent alone; To generate colonies of mice bearing human TNBC orthotopic implants (derived from TNBC cell lines or from TNBC patient biopsies) and to treat these mice with the combination of Velcade plus ABT-888. As a subaim, we will examine the tumors dissected from these mice with numerous pharmacodynamic markers (i.e., immunohistochemistry for pBRCA1, FANCD2-Ub, and RAD51).

3. 5P01HL048546-23 (PI: Grompe)
Pathophysiology and Treatment of Fanconi/Project 2 and Core 9/1/16-5/31/21
NIH
The major goals of this study are 1) determine the mechanism by which TGF-β inhibitors promote FA cellular growth and regulate DNA repair, 2) to determine whether inhibition of TGF-β pathway rescues hematopoietic defects in FA mouse models, and 3) to determine whether inhibition of TGF-β pathway rescues hematopoietic defects in primary bone marrow cells from FA patients.
Role: Project PI

4. Richard & Susan Smith Family Foundation (PI: D'Andrea) 9/1/17-3/31/21
Ovarian Tumor Living Biobank
Major goals are to establish the Susan F. Smith Center Living Biobank of recurrent and primary ovarian tumors at Dana-Farber, and to disseminate the knowledge gained from the initiative to investigators with specialties that span multiple cancer types.
Role: PI

5. SU2C/Lustgarten Foundation (PI: D'Andrea) 11/1/18/-12/31/22
Exploiting DNA Repair Gene Mutations in Pancreatic Cancer
Major goal of this study is to identify the most potent combinatorial strategy, along with predictive biomarkers, that will inform the design of a clinical trial for PDAC patients.
Role: PI

6. Bridge Project (PI: D'Andrea) 3/1/19-2/28/21
Nanoparticle-mediated Drug Delivery to Ovarian Cancer Organoid Cultures
Major goal of this study is to understand; 1) the mechanisms that lead to HGSC sensitivity and resistance to various immune and DNA damage repair therapies and 2) what the best method of delivering specific therapies for these defects is.
Role: PI

7. NIH 1P01CA228696-01A1 (PI: Kantoff) 7/1/19-6/30/24
Project 3: Functional Evaluation and Interpretation of DDR Variants in Prostate Cancer
Major goal of this proposal is to determine whether specific DNA repair gene mutations in prostate cancers are deleterious and vulnerable to targeted therapy.
Role: Project Leader

8. DF/HCC SPORE in Gastrointestinal Cancer 7/1/19-6/30/24
P50CA127003 (Bass/El-Bardeesy) Project 3
Major goal of this proposal is to define optimal strategies for identifying PDAC patients with DDR deficiency and to conduct innovative PARP inhibitor treatment trials to improve care for these patients.
Role: Co-Project Leader

9. SU2C Phillip A. Sharp Innovation in Collaboration Award (D'Andrea/Cubillos-Ruiz) 6/1/19-5/31/2021
Resistance to PARP inhibitor plus anti-PD1 therapy driven by ER stress and bioactive lipids in ovarian cancer
Major goals of this proposal are to evaluate the status of LPA/PERK-controlled signatures in responder vs. non-responder patients, and to determine whether inhibition of LPA-PERK sensitizes OC to Niraparib and Pembrolizumab.
Role: Co-PI

10. Basser Initiative at Gray Foundation, Team Science Grants (PI: Sung) 7/1/19-6/30/23
Dissection of BRCA-mediated Tumor Suppression Pathways

Major goals of this study are to evaluate the PARPi and platinum-resistant mechanisms, and to determine their impact on patient drug response.
Role: Specific 4 Project Leader

11. P50CA168504-06A1 (Winer) 7/01/19 – 06/30/24
NIH/NCI
Dana-Farber/ Harvard Cancer Center SPORE in Breast Cancer – Developmental Research Program
The overall goal of the Developmental Research Program (DRP) is to advance high quality, novel, early-phase research, to foster new ideas in breast cancer research and to move research projects from a pilot stage to a level where external funding for more mature research is possible. A secondary goal of the DRP is to create opportunities for the career development of junior faculty or senior investigators who are interested in transitioning into breast cancer research. The DRP will be co-led by Alan D'Andrea, MD and Nikhil Wagle, MD. They will work with a standing committee made up of SPORE and DF/HCC investigators to review applications.
Role: Co-Director, DRP

12. DF/HCC Ovarian SPORE (D'Andrea/Spriggs/Matulonis) 8/5/20-7/31/2025
NIH 1P50CA240243-01A1
DF/HCC Ovarian SPORE
Major goal of this proposal is to make translational progress in understanding resistance mechanisms that will lead to meaningful and improved outcomes of women with newly diagnosed ovarian cancer and those with relapsed ovarian cancer.
Role: Contact PI: D'Andrea

What other organizations were involved as partners?

Nothing to report

8. SPECIAL REPORTING REQUIREMENTS:

COLLABORATIVE AWARDS: We have marked the tasks assigned to us and accordingly we have provided a progress made for each task.

QUAD CHARTS: Not applicable

9. APPENDICES: Nothing to report

The Dystonia Gene THAP1 Controls DNA Double Strand Break Repair Choice

Kenta Shinoda[1], Dali Zong[1], Elsa Callen[1], Wei Wu[1], Lavinia C. Dumitrache[2], Frida Belinky[1], Nancy Wong[1], Momoko Ishikawa[1], Andre Stanlie[1], Michelle Ehrlich[3], Peter J. McKinnon[2], Andre Nussenzweig[1‡]

[1] Laboratory of Genome Integrity, National Cancer Institute, NIH, Bethesda, MD, USA

[2] St. Jude Translational Neuroscience, Center for Pediatric Neurological Disease Research, Depts. Genetics & Cell Mol. Biology, St. Jude Children's Research Hospital, Memphis, TN, USA

[3] Department of Neurology, Icahn School of Medicine at Mount Sinai, New York, NY, USA

‡*Correspondence:* andre_nussenzweig@nih.gov

SUMMARY

The Shieldin complex, consisting of SHLD1, SHLD2, SHLD3 and REV7, shields double strand DNA breaks (DSBs) from nucleolytic resection. The end-protecting activity of Shieldin promotes productive non-homologous end joining (NHEJ) in G1 but can threaten genome integrity during S-phase by blocking homologous recombination (HR). Curiously, the penultimate Shieldin component, SHLD1 is one of the least abundant mammalian proteins. Here, we report that the transcription factors THAP1, YY1 and HCF1 bind directly to the *SHLD1* promoter, where they cooperatively maintain the low basal expression of *SHLD1*. Functionally, this transcriptional network ensures that SHLD1 protein levels are kept in check to enable a proper balance between end protection and end resection during physiological DSB repair. In the context of BRCA1 deficiency, loss of THAP1 dependent SHLD1 expression confers cross resistance to PARP inhibitor and cisplatin, and shorter progression free survival in ovarian cancer patients. In contrast, loss of THAP1 in BRCA2 deficient cells increases genome instability and correlates with improved responses to chemotherapy. Pathogenic THAP1 mutations are causatively linked to adult-onset torsion dystonia type 6 (DYT6) movement disorder, but the critical disease targets are unknown. We further demonstrate that murine models of *Thap1*-associated dystonia show reduced *Shld1* expression concomitant with elevated levels of unresolved DNA damage in the brain. In summary, our study provides the first example of a transcriptional network that directly controls DSB repair choice and reveals a previously unsuspected link between DNA damage and dystonia.

INTRODUCTION

Direct joining of double strand breaks (DSBs) by non-homologous end joining (NHEJ) provides eukaryotes with a fast and versatile mechanism for repairing genomic damage (Pannunzio et al., 2018). Although not as precise as homologous recombination (HR), NHEJ is essential for genome maintenance in non-dividing cells (such as post-mitotic neurons), in which sister chromatid templates for high fidelity HR are unavailable. In addition, NHEJ supports physiological DNA rearrangement processes where genetic diversity is desirable, such as V(D)J recombination and immunogolobulin class switching (CSR) in developing lymphocytes (Rooney et al., 2004). Accordingly, inactivation of NHEJ factors have been linked to neurodegeneration and immunodeficiency in model organisms as well as in human patients (McKinnon, 2009, 2017; Slatter and Gennery, 2010).

Once cells enter S-phase to commence DNA replication, HR gradually becomes an option as sister chromatids become available (Karanam et al., 2012). Extensive 5'-nucleolytic resection of DSBs is required to license HR, and is thought to simultaneously preclude NHEJ. The 3'-single-stranded (ss) overhang generated by resection is initially stabilized by RPA. The PALB2/BRCA2 mediator complex subsequently displaces RPA from the ssDNA and assembles polymers of the recombinase RAD51, which invade the intact sister chromatid to promote templated DNA synthesis (Zhao et al., 2019). HR is uniquely suitable for healing replication-associated DSBs, as most of them are one-ended. By contrast, NHEJ is either unable to repair such lesions or would give rise to translocations when disparate one-ended DSBs are inappropriately joined together. Consistently, deficiencies in HR not only facilitate tumorigenesis (Tarsounas and Sung,

74 2020), but also confer hypersensitivity to chemotherapeutic agents such as PARP
75 inhibitors (PARPi) and cisplatin (Byrum et al., 2019).
76 53BP1 has emerged as a master coordinator of anti-HR activities in mammalian
77 cells (Mirman and de Lange, 2020; Setiaputra and Durocher, 2019). 53BP1 recruits a
78 number of effector molecules, including PTIP (Callen et al., 2013), the Shieldin complex
79 (consisting of SHLD1, SHLD2, SHLD3 and REV7) (Boersma et al., 2015; Dev et al.,
80 2018; Findlay et al., 2018; Gao et al., 2018; Ghezraoui et al., 2018; Gupta et al., 2018;
81 Mirman et al., 2018; Noordermeer et al., 2018; Tomida et al., 2018; Xu et al., 2015), the
82 CST complex (Barazas et al., 2018; Mirman et al., 2018) and DYNLL1 (He et al., 2018),
83 to block DSB end resection by multiple nucleases. In addition, the 53BP1-RIF1-Shieldin
84 subpathway is capable of directly impeding PALB2/BRCA2-mediated RAD51 loading on
85 ssDNA (Callen et al., 2020). These barriers are normally overcome by the tumor
86 suppressor BRCA1, which counteracts 53BP1 during S-phase by stimulating both
87 nucleolytic resection and RAD51 assembly through still poorly defined mechanisms
88 (Scully et al., 2019; Tarsounas and Sung, 2020). In addition to BRCA1, the AAA+ family
89 ATPase TRIP13 promotes HR by catalyzing an inactivating conformational change in the
90 REV7 component of the Shieldin complex (Clairmont et al., 2020).
91 Consistent with the antagonism between BRCA1 and 53BP1, much of the HR
92 defects observed in BRCA1-deficient cells can be attributed to excessive 53BP1-mediated
93 end-protection, which in turn promotes the toxic joining of replication-associated DSBs
94 (Bouwman et al., 2010; Bunting et al., 2010; Callen et al., 2020; Jaspers et al., 2013).
95 Notably, depletion of any Shieldin complex component restores HR in BRCA1-deficient
96 cells (Dev et al., 2018; Gao et al., 2018; Ghezraoui et al., 2018; Gupta et al., 2018;

Mirman et al., 2018; Noordermeer et al., 2018; Tomida et al., 2018), and reduced expression of Shieldin is associated with the acquisition of PARPi resistance in PDX models of BRCA1-deficient breast cancer (Dev et al., 2018).

These observations suggest that the level of Shieldin activity is a major determinant of HR competency and chemosensitivity in mammalian cells. Remarkably, however, Shieldin components are undetectable by mass spectrometry based proteome analysis (Gupta et al., 2018) and are amongst the lowest expressed genes (Ochs et al., 2019). Here, using a combination of CRISPR screens and biochemical assays, we addressed how changes in the levels of Shieldin affect DSB repair and genome stability. Our work reveals that the basal expression of SHLD1 is maintained by direct binding of a network of transcription factors consisting of THAP1, HCF1 and YY1 to the *SHLD1* promoter. Mutations in *THAP1* cause an inherited form of dystonia (DYT6), a neurodegenerative disorder that is characterized by sustained involuntary muscle contractions (Bressman et al., 2009; Djarmati et al., 2009; Fuchs et al., 2009), although critical disease target(s) of THAP1 are unknown. We find that inactivation of THAP1 abolishes *SHLD1* expression, leading to HR restoration and chemoresistance in BRCA1-deficient cell lines and BRCA1-mutated human ovarian cancers. Conversely, overexpression of either SHLD1 or THAP1 compromises the repair of replication-associated DSBs in BRCA1-proficient cells. Moreover, THAP1 is essential for CSR and helps maintain genome integrity in the developing brain. Together, these results help explain why Shieldin expression must be kept at low levels, and reveal a hitherto unknown physiological function of THAP1 in maintaining genome stability in all cell types.

RESULTS

Identification of THAP1 as a modifier of chemosensitivity in mouse models of BRCA1 deficiency

The Shieldin complex is recruited to DSBs by 53BP1 and RIF1 where it blocks HR by inhibiting nucleolytic end resection (Mirman and de Lange, 2020; Setiaputra and Durocher, 2019) as well as suppressing the loading of RAD51 onto ssDNA post-resection (Belotserkovskaya et al., 2020; Callen et al., 2020). These barriers to HR are normally overcome by BRCA1, but become insurmountable in BRCA1-deficient cells and cause PARP inhibitor hypersensitivity. Here, we undertook whole-genome CRISPR-Cas9 screens to search for genes whose mutation confer PARPi resistance in BRCA1-deficient cells. Human BRCA1-deficient tumors, particularly those that harbor pathogenic exon 11 mutations, often express a hypomorphic BRCA1 protein (BRCA1-Δ11q) (Wang et al., 2016). In order to model acquired PARPi resistance in a clinically relevant manner, we elected to use the PARPi-hypersensitive *Brca1*$^{\Delta 11}$ and *Brca1*$^{\Delta 11}$*Trp53bp1*^{S25A} murine models, in which a truncated BRCA1 protein (BRCA1-Δ11) highly similar to human BRCA1-Δ11q is expressed (Bunting et al., 2010; Callen et al., 2020; Wang et al., 2016) (Figure 1A). Moreover, cells derived from *Brca1*$^{\Delta 11}$ and *Brca1*$^{\Delta 11}$*Trp53bp1*^{S25A} mice exhibit marked HR defects due to excessive Shieldin-mediated end protection pre- and post-resection, respectively (Callen et al., 2020), thereby representing murine model systems to search for putative regulators of Shieldin function.

Brca1$^{\Delta 11}$ and *Brca1*$^{\Delta 11}$*Trp53bp1*^{S25A} mouse embryonic fibroblasts (MEFs) expressing CAS9 were transduced with lentiviral libraries of single-guide RNAs (sgRNAs), and the resultant pools of edited cells were exposed to near-lethal doses of the

clinically used PARPi olaparib for two weeks (Figure 1A). The screens were performed twice in *Brca1*$^{\Delta 11}$ MEFs and once in *Brca1*$^{\Delta 11}$*Trp53bp1*^{S25A} MEFs. Gene enrichment or depletion was calculated for each cell line using the MAGeCK algorithm (Li et al., 2014), with positive and negative β-scores indicating positive and negative selection of a given gene deletion event, respectively (Figure 1B and Table S1). To obtain high-confidence hits, we combined the individual gene-based β-scores for both cell lines. Importantly, in both cell lines we identified multiple genes whose deletion have previously been shown to confer PARPi resistance in human BRCA1-deficient cells, including *Trp53bp1, Shld1*, *Shld2*, *Ctc1*, *Stn1* and *Dynll1* (Figure 1B and Table S1) (Dev et al., 2018; Findlay et al., 2018; Gao et al., 2018; Ghezraoui et al., 2018; Gupta et al., 2018; He et al., 2018; Mirman et al., 2018; Noordermeer et al., 2018; Tomida et al., 2018), thereby confirming the validity of our approach.

In addition, we identified *Thap1* as a candidate gene whose deletion strongly imparts PARPi resistance in both *Brca1*$^{\Delta 11}$ and *Brca1*$^{\Delta 11}$*Trp53bp1*^{S25A} MEFs (Figure 1B). *Thap1* encodes the murine homolog of Thanatos-associated protein 1 (THAP1), a transcription factor that is causatively linked to adult-onset torsion dystonia type 6 (DYT6) in humans (Bressman et al., 2009; Djarmati et al., 2009; Fuchs et al., 2009). To verify the results of the primary screens, we conducted competitive growth assays using sgRNAs targeting either *Thap1* or *Trp53bp1* and found that each led to the outgrowth of two independently-derived *Brca1*$^{\Delta 11}$ MEF cell lines (14223 and S12) treated with PARPi (Figure 1C and Figure S1A), while no difference in growth was observed without the drug (Figure S1B). In addition, we generated two independent clonal knockouts of *Thap1* in one of the two *Brca1*$^{\Delta 11}$ MEF cell lines (14223) (Figure S1C) and confirmed that deletion

of *Thap1* led to cross-resistance to PARPi and cisplatin, as measured by cell viability assays and mitotic chromosome aberrations (Figures 1D and 1E). Finally, deletion of *THAP1* in two-independently derived clones of *BRCA1*-null human RPE1 cells (Noordermeer et al., 2018) resulted in marked PARPi resistance (Figure 1F). In contrast to BRCA1 deficiency, *Thap1* deletion did not affect the PARPi sensitivity of BRCA1-proficient cells as measured by cell viability (Figures S1D, S1E). These results support THAP1 as a *bona fide* modifier of chemosensitivity in mouse and human cell lines with BRCA1 deficiency.

Loss of THAP1 confers chemoresistance in BRCA1-deficient human tumors

Analysis of publicly available gene expression databases revealed that *THAP1* is ubiquitously expressed in malignant tissues, with notable increased expression in testis, ovarian and breast cancers (Figure 2A). To test whether dysregulated *THAP1* has an impact on tumorigenesis, we extracted human tumor expression data from The Cancer Genome Atlas (TCGA) database (Cancer Genome Atlas Research, 2011). We found that low expression of *THAP1* was significantly correlated with shorter progression-free survival (PFS) of patients with *BRCA1*-mutated serous ovarian carcinoma (Figure 2B). Although most patients in this cohort have undergone platinum-based chemotherapy (Cancer Genome Atlas Research, 2011), no correlation with PFS was observed for patients with *BRCA1*-proficient tumors (Figure S2A). Together, our results suggest that *THAP1* loss alleviates PARPi and cisplatin hypersensitivity in BRCA1-deficient cells and tumors, and this effect is not species-, cell lineage- or *BRCA1* mutation-specific.

A recent study of prognostic biomarkers of acute myelogenous leukemia reported that low levels of *BRCA2* and *THAP1* correlated with improved patient survival (Wang et al., 2020). Since BRCA1 and BRCA2 have functionally non-redundant roles during HR, we also sought to determine how *THAP1* expression influences the response of BRCA2-deficient cells to chemotherapy. In contrast to its impact on BRCA1-deficiency (Figures 1C-F and 2B), we found that deletion of *Thap1* increased genome instability in *Brca2*^{Y3308X} mutant mouse embryonic stem cells (mESCs) treated with PARPi (Figure 2C). Consistent with increased therapeutic vulnerability, clinical data demonstrated that low *THAP1* expression was correlated with longer PFS in patients with *BRCA2*-mutant serous ovarian carcinoma (Figure 2D). In summary, our results indicate that expression of THAP1 modulates the responses to chemotherapy in opposite ways in BRCA1- and BRCA2-deficient cells (see discussion).

THAP1 is a direct transcriptional regulator of *Shld1* expression

Mutations in *THAP1* have been identified as the cause of DYT6, which is associated with abnormal muscle contractions without overt neuropathological lesions (Bressman et al., 2009; Djarmati et al., 2009; Fuchs et al., 2009). THAP1 has been proposed to regulate numerous targets in neurons, but the causative links to dystonic movements remains unclear. While several transcription factors promote DNA repair by acting directly at sites of DNA damage in a transcription-independent manner (Malewicz and Perlmann, 2014), we failed to observe THAP1 translocation to sites of DNA damage induced by γ-irradiation (not shown). Given that most of the dystonia-causing mutations in THAP1 disrupt its sequence-specific DNA binding activity (Blanchard et al., 2011; Charlesworth

et al., 2013), we therefore hypothesized that THAP1 may regulate the chemosensitivity of BRCA-deficient cells directly through transcriptional mechanisms.

To investigate this possibility, we performed nascent RNA-sequencing (RNA-seq) to systematically probe for genes that are differentially expressed in *Thap1*$^{-/-}$ and *Thap1*$^{-/-}$ *Brca1*$^{\Delta 11}$ MEFs, as compared to wild-type and *Brca1*$^{\Delta 11}$ parental cells, respectively. In total, we found 452 differentially expressed genes (221 downregulated and 231 upregulated (log2 fold-change >2 and FDR <0.05) between *Thap1*$^{-/-}$ and wild-type cells, while 1,337 genes were differentially expressed (675 downregulated and 662 upregulated) between *Thap1*$^{-/-}$*Brca1*$^{\Delta 11}$ and *Brca1*$^{\Delta 11}$ cells (Figure 3A). Of these, 98 genes (57 downregulated and 41 upregulated) were common to both *Thap1*$^{-/-}$ and *Thap1*$^{-/-}$*Brca1*$^{\Delta 11}$ cells (Table S2). To determine which of these genes are direct targets of THAP1, we performed THAP1 ChIP-sequencing (ChIP-seq) in MEFs which identified 2,137 unique THAP1 binding sites. Overall, only six differentially expressed genes in *Thap1*-deficient MEFs were also bound by THAP1 (Figure 3B). Among these genes we identified *Shld1*, the penultimate component of the Shieldin complex and a known effector of 53BP1. Re-examining our RNA-seq datasets, we confirmed that *Shld1* expression was significantly decreased in *Thap1*$^{-/-}$ and *Thap1*$^{-/-}$*BRCA1*$^{\Delta 11/\Delta 11}$ MEFs (Figure 3C). By contrast, other genes encoding 53BP1 pahway members exhibited only modest expression changes (Figures S3A and S3B).

Sequence-specific DNA binding of THAP1 is mediated by the THAP-type zinc finger located within its THAP domain and is disrupted by the C54Y point mutation, a known DYT6 causative mutation (Roussigne et al., 2003; Ruiz et al., 2015). To determine whether THAP1 DNA binding is required for *Shld1* expression, we re-analyzed a RNA-

234 seq dataset derived from *Thap1*$^{C54Y/C54Y}$ mESCs (Aguilo et al., 2017). We found that *Shld1*
235 expression was decreased in *Thap1*$^{C54Y/C54Y}$ cells at a level comparable to that found in
236 *THAP1*$^{-/-}$ mESCs (Figures 3D and S3C). Thus, sequence-specific DNA binding is required
237 for THAP1 regulation of SHLD1.
238
239 **A THAP1-HCF1-YY1 co-regulatory module promotes *Shld1* expression**
240 The *Shld1* promoter contains a consensus THAP1 binding sequence (THABS) motif
241 where a strong THAP1 ChIP-seq signal was detected (Figure 3E). Recent studies have
242 shown that THAP1 frequently co-occupy its target genes with the transcriptional co-
243 regulators HCF1 and YY1 (Mazars et al., 2010; Michaud et al., 2013; Yellajoshyula et al.,
244 2017), which are recruited to the promoters by THAP1 (Hollstein et al., 2017;
245 Yellajoshyula et al., 2017). Analyses of ChIP-seq datasets revealed that HCF1 and YY1
246 indeed co-occupy the *Shld1* promoter with THAP1 (Figure 3F). Moreover, we found that
247 inducible deletion of *Hcfc1* in hepatocytes (*Hcfc1*$^{HepKO/Y}$) led to downregulation of *Shld1*
248 expression (Figure S3D) (Minocha et al., 2019). Similarly, conditional knockout of *Yy1* in
249 mouse B cells (*Yy1*$^{flox/flox}$ TAT-Cre) significantly decreased *Shld1* expression (Figure S3E)
250 (Nandi et al., 2020). Strikingly, when we overlapped genes whose expression were
251 differentially regulated by HCF1, YY1 and THAP1 and that were also direct targets of
252 these transcription factors, only two genes *Shld1* and *Gm11520* (a putative noncoding
253 RNA of unknown function) showed significant enrichment (Figures 3A, 3G, S3F and
254 S3G). Thus, although the vast majority of genes appear to be independently regulated by
255 THAP1, HCF1 and YY1, these cofactors cooperatively bind to the *Shld1* promoter to
256 maintain its expression.

Given that Shld1 has been shown to be amongst the lowest expressed genes (Gupta et al., 2018; Ochs et al., 2016), we next investigated whether elevating SHLD1 protein levels cause cellular toxicity. To this end, we overexpressed exogenous SHLD1 or THAP1 in BRCA1-proficient MEFs (Figure S3H). We found that overexpression of SHLD1 and to a lesser extent THAP1 significantly potentiated PARPi-induced genomic instability (Figure 3H). These results are consistent with the notion that excessive expression of SHLD1 can overtly suppress HR, which explains why its expression must be fine-tuned by THAP1-mediated transcriptional regulation.

THAP1 inhibits DNA end resection and RAD51 nucleofilament formation

The precise function of SHLD1 within the 53BP1 pathway is unknown. Previous studies have implicated SHLD3 as the anchor that attaches the entire Shieldin complex to RIF1, enabling its recruitment to 53BP1-decorated chromatin (Dai et al., 2020; Noordermeer et al., 2018; Setiaputra and Durocher, 2019). We therefore tested whether THAP1 loss would impact the recruitment of 53BP1, RIF1 and SHLD3 to sites of DNA damage. We found that foci formation for 53BP1, RIF1, and SHLD3 was not impaired in *Thap1*$^{-/-}$*Brca1*$^{\Delta 11}$ cells but rather, was somewhat enhanced (Figures 4A, 4B, S4A and S4B). In BRCA1-deficient cells, unmitigated activation of the 53BP1-RIF1-Shieldin pathway impairs HR by blocking DSB end-resection and RAD51 nucleofilament formation (Dev et al., 2018; Findlay et al., 2018; Gao et al., 2018; Ghezraoui et al., 2018; Gupta et al., 2018; He et al., 2018; Mirman et al., 2018; Noordermeer et al., 2018; Tomida et al., 2018). Indeed, we found that deletion of THAP1 rescued both irradiation-induced RPA foci and RAD51 nucleofilament formation in *Brca1*$^{\Delta 11}$ MEFs to wild-type levels (Figures 4C and 4D).

Thus, loss of THAP1 promotes end-resection and HR proficiency in BRCA1-deficient cells, which may explain why doubly deficient cells acquire resistance to PARPi and cisplatin.

To provide definitive proof that the observed PARPi resistance in *Thap1*$^{-/-}$*Brca1*$^{\Delta 11}$ MEFs was due to decreased *Shld1* expression, we overexpressed SHLD1 in *Thap1*$^{-/-}$*Brca1*$^{\Delta 11}$ MEFs (Figure S4C). Overexpression of SHLD1 in *Thap1*$^{-/-}$*Brca1*$^{\Delta 11}$ MEFs restored their hypersensitivity to PARPi (Figure 4E). As expected, add-back of THAP1 in *Thap1*$^{-/-}$*Brca1*$^{\Delta 11}$ MEFs also restored PARPi hypersensitivity (Figure 4E). Moreover, SHLD1 or THAP1 overexpression in *Thap1*$^{-/-}$*Brca1*$^{\Delta 11}$ MEFs increased chromosomal aberrations to levels comparable to or higher than in *Brca1*$^{\Delta 11}$ parental cells (Figure 4F). In contrast to SHLD1, ectopic SHLD3 expression did not promote PARPi hypersensitivity in *Thap1*$^{-/-}$*Brca1*$^{\Delta 11}$ cells, despite normal focal accumulation (Figures S4A, S4B, and S4D). Thus, THAP1-dependent *Shld1* transcription is essential for maintaining the functionality of Shieldin, but is not required for localizing the complex to DNA damage sites.

THAP1 promotes Immunglobulin Class-Switch Recombination

Uncontrolled activation of the 53BP1-RIF1-Shieldin pathway compromises genome integrity during S/G2, but its function is uniquely required for NHEJ during class switch recombination (CSR), a G1-restricted genome rearrangement process that changes antibody effector functions (Findlay et al., 2018; Ghezraoui et al., 2018; Gupta et al., 2018; Manis et al., 2004; Noordermeer et al., 2018; Petersen et al., 2001; Ward et al., 2004). Notably, we found that loss of THAP1 severely compromised CSR in cytokine-stimulated CH12F3-2 mouse B cells, similar to inactivation of 53BP1, SHLD1 or SHLD3

(Figure 5A). These results revealed a hitherto unknown physiological function THAP1 outside of the developing nervous system in which it maintains *Shld1* expression to promote NHEJ.

Although Shieldin genes are epistatic with 53BP1 and RIF1 with respect to CSR, they are phylogenetically younger than the upstream components of the DNA end-protection pathway (Gupta et al., 2018). Indeed, it has been speculated that the evolution of Shieldin may have contributed to the emergence of a primitive form of CSR, with SHLD1 being the most recently evolved regulator (Gupta et al., 2018). Interestingly, based on homology data across diverse eukaryotes, we found that THAP1 evolved recently with components of the Shieldin complex (Figure 5B). It is therefore tempting to speculate that THAP1 might have co-evolved with Shieldin to regulate developmentally programmed NHEJ during CSR.

THAP1 deficiency leads to unresolved DNA damage in the developing brain

Given previously established links between NHEJ deficiency and various neurodevelopmental abnormalities (McKinnon, 2017), we next investigated whether a perturbed DNA damage response could partially underlie the pathogenesis of DYT6. While germline deletion of murine *Thap1* is embryonic lethal (Ruiz et al., 2015), conditional deletion of *Thap1* in the central nervous system has been achieved using Nestin-Cre (*Thap1*$^{flox/flox}$ *Nestin-Cre*) (Hollstein et al., 2017)). Re-analysis of RNA-seq dataset of *Thap1*$^{flox/flox}$ *Nestin-Cre* mice showed that *Shld1* expression was significantly decreased in both the cerebellum and striatal tissues (Figures 6A and S5A), suggesting that loss of THAP1 may result in defective DNA repair in the brain.

To explore this possibility, we generated an independently derived *Thap1*$^{flox/flox}$ *Nestin-Cre* murine model, which exhibited motor abnormalities, including abnormal limb clasping during tail suspection similar to previous reports (Hollstein et al., 2017; Melis et al., 2020). Histological examinations of various brain regions of 6-week old mice revealed significant increased immune-positivity of γH2ax and serine-824 phosphorylated Kap1 (p-Kap1) in THAP1-deficient cerebellum (Figure 6B). For example, sporadic γH2ax immunoreactivity was observed throughout the inner granule layer of the of 6-weeks-old *Thap1*$^{Nes\text{-}cre}$ cerebellum, while in control tissue no DNA damage was found (Figure 6B). Within the *Thap1*$^{flox/flox}$ *Nestin-Cre* molecular layer, interneurons (Figure 6B) also showed clear evidence of γH2ax as did calbindin positive Purkinje cells (PCs) (Figure 6C). We also used p-Kap1, indicative of ATM activation, to confirm DNA damage accumulation (Figure 6D). p-Kap1 foci showed a distribution similar to γH2ax with distinct, multiple individual nuclear foci present in *Thap1*$^{flox/flox}$ *Nestin-Cre* neurons but not in controls, as shown for Calbindin positive PCs (Figure 6D). Elsewhere in the *Thap1*$^{flox/flox}$ *Nestin-Cre* brain, DNA damage accumulation was apparent in the cortex and hippocampal region as shown with γH2ax immunohistochemistry, while no sign of DNA damage was found in comparative regions of control tissue (Figures S5B and S5C). Thus, proper THAP1 function ensures neural genome stability in the brain.

DISCUSSION

The recently discovered Shieldin complex is a key enforcer of 53BP1-mediated DSB end protection, which blocks HR and confers chemotherapy hypersensitivity in BRCA1-deficient cells (Mirman and de Lange, 2020; Setiaputra and Durocher, 2019). Yet, little is

known about how Shieldin is functionally regulated *in vivo*. Here, we reveal that the evolutionarily conserved transcription factor THAP1 promotes transcription of the *Shld1* gene in mammalian cells. Our observation that loss of THAP1 restores HR and chemoresistance in both *Brca1*$^{\Delta 11}$ (resection-deficient) and *Brca1*$^{\Delta 11}$*Trp53bp1*^{S25A} (RAD51 loading-deficient) MEFs is consistent with a critical function of Shieldin pre- and post-resection (Belotserkovskaya et al., 2020; Callen et al., 2020). Moreover, the THAP1-SHLD1 axis supports productive CSR in murine B lymphocytes and maintains genome stability in the developing mouse brain. THAP1 acts in concert with co-regulators HCF1 and YY1, previously unknown to function in 53BP1-end protection pathways. Altogether, these observations highlight the need for cells to fine-tune Shieldin activity depending on the desired DSB repair outcome.

Shieldin complex components, particularly SHLD1 and SHLD3, are expressed at extremely low abundance compared to other DNA repair factors (Gupta et al., 2018; Ochs et al., 2019). It is notable that a ubiquitious transcription factor module (THAP1-YY1-HCF1) would be specifically required to promote ultra low levels of *Shld1* expression. While additional studies are needed to determine whether the *Shld1* locus receives additional regulatory inputs to ensure its low steady-state transcriptional output, our data suggest that relatively low levels of 53BP1-RIF1-Shieldin activity may suffice for physiological NHEJ and that hyper-activation of this pathway could disproportionately inhibit HR in replicating cells. Consistent with this idea, we have demonstrated earlier that the efficiency of CSR was unaffected by increased 53BP1 expression or 53BP1 heterozygosity, whereas 53BP1 overexpression clearly exacerbated genomic instability induced by PARPi treatment not only in the BRCA1-deficient setting but also in BRCA1-

proficient cells (Zong et al., 2015). Intriguingly, organisms with no clear orthologs for 53BP1, RIF1 or Shieldin also generally lack a recognizable orthologous *Brca1* gene (Figure 5B). Given that excessive end protection can potentially compromise genome stability, we propose that orthologs of BRCA1 might have evolved specifically to counteract 53BP1-RIF1-Shieldin activity, thereby balancing the requirement for inherently error-prone NHEJ against the overarching need for maximizing genome stability through high-fidelity DSB repair.

Surprisingly, unlike BRCA1-deficient MEFs and human cells, in which loss of THAP1 largely rescued PARPi-induced genomic instability, we observed the opposite effect in BRCA2-mutant mESCs. Accordingly, clinical data suggested that low expression of THAP1 correlated with better outcome in *BRCA2*-mutated patients with serious ovarian cancer (Figure 2D). Consistently, we found that low SHLD1 expression also correlates with improved survival in BRCA2 deficient ovarian cancer (Figure S2B). We note that a similar observation has previously been made in mouse cells lacking PALB2, in which the existing vulnerability to PARPi was further exacerbated upon loss of 53BP1 (Bowman-Colin et al., 2013). Therefore, HR restoration in cells with an inhibited 53BP1-RIF1-Shieldin pathway seems to be contingent on RAD51 nucleofilament assembly. The role of BRCA1 in loading RAD51 can be partly compensated for by RNF168 (Zong et al., 2019), whereas PALB2 and BRCA2 are absolutely required (Belotserkovskaya et al., 2020; Callen et al., 2020). Productive assembly of the recombinase has been shown to limit the extent of end resection (Huang et al., 2018; Paiano et al., 2020). Therefore, it is possible that deletion of THAP1/SHLD1 or 53BP1 would lead to excessive single-strand DNA in BRCA2-deficient cells, which in turn could increase genomic instability by engaging

highly mutagenic RAD51-independent single strand annealing pathways (Ochs et al., 2016).

Clinically, loss-of-function mutations in THAP1 cause the autosomal dominant dystonia disorder DYT6 (Bressman et al., 2009; Djarmati et al., 2009; Fuchs et al., 2009). Employing a mouse model of DYT6, we demonstrated for the first time that THAP1 deficiency leads to unresolved DNA breaks in the developing mouse brain, particularly in the cerebellum. These DNA DSBs could arise from unrepaired oxidative DNA lesions that occur at preferentially at the promoter regions of neuronal genes during normal cellular aging (Cassagnes et al., 2018; Madabhushi et al., 2015)). NHEJ, which is promoted by Shieldin-mediated end protection, is likely to be the dominant DSB repair pathway in post-mitotic neurons. Therefore, it is plausible that impairment of THAP1-dependent *Shld1* expression could result in the accumulation of neuronal DSBs. In that context it is notable that ATM mutations, which abrogate 53BP1-Shieldin activity frequently present as generalized dystonia (Meissner et al., 2013; Saunders-Pullman et al., 2012). While additional studies are needed to test the role of SHLD1 in neurodevelopmental DNA repair, we speculate that unresolved DNA damage in the nervous system could be an important contributing factor to dystonia.

Acknowledgements

We thank Sergio Ruiz Macias, Eros Lazzerini Denchi, and Nussenzweig lab members for stimulating discussions; Daniel Durocher for the *TP53*$^{-/-}$*BRCA1*$^{-/-}$ human RPE1 cell line; Panagiotis A. Konstantinopoulos for assistance with TCGA analysis; K.S. was supported by a fellowship from the Uehara Memorial Foundation; Sequencing were supported from the NCI CCR Genomics Core; Flow cytometry analyses and sorting were performed at the NCI CCR LGI Flow Cytometry Core; The A.N. laboratory is supported by the Intramural Research Program of the NIH, an Ellison Medical Foundation Senior Scholar in Aging Award (AG-SS- 2633-11), the Department of Defense Idea Expansion (W81XWH-15-2-006) and Breakthrough (W81XWH-16-1-599) Awards, the Alex Lemonade Stand Foundation Award, and an NIH Intramural FLEX Award.

Author Contributions

K.S., D.Z., E.C., and A.N. conceived and planned the study. K.S., D.Z., E.C., L.C.D., N.W., M.I., A.S. designed and performed experiments; K.S., D.Z., E.C., W.W. and F.B. analyzed the data; M.E. and P.J.M. supervised and provided advice; A.N. supervised the study. K.S., D.Z., and A.N. wrote the manuscript with comments from the authors.

Declaration of Interests

The authors declare no competing interests.

Figure Legends:

Figure 1. Identification of THAP1 as a modifier of chemosensitivity in BRCA1-deficient cells.

(**A**) Schematic of CRISPR-based screen for genes whose deletion confer resistance to the clinical PARP inhibitor (PARPi) olaparib in *BRCA1*-mutant MEFs (*Brca1*$^{\Delta 11}$ and *Brca1*$^{\Delta 11}$*Trp53bp1*^{S25A}).

(**B**) Analysis of guide RNA enrichment in PARPi survivors by MAGeCK. A false discovery rate (FDR) of 0.1 is indicated by dotted line. The combined top hits for *Brca1*$^{\Delta 11}$ (n=2) and *Brca1*$^{\Delta 11}$*Trp53bp1*^{S25A} (n=1) screens are tabulated on the right. summary of n = 3 biological replicates. Previously identified resistance genes are shown in blue and *Thap1* is shown in red.

(**C**) Multicolor Competition Assay (MCA): *Cas9*$^{+}$*Brca1*$^{\Delta 11}$ MEFs (14223 and S12) transduced with a specific guide RNA targeting *Thap1*, *Trp53bp1* or an empty vector (sgEmpty, all GFP-positive) were co-incubated (1:1 ratio) with *Cas9*$^{+}$*Brca1*$^{\Delta 11}$ MEFs transduced with non-targeting guides (sgLacZ, mCherry-positive). Data represent mean fraction of GFP-positive cells ± s.d., normalized to day 0 (n = 3). Deletion of *Thap1* significantly enhanced the outgrowth of *Brca1*$^{\Delta 11}$ cells following PARPi (100 nM) treatment ($p<0.0001$).

(**D**) Viability of *WT*, *Brca1*$^{\Delta 11}$ (14223) and two individual clones (#1 and #2) of *Thap1*$^{-/-}$ *Brca1*$^{\Delta 11}$ MEFs (derived from 14223), as measured by CellTiter-Glo seven days after treatment with PARPi (left) and cisplatin (right).

(**E**) Genomic instability detected in the metaphase spreads of *WT*, *Brca1*$^{\Delta 11}$ and two individual clones (#1 and #2) of *Thap1*$^{-/-}$*Brca1*$^{\Delta 11}$ MEFs. Cells were treated for 16 h with 1

μM PARPi or with 0.5 μM cisplatin. At least 50 metaphase spreads were scored per genotype and condition. The experiments were repeated four (PARPi) and three (cisplatin) times, respectively. Statistical significance was determined by Welch's t-test.

(**F**) Viability of *TP53*$^{-/-}$, *TP53*$^{-/-}$*BRCA1*$^{-/-}$ and two individual clones (#1 and #2) of *THAP1*$^{-/-}$*TP53*$^{-/-}$*BRCA1*$^{-/-}$ human RPE1 cells, as measured by CellTiter-Glo seven days after PARPi treatment. See also Figure S1.

Figure 2. Loss of THAP1 expression affects the chemosensitivity of BRCA-deficient human tumors.

(**A**) Expression of the *THAP1* gene (FPKM) in 17 different cancer types. Data were obtained from The Cancer Genome Atlas (TCGA) project.

(**B**) Progression-free survival (PFS) of BRCA1-mutated ovarian serous adenocarcinoma patients with standard platinum-based regimens (Cancer Genome Atlas Research, 2011). Patients were defined as having *THAP1* low- or high-expressing tumors on the basis of the quintile of *THAP1* expression (z-scores < -0.67). PFS curves were generated by the Kaplan–Meier method. The difference between the PFS of *THAP1* low- versus *THAP1* high-expressing patients was assessed by the log-rank test ($P < 0.01$).

(**C**) Genomic instability detected in the metaphase spreads of *Brca2*^{Y3308X} and *Thap1*$^{-/-}$*Brca2*^{Y3308X} mouse embryonic ES cells (mESCs) after 16 h of PARPi treatment (0.5 μM). At least 50 cells were scored per genotype and condition. The experiment was repeated three times. Statistical significance was determined by Welch's t-test.

(**D**) PFS of BRCA2-mutated ovarian serous adenocarcinoma patients with standard platinum-based regimens (Cancer Genome Atlas Research, 2011). Patients were defined

as having *THAP1* low- or high-expressing tumors on the basis of the same *THAP1* expression (z-scores < -0.67) used in Figure 2B. PFS curves were generated by the Kaplan–Meier method. The difference between the PFS of *THAP1* low- versus *THAP1* high-expressing patients was assessed by the log-rank test ($P < 0.05$). See also Figure S2.

Figure 3. THAP1 co-regulates *Shld1* expression with HCF1 and YY1.

(**A**) Top: Venn diagram depicting differentially expressed genes (nascent RNA-seq, log2 fold-change >2 and FDR <0.05) in *Thap1*$^{-/-}$ versus *WT* and *Thap1*$^{-/-}$*Brca1*$^{\Delta 11}$ versus *Brca1*$^{\Delta 11}$ MEFs in relation to THAP1-bound genes (ChIP-seq). The number of genes that were shown to be bound by THAP1 and were either downregulated or upregulated in THAP1-deficient MEFs are shown in blue and red, respectively.
Bottom: Venn diagram showing the number of differentially expressed THAP1-bound genes that were common to both *Thap1*$^{-/-}$ and *Thap1*$^{-/-}$*Brca1*$^{\Delta 11}$ MEFs.
(**B**) List of differentially expressed THAP1-bound genes that were common to both *Thap1*$^{-/-}$ and *Thap1*$^{-/-}$*Brca1*$^{\Delta 11}$ MEFs.
(**C**) Levels of *Shld1* gene expression in *WT, Thap1*$^{-/-}$*, Brca1*$^{\Delta 11}$ and *Thap1*$^{-/-}$*Brca1*$^{\Delta 11}$ MEFs, as detected by nascent RNA-seq. Data represent mean counts per million reads (CPM) ± s.d. (n = 3, **p<0.01, ***p<0.001). Statistical significance were determined by one-way ANOVA.
(**D**) Levels of *Shld1* gene expression in *WT, Thap*$^{C54Y/C54Y}$ *and Thap1*$^{-/-}$ mESCs. Data are from a publicly available RNA-seq dataset (GSE86911). Data represent mean CPM.
(**E**) The location of the THAP1 binding sequence (THABS) motif within the *Shld1* promoter. THAP1 ChIP-seq signal at the Shld1 promoter region is shown above.

(**F**) Binding of THAP1, HCF1 and YY1 at the *Shld1* gene locus. Data are from publicly available ChIP-seq datasets in ES cells (THAP1, GSE86911; HCF1, GSE36030; YY1, GSE68195).

(**G**) Venn diagram depicting differentially expressed genes (log2 fold-change >2 and FDR <0.05) in *Thap1*$^{-/-}$ MEFs, *Yy1*$^{flox/flox}$ TAT-Cre, and *Hcfc1*$^{HepKO/Y}$ cells in relation to THAP1-bound genes and previously published YY1- and HCF1-bound genes. Only two genes bound by all three transcription factors also exhibited differential expression: *Shld1* (downregulated) and *Gm11520* (upregulated).

(**H**) Genomic instability detected in the metaphase spreads of *WT* MEFs overexpressing SHLD1 or THAP1 with or without 16 h of PARPi treatment (1 μM). Ectopic SHLD1 and THAP1 expression was induced by doxycycline (Dox) treatment. At least 50 cells were scored per genotype and condition. The experiments were repeated four times for PARPi treated condition and twice for untreated condition. Statistical significance was determined by Welch's t-test. See also Figure S3.

Figure 4. THAP1-dependent *Shld1* expression inhibits HR and drives PARPi hypersensitivity in BRCA1-deficient cells.

(**A-B**) Quantification of 53BP1 (**A**) and RIF1 (**B**) foci in individual nuclei of *WT, Brca1*$^{\Delta 11}$*, Trp53bp1*$^{-/-}$*Brca1*$^{\Delta 11}$ and two individual clones of *Thap1*$^{-/-}$*Brca1*$^{\Delta 11}$ MEFs. Cells were irradiated with 10 Gy and analyzed 1 h post-IR.

(**C-D**) Quantification of RPA2 (**C**) and RAD51 (**D**) foci in individual EdU-positive (S-phase) nuclei of *WT, Brca1*$^{\Delta 11}$*, Trp53bp1*$^{-/-}$*Brca1*$^{\Delta 11}$ and two individual clones of *Thap1*$^{-/-}$ *Brca1*$^{\Delta 11}$ MEFs. Cells were irradiated with 10 Gy and analyzed 4 h post-IR.

(**E**) Viability of *WT*, *Brca1*$^{\Delta 11}$, *Thap1*$^{-/-}$*Brca1*$^{\Delta 11}$ and *Thap1*$^{-/-}$*Brca1*$^{\Delta 11}$ MEFs complemented with either THAP1 (*Thap1*$^{-/-}$*Brca1*$^{\Delta 11}$+THAP1) or SHLD1 (*Thap1*$^{-/-}$ *Brca1*$^{\Delta 11}$+SHLD1) cDNA, as measured by CellTiter-Glo seven days after PARPi treatment.

(**F**) Genomic instability (chromosome breaks and radials) detected in the metaphase spreads of *WT*, *Brca1*$^{\Delta 11}$, *Thap1*$^{-/-}$*Brca1*$^{\Delta 11}$, *Thap1*$^{-/-}$*Brca1*$^{\Delta 11}$+THAP1, and *Thap1*$^{-/-}$ *Brca1*$^{\Delta 11}$+SHLD1 MEFs after 16 h of PARPi treatment (1 μM). At least 50 cells were scored per genotype and condition. The experiment was repeated five times. Statistical significance was determined by Welch's t-test. See also Figure S4.

Figure 5. THAP1 participates in physiological NHEJ.

(**A**) Representative flow cytometry plots of IgM-to-IgA class switch recombination (CSR) in *WT*, *Trp53bp1*$^{-/-}$, *Shld1*$^{-/-}$, *Shld3*$^{-/-}$ and two individual clones of *Thap1*$^{-/-}$ (#4 and #12) CH12-F3 cells 24 hours after cytokine stimulation (IL-4, CD40L and TGFβ). Unstimulated WT cells are shown as a negative control. Quantification of IgM-to-IgA CSR is shown on the right and represents mean ± s.d., n=3.

(**B**) Presence/absence table of selected genes in representative organisms, based on orthology data from the Ensembl database. Genes encoding THAP1, the Shieldin components and additional DSB repair factors are shown in relation to antibody CSR.

Figure 6. THAP1-deficiency leads to ongoing DNA damage signaling in the brain.

(**A**) Analysis of *Shld1* gene expression in the striatum and cerebellum of *WT* and *Thap*$^{flox/flox}$*Nestin-Cre* mice from a publicly available RNA-seq dataset (GSE123880). Data depict the mean value of counts per million reads (CPM) ± s.d.. (n = 3 for striatum and n =

4 for cerebellum, ***p<0.001). Statistical significance were determined by one-way ANOVA.

(**B**) THAP1 loss causes genome instability in the cerebellum of 6-week-old *Thap1*$^{Nes\text{-}cre}$ mice. Unresolved DNA damage indicated by γH2AX (pSer139) is readily detectable in the inner granule layer (IGL) and in interneurons (IN) of the molecular layer (ML). The IGL is marked by NeuN immunopositivity. Expanded inset images show multiple nuclear puncta of γH2AX.

(**C**) Purkinje cells (PC) marked by Calbindin immunopositivity show γH2AX nuclear puncta.

(**D**) Phosphorylated KAP1 (pSer824)), a marker of DSB-induced ATM activation, is readily detectable in the nuclei of both PCs and interneurons (IN). Expanded inset images show multiple nuclear puncta of DNA damage-induced pKAP1 foci. See also Figure S5.

Supplementary Figure Legends:

Figure S1. Generation and characterization of *Thap1* knockout cells. Related to Figure 1.

(**A**) Genomic sequences within exon 2 of the mouse *Thap1* gene. Two independent sgRNAs used to generate *Thap1* knockouts (sgRNA1;red and sgRNA2;blue) are shown along with their respective PAM sequences (green).

(**B**) Multicolor Competition Assay (MCA): *Cas9*$^{+}$*Brca1*$^{\Delta 11}$ MEFs (14223 and S12) transduced with a specific guide RNA targeting *Thap1*, *Trp53bp1* or an empty vector (sgEmpty, all GFP-positive) were co-incubated (1:1 ratio) with *Cas9*$^{+}$*Brca1*$^{\Delta 11}$ MEFs transduced with non-targeting guides (sgLacZ, mCherry-positive). Data represent mean

fraction of GFP-positive cells ± s.d., normalized to day 0 (n = 3). Vehicle treated control for Figure 1C are shown.

(**C**) Confirmation of successful editing of the *Thap1* locus in two independent *Thap1*$^{-/-}$ *Brca1*$^{\Delta 11}$ MEF clones (#1 and #2) by either sgRNA1 (red) or sgRNA2 (blue).

(**D**) Confirmation of successful editing of the *Thap1* locus in a *Thap1*$^{-/-}$ MEF clone by sgRNA1 (red).

(**E**) Viability of *WT*, *Thap1*$^{-/-}$ and *Brca1*$^{\Delta 11}$ MEFs, as measured by CellTiter-Glo seven days after PARPi treatment.

Figure S2. Prognostic implications of THAP1 expression in ovarian serious adenocarcinoma. Related to Figure 2.

(**A**) Progression-free survival (PFS) of *BRCA1*-WT ovarian serous adenocarcinoma patients on standard platinum-based regimens (Cancer Genome Atlas Research, 2011). Patients were defined as having *THAP1* low- or high-expressing tumors on the basis of the quintile of *THAP1* expression (z-scores < -0.67). PFS curves were generated by the Kaplan–Meier method. The difference between the PFS of *THAP1* low- versus *THAP1* high-expressing patients was assessed by the log-rank test ($P > 0.05$).

(**B**) PFS of *BRCA2*-mutated ovarian serous adenocarcinoma patients on standard platinum-based regimens (Cancer Genome Atlas Research, 2011). The median expression of *SHLD1* for all tumors in this cohort was used to separate patients into *SHLD1* low- or high-expressors. The difference between the PFS of *SHLD1* low- versus *SHLD1* high-expressing patients was assessed by the log-rank test ($P < 0.05$).

Figure S3. Analysis of nascent RNA-seq, RNA-seq and ChIP-seq datasets for THAP1, HCF1, and YY1. Related to Figure 3.

(**A** and **B**) Nascent RNA-seq showing relative differences (log2 fold-change) in the expression (CPM, counts per million mapped reads) of 53BP1 pathway-related genes in *Thap1*$^{-/-}$ versus WT MEFs (**A**) or *Thap1*$^{-/-}$*Brca1*$^{\Delta 11}$ versus *Brca1*$^{\Delta 11}$ MEFs (**B**). Data represent mean relative difference ± s.d. (n = 3).

(**C**) Scatter plot depicting relative differences (log2 fold-change) in gene expression (CPM) in *Thap1*$^{C54Y/C54Y}$ versus *WT* and *Thap1*$^{-/-}$ versus *WT* mESCs (RNA-seq, GSE86911) in relation to THAP1-bound genes (ChIP-seq, GSE86911). Genes that were bound by THAP1 and were either downregulated or upregulated in THAP1-deficient cells (both *Thap1*$^{C54Y/C54Y}$ and *Thap1*$^{-/-}$) are shown in blue and red, respectively. *Shld1* is indicated by the arrow.

(**D**) Levels of *Shld1* gene expression in *Hcfc1*$^{+/Y}$ and *Hcfc1*$^{HepKO/Y}$ hepatocytes. Data are from a publicly accessible RNA-seq dataset (GSE115768) and represent mean CPM ± s.d. (n = 2).

(**E**) Levels of *Shld1* gene expression in *Yy1*$^{flox/flox}$ and *Yy1*$^{flox/flox}$ TAT-Cre transfected B cells. Data are from a publicly accessible RNA-seq dataset (GSE145161) and represent mean CPM ± s.d. (n = 3, *p<0.05). Statistical significance was determined by the unpaired two-tailed Student's t test.

(**F**) Venn diagram depicting differentially expressed genes (log2 fold-change >2 and FDR <0.05) in *Hcfc1*$^{+/Y}$ versus *Hcfc1*$^{HepKO/Y}$ hepatocytes (RNA-seq, GSE115768) in relation to HCF1-bound genes in mESCs (ChIP-seq, GSE36030). The number of genes that were

previously shown to be bound by HCF1 and were either downregulated or upregulated in HCF1-deficient cells are shown in blue and red, respectively.

(**G**) Venn diagram depicting differentially expressed genes (log2 fold-change >2 and FDR <0.05) in *Yy1*$^{flox/flox}$ versus *Yy1*$^{flox/flox}$ TAT-Cre transfected B cells (RNA-seq, GSE145161) in relation to YY1-bound genes (ChIP-seq, GSE145161). The number of genes that were previously shown to be bound by YY1 and were either downregulated or upregulated in YY1-deficient cells are shown in blue and red, respectively.

(**H**) Western blot analysis of doxycycline-dependent expression of exogenous SHLD1 and THAP1 proteins in *WT* MEFs 24 to 96 hours after induction with doxycycline (Dox) as detected by anti-Flag antibody.

Figure S4. Ectopic SHLD3 expression does not restore PARPi sensitivity in *Thap1*$^{-/-}$ *Brca1*$^{\Delta 11}$ cells, despite normal focal accumulation. Related to Figure 4.

(**A**) Representative immunofluorescence images depicting normal accrual of ectopically expressed GFP-SHLD3 in three independent *Thap1*$^{-/-}$ *Brca1*$^{\Delta 11}$ MEF clones. Cells were irradiated with 10 Gy and analyzed 1 h post-IR.

(**B**) Quantification of the percentage of *Thap1*$^{-/-}$ *Brca1*$^{\Delta 11}$ MEFs containing >5 nuclear GFP-SHLD3 foci, as depicted in (A).

(**C**) Western blot analysis of doxycycline-dependent expression of exogenous SHLD1 and THAP1 proteins in *Thap1*$^{-/-}$ *Brca1*$^{\Delta 11}$ MEFs 96 hours (SHLD1) and 72 hours (THAP1) after induction with doxycycline (Dox), as detected by anti-Flag antibody.

(**D**) Viability of *WT*, *Brca1*$^{\Delta 11}$, *Thap1*$^{-/-}$*Brca1*$^{\Delta 11}$ MEFs transduced with an empty vector (*Thap1*$^{-/-}$ *Brca1*$^{\Delta 11}$+Mock) and two individual clones of *Thap1*$^{-/-}$*Brca1*$^{\Delta 11}$ MEFs

complemented with SHLD3 cDNA (*Thap1*$^{-/-}$*Brca1*$^{\Delta 11}$+SHLD3#1 and *Thap1*$^{-/-}$ *Brca1*$^{\Delta 11}$+SHLD3#2), as measured by CellTiter-Glo seven days after PARPi treatment.

Figure S5. Decreased *Shld1* expression and accumulation of unresolved DNA damage in the brain of *Thap1*$^{flox/flox}$ *Nestin-Cre* mice. Related to Figure 6.

(**A**) Volcano plot depicting relative differences (log2 fold-change) in gene expression (CPM) and the associated FDR in the striatum or cerebellum of *Thap1*$^{flox/flox}$ *Nestin-Cre* mice as compare to WT. Data are from a publicly accessible RNA-seq dataset (GSE123880). Genes that were shown to be bound by THAP1 (ChIP-seq, GSE: GSE86911) and were either downregulated or upregulated in *Thap1*$^{flox/flox}$ *Nestin-Cre* mice are shown in blue and red, respectively.

(**B**) DNA damage resulting from THAP1 loss, as visualized by positive γH2AX immunostaining, is detected in multiple brain regions. In the 6-week-old *Thap1*$^{Nes\text{-}cre}$ cortex and CA1 area of the hippocampal region, γH2AX (pSer139) is observed as nuclear puncta sporadically in neurons as shown by NeuN immunostaining.

(**C**) DNA damage, as indicated by phosphorylated KAP1 (pSer824), is also present sporadically throughout multiple brain regions including the cortex.

Material & Methods.

Cell lines and gene editing.

Mouse embryonic fibroblasts (MEFs) were generated as previously described (Callen et al., 2020). The *Brca1*$^{\Delta 11}$ and *Brca1*$^{\Delta 11}$*Trp53bp1*^{S25A} MEFs were generated from E13.5 embryos and grown in Dulbecco's Modified Eagle's Medium (DMEM, GIBCO)

supplemented with 15% heat-inactivated fetal bovine serum (FBS, Gemini Bio-Products) and 1% penicillin + strep- tomycin (GIBCO). To establish immortalized MEF cell lines, primary MEFs between passages 2-4 were transiently transfected with a vector encoding SV40 T-antigen (pCMV-SV40T). SV40-immortalized MEFs were routinely cultured in DMEM supplemented with 10 or 15% FBS. For generation of Cas9-expressing *Brca1*$^{\Delta 11}$ and *Brca1*$^{\Delta 11}$*Trp53bp1*^{S25A} MEFs, cells were transduced with the lentiCas9-Blast vector (Addgene #52962) and transductants were selected with blasticidin. Cells were then seeded at low densities (200–400 cells, depending on the cell line) on 15-cm dishes and single colonies were isolated using glass cylinders. Cas9 expression was confirmed by immunoblotting.

Generation of CRISPR knockout MEF and human cell lines were carried out as follows: Individual sgRNA (see Table S3 for sgRNA sequence) were cloned into LentiCRISPRv2 (Addgene #52961). To produce and package infectious lentivirus, HEK293T cells (American Type Culture Collection) at 80% confluency were co-transfected with LentiCRISPRv2-sgRNA, pRSV-Rev, pMDLg-pRRE and pHCMVG. After 72 h, viral supernatant was collected and passed through a 0.45 μm filter. Wild-type MEFs, *Brca1*$^{\Delta 11}$ MEFs or *TP53*$^{-/-}$*BRCA1*$^{-/-}$ hTERT-RPE1 cells (1×10^5) were seeded in 6-well plates and transduced using the filtered viral supernatant along with polybrene (10 μg/ml). RPE1-hTERT *TP53*$^{-/-}$*BRCA1*$^{-/-}$ and RPE1-hTERT *TP53*$^{-/-}$ cells were kind gifts from Dr. Daniel Durocher (Lunenfeld-Tanenbaum Research Institute) and the packaging vectors were kind gifts from Dr. Shyam K. Sharan (NCI-Frederick, NIH). After 48 h, cells were subjected to puromycin selection (MEFs, 1.5 μg/ml; RPE1, 15 μg/ml) for five days. Thereafter, single cells were sorted into 96-well plates on a BD FACSAria UV instrument

694 (BD Biosciences) and grown until colonies formed . Knockout clones were confirmed by
695 Sanger sequencing of PCR amplicons of the targeted region that underwent editing. When
696 different alleles of mutant clones had to be analysed, target regions were amplified by
697 PCR using 50 ng genomic DNA in MyTaq Red mix (Bioline) according to manufacture
698 instructions (primers are shown in Table S3). The amplicons were cloned using Zero Blunt
699 TOPO PCR Cloning Kit (Thermo Fisher Scientific), and transformed into STBL3
700 chemically competent E. coli. (Thermo Fisher Scientific). Bacterial colonies containing
701 plasmids with inserts were selected by colony-direct PCR, and were subsequently used for
702 direct Sanger sequencing.

703 Generation of *Thap1*$^{-/-}$*Brca2*^{Y3308X} clones was performed as follows. Guide RNA
704 against Thap1(see Table S3 for sgRNA sequence) was cloned into LentiCRISPRv2 with a
705 Neo selection marker (Addgene #98292). *Brca2*^{Y3308X} mES cells (Kuznetsov et al., 2008)
706 were grown on 0.1% gelatin (Sigma) coated dishes in Dulbecco's Modified Eagle's
707 Medium (DMEM, GIBCO) supplemented with 15% heat-inactivated fetal bovine serum
708 (ESC-tested, Thermo Fisher Scientific), LIF (1000 U/ml) (Millipore), 0.1 mM MEM non-
709 essential aminoacids, 1% glutamax and 55 μM β-mercaptoethanol. *Brca2*^{Y3308X} mES cells
710 (1×10^5) cells were seeded in 6-well plates and transduced using virus produced from
711 HEK293T cells along with polybrene (10 μg/ml). Thereafter, cells were selected with
712 G418 (0.2 mg/ml). *Thap1*$^{-/-}$ clones were confirmed by Sanger sequencing as described
713 above.

714 *Trp53bp1*$^{-/-}$, *Shld1*$^{-/-}$, and *Shld3*$^{-/-}$ CH12F3-2 cells were generated as previously
715 described (Gupta et al., 2018). *Thap1* mutant clones were edited through transient
716 transfection with the pX330 (Addgene #42230) plasmid constructs expressing Cas9 and

717 sgRNAs against *Thap1* (sgRNA: *Thap1*-sgRNA2, see Table S3), along with a plasmid
718 encoding hCD4 (pcDNA3.1-Hygro-delta-hCD4), using an Amaxa Nucleofector II
719 instrument (Lonza). Magnetic beads (STEMCELL Technologies 18052) were used to
720 isolate $hCD4^+$ cells 48 hour post-transfection. Thereafter, cells were grown for ~7 days
721 and subcloned by single cell sorting on a BD FACSAria UV instrument (BD Biosciences).
722 *Thap1*$^{-/-}$ clones were confirmed by Sanger sequencing as described above.
723 To generate GFP-SHLD3 expressing clones, *Thap1*$^{-/-}$*Brca1*$^{\Delta 11}$ MEFs were
724 transduced with pCW-eGFP-SHLD3 (Addgene #114126) along with polybrene (10
725 μg/ml). Forty-eight hours after infection, the cells were treated with 1 μg/ml doxycycline
726 for 24 h and single GFP^+ cells were sorted into 96-well plates on a BD FACSAria UV
727 instrument (BD Biosciences) and grown until colonies formed.
728 *WT* and *Thap1*$^{-/-}$*Brca1*$^{\Delta 11}$ MEFs expressing inducible exogenous THAP1 or
729 SHLD1 were generated as follows. The PB-TRE-dCas9-VPR plasmid (Addgene #63800)
730 was modified in which dCas9 was replaced by SnaBI, PacI and NotI restriction enzyme
731 recognition sites. Thereafter, C-terminal Myc-Flag-tagged sgRNA target site-mutated
732 version of the mouse *Thap1* cDNA (NM_199042) was inserted using SnaBI and PacI (PB-
733 TRE-mTHAP1). Alternatively, C-terminal Flag-tagged mouse SHLD1 cDNA
734 (NM_001358260.1) was inserted using SnaBI and NotI (PB-TRE-mSHLD1). sgRNA
735 target site mutation in mouse *Thap1* cDNA was generated by Q5 Site-Directed
736 Mutagenesis Kit (NEB, see Table S3 for primers). *WT* or *Thap1*$^{-/-}$*Brca1*$^{\Delta 11}$ MEFs (1×10^6)
737 were seeded in 10 cm dish and cultured for 24 hours. Cells were subsequently transfected
738 with 7 μg of PB-TRE-mTHAP1 or PB-TRE-mSHLD1 along with 1.4 μg of a transposase
739 expressing plasmid. Fourty-eight hours after transfection, cells were subjected to selection

with 250 μg/ml hygromycin (Thermo Fisher Scientific) for 5 days. Cells were then seeded at low densities (200–400 cells) on 15-cm dishes and single colonies were isolated using glass cylinders. Selected clones were cultured in medium containing 1 μg/ml doxycycline for the indicated amount of time, and THAP1/SHLD1 expression was confirmed by immunoblotting.

CRISPR–Cas9 screen

CRISPR–Cas9 screen was performed using genome-scale mouse Brie CRISPR knockout pooled library (Addgene #73633) (Doench et al., 2016). *Cas9*$^{+}$*Brca1*$^{\Delta 11}$ and *Cas9*$^{+}$*Brca1*$^{\Delta 11}$*Trp53bp1*^{S25A} MEFs were transduced at a multiplicity of infection (MOI) of 0.3 and 400-fold coverage of the library. Cells were then selected with puromycin for 3 days before treatment with the clinically-approved inhibitor Olaparib (PARPi, 100nM) for a further 14 days. This dose corresponds with to the IC95 for the BRCA1-deficient cells. Surviving clones from each condition were collected, and genomic DNA (gDNA) was isolated (Blood & Cell Culture DNA Midi Kit, Qiagen) and subjected to PCR with Illumina-compatible primers, followed by Illumina sequencing. Genes enriched or depleted in the PARPi-treated samples were determined with the MAGeCK software package version 0.5.9.2.

Multicolor growth competition assay (MCA)

One hundred thousand cells of two independent clones of Cas9-expressing *Brca1*$^{\Delta 11}$ MEFs (14223 and S12) were infected with either virus particles of NLS-mCherry LacZ-sgRNA or NLS-GFP GOI-sgRNA (*Thap1*, *Trp53bp1* or the empty vector), respectively. Ninety-

six hours after transduction, mCherry- and GFP-expressing cells were mixed 1:1 and 5 × 10^5 cells were seeded with or without PARPi (Olaparib, 100 nM) in 100 mm dishes. During the course of the experiment, cells were subcultured when they approach confluency was reached. PARPi-containing medium was replaced every three days. Cells were analyzed on the day of initial plating (t = 0) and on days 3, 6, 9 and 14 using the flow cytometer (LSRFortessa; BD) and the percentage of GFP- and mCherry-positive cells were analyzed using FlowJo software (Tree Star). LIVE/DEAD discrimination was performed by staining with 1 μg/ml of 4'6-diamidino-2-phenylindole (DAPI, Invitrogen).

Immunoblotting and Immunofluorescence

Western blotting was performed as described previously (Zong et al., 2019). Briefly, cells were collected and lysed in a buffer containing 50 mM Tris-HCl (pH 7.5), 200 mM NaCl, 5% Tween-20, 0.5% NP-40, 2 mM PMSF, 2.5 mM b-glycerophosphate (all from Sigma) and protease inhibitor cocktail tablet (complete Mini, Roche Diagnostics). Equal amounts of protein were loaded into precast mini-gels (Invitrogen) and resolved by SDS-PAGE. Proteins were blotted onto a nitrocellulose membrane, blocked with 5% membrane blocking agent (GE Healthcare) in TBS and incubated with the corresponding primary antibody. Primary antibodies were used at the following dilutions: anti-Flag (1:1000, Sigma) and anti-Tubulin (1:10,000, Sigma). Fluorescent secondary antibodies were used at a dilution of 1:15,000 (Li-Cor Biosciences). Detection of protein bands was performed by fluorescence imaging using a Li-Cor Odyssey CLx imaging system (Li-Cor Biosciences).

For immunofluorescence staining, MEFs were grown on 18 mm x 18 mm glass

coverslips. Prior to γ-irradiation (^{137}Cs Mark 1 irradiator, JL Shepherd), cells were incubated with 10 μM EdU (Invitrogen) for 20 min. Following irradiation, cells were allowed to recover for 1 hour or 4 hours. Thereafter, cells were pre-extracted (20 mM HEPES, 50 mM NaCl, 3 mM MgCl2, 0.3 M sucrose, 0.2% Triton X-100) on ice for 5 min to remove soluble nuclear proteins. Extracted samples were fixed (4% para-formaldehyde), permeabilized (0.5% Triton X-100), incubated with the indicated primary antibodies followed by appropriate fluorochrome-conjugated secondary antibodies (Invitrogen). Next, click-IT chemistry was performed as per manufacturer's instructions (Thermo Fisher Scientific) and DNA was counterstained with DAPI (Thermo Fisher Scientific). Images were captured at 63X magnification with an AxioCam MRc5 mounted on an Axio Observer Z1 epifluorescence microscope (Zeiss) or at 40x magnification on a Lionheart LX automated microscope (BioTek Instruments, Inc.). Quantification of nuclear foci and total nuclear intensity was performed using the Gen5 spot analysis software (BioTek). The antibodies used for standard immunofluorescence experiments were anti-RIF1 (1:5,000, gift of Davide Robbiani, Rockefeller University), anti-53BP1 (1:1000, Novus), anti-RAD51 (1:250, Abcam), anti-RPA (1:5,000, Abcam), and anti-GFP (1:500, Roche).

Metaphase spreads and cell viability assays

MEFs and mESCs were treated with 0.5 μM or 1 μM PARPi (Olaparib, Selleckchem) for 16 hours, subsequently arrested at mitosis with 0.1 mg/ml colcemid (Roche) and metaphase chromosome spreads were prepared as previously described (Zong et al., 2019). Images were acquired using a Metafer automated scanning and imaging platform

(MetaSystems). Clonogenic survival for a given treatment was calculated relative to the plating efficiency in non-treated controls. To determine cell growth and viability, MEFs were plated in 6-well plates (10,000 per well) and treated continuously with different doses of PARPi or cisplatin (Sigma) for 7 days. The drug-containing medium was replenished every three days and cells were subcultured when they approach confluency. On day 10, cell viability was determined using the CellTiter-Glo Luminescent Cell Viability Assay (Promega) as per manufacturer's instructions.

Nascent RNA-seq

Four million MEFs were labeled with 0.5 mM 5-ethynyl uridine (EU) for 30 min. Total RNA was extracted using TRIzol (Ambion). The NEBNext rRNA Depletion kit (human/mouse/rat) (New England Biosciences) was used to deplete rRNA from 1 μg of total RNA prior to sample biotinylation through Click-it reactions (Click-iT Nascent RNA Capture Kit, ThermoFisher C10365) as per the manufacturer's specification. First-strand cDNA synthesis of the captured nascent RNA was done using the SuperScript VILO cDNA synthesis kit (Invitrogen), followed by AMPure XP purification (1.8X) and elution in (20 μl), Thereafter, second-strand cDNA synthesis was carried out in a total reaction volume of 30 μl that contained 0.6 mM dNTP, 1.2 mM of dUTP, 2 units of RNase H (Invitrogen) and 20 units of E. coli DNA polymerase I (Invitrogen), for 2.5 hr at 16°C. Double stranded cDNA was cleaned using 1.8X Agencourt AMPure XP beads and eluted in 20 μl of EB that was used for end-repair. End-repair was performed in 50 μl of T4 ligase reaction buffer containing 0.4 mM of dNTPs, 3 units of T4 DNA polymerase (NEB), 9 units of T4 Polynucleotide Kinase (NEB) and 1 unit of Klenow fragment (NEB)

832 at 24°C for 30 min in a ThermoMixer C at 400 rpm. End-repair reaction was cleaned using
833 1.8X Agencourt AMPure XP beads and eluted in 15 µl of EB that was used for A-tailing
834 reaction in 30 µl of NEBNext dA-Tailing reaction buffer (NEB) with 7.5 units of Klenow
835 fragment exo- (NEB) at 37°C for 30 min. The 30 µl of the A-tailing reaction were mixed
836 with Quick Ligase buffer 2X (NEB), 3,000 units of Quick ligase and 5 nM of annealed
837 adaptor (Illumina truncated adaptor) in a volume of 75 µl and incubated at 25°C for 20
838 min. The adaptor was prepared by annealing the following HPLC-grade oligos: 5′-
839 Phos/GATCGGAAGAGCACACGTCT-3′and 5′-
840 ACACTCTTTCCCTACACGACGCTCTTCCGATC∗T-3′ (∗phosphorothioate bond). The
841 ligation reaction was terminated by adding 50mM of EDTA and cleaned with 1.8X
842 Agencourt AMPure XP beads and eluted in 15 µl of EB. Thereafter, samples were treated
843 with 0.5 units of Uracil-DNA glycosylase (Thermofisher) for 15 min at 37°C and used for
844 PCR amplification in a 50 µl reaction volume containing 1 µM of TruSeq barcoded primer
845 p5,
846 AATGATACGGCGACCACCGAGATCTACACNNNNNNNNACACTCTTTCCCTACA
847 CGACGCTCTTCCGATC*T, TruSeq barcoded primer p7,
848 CAAGCAGAAGACGGCATACGAGANNNNNNNNGTGACTGGAGTTCAGACGTGT
849 GCTCTTCCGATC*T, (NNNNNNNN represents barcode and * a phosphothiorate bond),
850 and 2X Kapa HiFi HotStart Ready mix (Kapa Biosci-ences). The temperature settings
851 during the PCR amplification were 45 s at 98°C followed by 15 cycles of 15 s at 98°C, 30
852 s at 63°C, 30 s at 72°C and a final 5 min extension at 72°C. PCR reactions were cleaned
853 with Agencourt AMPure XP beads (Beckman Coulter), run on a 2% agarose gel and a
854 smear of 200-500bp was cut and gel purified using QIAquick Gel Extraction Kit

855 (QIAGEN). Library concentration was determined with KAPA Library Quantification Kit
856 for Illumina Platforms (Kapa Biosystems). Sequencing was performed on the Illumina
857 Nextseq500 (75 bp single end reads).
858
859 **ChIP-Seq**
860 ChIP-seq was performed as described previously (Shinoda et al., 2019) with a rabbit
861 polyclonal antibody against THAP1 (Proteintech, 12584-1-AP). Twenty million MEFs
862 were fixed in fresh media by adding 37% formaldehyde (F1635, Sigma) to attain a final
863 concentration of 1% and incubated at 37°C for 10 min. Fixation was quenched by addition
864 of 1 M glycine (Sigma) in PBS at a final concentration of 125 mM. Cells were washed
865 twice with cold PBS and pellets were snap frozen in dry ice and stored at -80°C. Fixed cell
866 pellets were thawed on ice and resuspended in 2 ml of cold RIPA buffer (10 mM TrisHCl
867 pH 7.5, 1 mM EDTA, 0.1% SDS, 0.1% sodium deoxycholate, 1% Triton X-100,
868 supplemented with proteinase inhibitor (Complete Mini EDTA free, Roche)). Sonication
869 was performed using the Covaris S220 sonicator at duty cycle 20%, peak incident power
870 175, cycle/burst 200 for 30 cycles of 60 s sonication and 30 s of pause at 4°C. Chromatin
871 was clarified by centrifugation at 21,000 g at 4°C for 10 min and precleared with 80 µl
872 prewashed Dynabeads protein A (ThermoFisher) for 30 min at 4°C. 40 µl of prewashed
873 Dynabeads protein A were incubated with 10 µg of antibody in 100 µl of PBS for 30 min
874 at room temperature under continuous mixing, washed twice in PBS for 5 min and added
875 to 1 ml of chromatin followed by overnight incubation at 4°C on a rotator. Beads were
876 then collected in a magnetic separator (DynaMag-2 Invitrogen), washed twice with cold
877 RIPA buffer, twice with RIPA buffer containing 0.3 M NaCl, twice with LiCl buffer (0.25

878 M LiCl, 0.5% Igepal-630, 0.5% sodium deoxycholate), once with TE (10 mM Tris pH 8.0,
879 1mM EDTA) plus 0.2% Triton X-100, and once with TE. Crosslinking was reversed by
880 incubating the beads at 65°C for 4 hr in the presence of 0.3% SDS and 1 mg/ml of
881 Proteinase K (Ambion). DNA was purified using Zymo ChIP DNA clean and concentrator
882 kit (Zymo Research) and eluted in 20 µl. The entire ChIP DNA was used to prepare
883 Illumina sequencing libraries. End-repair was performed in 75 µl of T4 ligase reaction
884 buffer, 0.4 mM of dNTPs, 4 U of T4 DNA polymerase (NEB), 13.5 U of T4
885 Polynucleotide Kinase (NEB) and 1.5 U of Klenow fragment (NEB) at 24°C for 30 min in
886 a ThermoMixer C at 400 rpm. End-repair reaction was cleaned using 2X Agencourt
887 AMPure XP beads and eluted in 16.5 µl of EB that was used for A-tailing reaction in 20 µl
888 of NEBNext dA-Tailing reaction buffer (NEB) with 7.5 U of Klenow fragment exo-
889 (NEB) at 37°C for 30 min. The 20 µl of the A-tailing reaction was mixed with Quick
890 Ligase buffer 2X (NEB), 3,000 U of Quick ligase and 5 nM of annealed adaptor (Illumina
891 truncated adaptor) in a volume of 50 µl and incubated at 25°C for 20 min. Ligation was
892 stopped by adding 50 mM of EDTA and cleaned with 1.8X Agencourt AM- Pure XP
893 beads and eluted in 15 µl of EB that was used for PCR amplification in a 50 µl reaction
894 with 1 µM primers TruSeq barcoded primer p5, TruSeq barcoded primer p7, and 2X Kapa
895 HiFi HotStart Ready mix (Kapa Biosciences). The temperature settings during the PCR
896 amplification were 45 s at 98°C followed by 14 cycles of 15 s at 98°C, 30 s at 63°C, 30 s
897 at 72°C and a final 5 min extension at 72°C . PCR reactions were cleaned with Agencourt
898 AMPure XP beads (Beckman Coulter), run on a 2% agarose gel and a smear of 200-500bp
899 was cut and gel purified using QIAquick Gel Extraction Kit (QIAGEN). Library
900 concentration was determined with KAPA Library Quantification Kit for Illumina

Platforms (Kapa Biosystems). Sequencing was performed on the Illumina Nextseq500 (75bp single-end reads).

Class switch recombination (CSR) assay in CH12-F3 cells

The mouse B lymphocyte cell line CH12-F3 cells was cultured and stimulated with 250 ng/ml CD40L, 10 ng/ml IL4 and 1 ng/ml TGFβ (CIT) to induce immunoglobulin class switching from IgM to IgA. For FACS analysis of CSR efficiency, live CH12 cells were labeled with FITC-conjugated anti-IgM (ThermoFisher Scientific) and PE-conjugated anti-IgA (Southern Biotech) antibodies. All analyses were performed on FACS CantoII (BD Biosciences).

TCGA analysis

Publicly available TCGA ovarian serous cystadenocarcinoma data on cBioportal (Cerami et al., 2012; Gao et al., 2013) were queried for *THAP1* and *SHLD1* expression data. Other clinical characteristics and information on *BRCA1* and *BRCA2* gene or expression alteration were also extracted for the 316 ovarian carcinoma tumors in the cohort (Cancer Genome Atlas Research, 2011). In total, 38 carcinomas with a *BRCA1* mutation and 34 carcinomas harboring a *BRCA2* mutation were isolated from the cohort and used for further analysis. All calculations were performed using and Graph Pad Prism v.7.0.

Analysis of orthologous genes between species

Orthology presnce/absence data is based on Ensembl orthology data versions 98, 99 and 100 including human Ensembl genes used to mine Ensembl orthologs. Drosophila and

Nematostella genes that were found to be orthologs of the human genes were used to further mine Ensembl Metazoa for additional animal orthologs, and human genes were further used to mine Pan-compara orthologs which includes taxa not present in the standard Ensembl dataset. A minimum of 15% identity of both query gene and target gene were required to avoid false positives.

Histology and immunodetection.

For cryosectioning, paraformaldehyde-perfused tissue was cryoprotected in buffered 25% sucrose (w/v) solution. Brains were embedded in Tissue-Tek OCT compound (Sakura), and sectioned sagittally at 9 μm using a CM3050S cryostat (Leica). Prior to immunodetection, sections were subjected to antigen retrieval according to the manufacturer's directions (HistoVT One, Nacalai Tesque). Immunodetection was done with the following primary antibodies: Calbindin (1:500, Sigma, C9848), phospho-KAP-1 (S824) (1:500, Bethyl, A300-767A), NeuN (1:500, Chemicon, MAB377) and phospho-H2AX-Ser-139 (1:200; Cell Signaling, #2577). For fluorescence immunodetection, FITC- and Cy3-conjugated secondary antibodies (Jackson Immunologicals) were used and counterstained with DAPI (Vector Laboratories).

Previously published RNA-seq and ChIP-seq datasets used in this study

RNA-seq and ChIP-seq datasets for *Thap1*$^{-/-}$ mESC were obtained from GSE86911(Aguilo et al., 2017). RNA-seq dataset for cerebellum and striatal tissues of *Thap1*$^{flox/flox}$ *Nestin-Cre* mice were obtained from GSE123880 (Frederick et al., 2019). RNA-seq and ChIP-seq datasets for *Yy1*$^{flox/flox}$ TAT-Cre primary B cells were obtained from GSE145161(Nandi et

al., 2020). ChIP-seq dataset for YY1 in mESC were from GSE68195 (Sigova et al., 2015). RNA-seq dataset for hepatocytes of *Hcfc1*$^{HepKO/Y}$ mice were obtained from GSE115768 (Minocha et al., 2019). ChIP-seq dataset for HCF1 in mESC were from the Mouse ENCODE Project (GSE36030) (Mouse et al., 2012; Yue et al., 2014).

Statistical analyses

Unless indicated, all data are presented as individual replicates. The total number of replicates, mean and error bars are explained in the figure legends. The statistical tests (Student's, Welch's, one-way ANOVA and log-rank) and resultant P values (represented by asterisks) are indicated in the figure legends and/or figure panels and were calculated using GraphPad Prism and R software (ns = $p > 0.05$; * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$).

DATA AND CODE AVAILABILITY

The accession number for the datasets reported in this paper is available at GEO with accession number: GSE154729.

REFERENCES

Aguilo, F., Zakirova, Z., Nolan, K., Wagner, R., Sharma, R., Hogan, M., Wei, C., Sun, Y., Walsh, M.J., Kelley, K.*, et al.* (2017). THAP1: Role in Mouse Embryonic Stem Cell Survival and Differentiation. Stem Cell Reports *9*, 92-107.

Barazas, M., Annunziato, S., Pettitt, S.J., de Krijger, I., Ghezraoui, H., Roobol, S.J., Lutz, C., Frankum, J., Song, F.F., Brough, R.*, et al.* (2018). The CST Complex Mediates End Protection at Double-Strand Breaks and Promotes PARP Inhibitor Sensitivity in BRCA1-Deficient Cells. Cell Rep *23*, 2107-2118.

Belotserkovskaya, R., Raga Gil, E., Lawrence, N., Butler, R., Clifford, G., Wilson, M.D., and Jackson, S.P. (2020). PALB2 chromatin recruitment restores homologous recombination in BRCA1-deficient cells depleted of 53BP1. Nat Commun *11*, 819.

Blanchard, A., Ea, V., Roubertie, A., Martin, M., Coquart, C., Claustres, M., Beroud, C., and Collod-Beroud, G. (2011). DYT6 dystonia: review of the literature and creation of the UMD Locus-Specific Database (LSDB) for mutations in the THAP1 gene. Hum Mutat *32*, 1213-1224.

Boersma, V., Moatti, N., Segura-Bayona, S., Peuscher, M.H., van der Torre, J., Wevers, B.A., Orthwein, A., Durocher, D., and Jacobs, J.J.L. (2015). MAD2L2 controls DNA repair at telomeres and DNA breaks by inhibiting 5' end resection. Nature *521*, 537-540.

Bouwman, P., Aly, A., Escandell, J.M., Pieterse, M., Bartkova, J., van der Gulden, H., Hiddingh, S., Thanasoula, M., Kulkarni, A., Yang, Q.*, et al.* (2010). 53BP1 loss rescues BRCA1 deficiency and is associated with triple-negative and BRCA-mutated breast cancers. Nat Struct Mol Biol *17*, 688-695.

Bowman-Colin, C., Xia, B., Bunting, S., Klijn, C., Drost, R., Bouwman, P., Fineman, L., Chen, X., Culhane, A.C., Cai, H.*, et al.* (2013). Palb2 synergizes with Trp53 to suppress mammary tumor formation in a model of inherited breast cancer. Proc Natl Acad Sci U S A *110*, 8632-8637.

Bressman, S.B., Raymond, D., Fuchs, T., Heiman, G.A., Ozelius, L.J., and Saunders-Pullman, R. (2009). Mutations in THAP1 (DYT6) in early-onset dystonia: a genetic screening study. Lancet Neurol *8*, 441-446.

Bunting, S.F., Callen, E., Wong, N., Chen, H.T., Polato, F., Gunn, A., Bothmer, A., Feldhahn, N., Fernandez-Capetillo, O., Cao, L.*, et al.* (2010). 53BP1 inhibits homologous recombination in Brca1-deficient cells by blocking resection of DNA breaks. Cell *141*, 243-254.

Byrum, A.K., Vindigni, A., and Mosammaparast, N. (2019). Defining and Modulating 'BRCAness'. Trends Cell Biol *29*, 740-751.

Callen, E., Di Virgilio, M., Kruhlak, M.J., Nieto-Soler, M., Wong, N., Chen, H.T., Faryabi, R.B., Polato, F., Santos, M., Starnes, L.M.*, et al.* (2013). 53BP1 mediates productive and mutagenic DNA repair through distinct phosphoprotein interactions. Cell *153*, 1266-1280.

Callen, E., Zong, D., Wu, W., Wong, N., Stanlie, A., Ishikawa, M., Pavani, R., Dumitrache, L.C., Byrum, A.K., Mendez-Dorantes, C.*, et al.* (2020). 53BP1 Enforces Distinct Pre- and Post-resection Blocks on Homologous Recombination. Mol Cell *77*, 26-38 e27.

Cancer Genome Atlas Research, N. (2011). Integrated genomic analyses of ovarian carcinoma. Nature *474*, 609-615.
Cassagnes, L.E., Chhour, M., Perio, P., Sudor, J., Gayon, R., Ferry, G., Boutin, J.A., Nepveu, F., and Reybier, K. (2018). Oxidative stress and neurodegeneration: The possible contribution of quinone reductase 2. Free Radic Biol Med *120*, 56-61.
Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E.*, et al.* (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov *2*, 401-404.
Charlesworth, G., Bhatia, K.P., and Wood, N.W. (2013). The genetics of dystonia: new twists in an old tale. Brain *136*, 2017-2037.
Clairmont, C.S., Sarangi, P., Ponnienselvan, K., Galli, L.D., Csete, I., Moreau, L., Adelmant, G., Chowdhury, D., Marto, J.A., and D'Andrea, A.D. (2020). TRIP13 regulates DNA repair pathway choice through REV7 conformational change. Nat Cell Biol *22*, 87-96.
Dai, Y., Zhang, F., Wang, L., Shan, S., Gong, Z., and Zhou, Z. (2020). Structural basis for shieldin complex subunit 3-mediated recruitment of the checkpoint protein REV7 during DNA double-strand break repair. J Biol Chem *295*, 250-262.
Dev, H., Chiang, T.W., Lescale, C., de Krijger, I., Martin, A.G., Pilger, D., Coates, J., Sczaniecka-Clift, M., Wei, W., Ostermaier, M.*, et al.* (2018). Shieldin complex promotes DNA end-joining and counters homologous recombination in BRCA1-null cells. Nat Cell Biol *20*, 954-965.
Djarmati, A., Schneider, S.A., Lohmann, K., Winkler, S., Pawlack, H., Hagenah, J., Bruggemann, N., Zittel, S., Fuchs, T., Rakovic, A.*, et al.* (2009). Mutations in THAP1 (DYT6) and generalised dystonia with prominent spasmodic dysphonia: a genetic screening study. Lancet Neurol *8*, 447-452.
Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R.*, et al.* (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat Biotechnol *34*, 184-191.
Findlay, S., Heath, J., Luo, V.M., Malina, A., Morin, T., Coulombe, Y., Djerir, B., Li, Z., Samiei, A., Simo-Cheyou, E.*, et al.* (2018). SHLD2/FAM35A co-operates with REV7 to coordinate DNA double-strand break repair pathway choice. EMBO J *37*.
Frederick, N.M., Shah, P.V., Didonna, A., Langley, M.R., Kanthasamy, A.G., and Opal, P. (2019). Loss of the dystonia gene Thap1 leads to transcriptional deficits that converge on common pathogenic pathways in dystonic syndromes. Hum Mol Genet *28*, 1343-1356.
Fuchs, T., Gavarini, S., Saunders-Pullman, R., Raymond, D., Ehrlich, M.E., Bressman, S.B., and Ozelius, L.J. (2009). Mutations in the THAP1 gene are responsible for DYT6 primary torsion dystonia. Nat Genet *41*, 286-288.
Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E.*, et al.* (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal *6*, pl1.
Gao, S., Feng, S., Ning, S., Liu, J., Zhao, H., Xu, Y., Shang, J., Li, K., Li, Q., Guo, R.*, et al.* (2018). An OB-fold complex controls the repair pathways for DNA double-strand breaks. Nat Commun *9*, 3925.

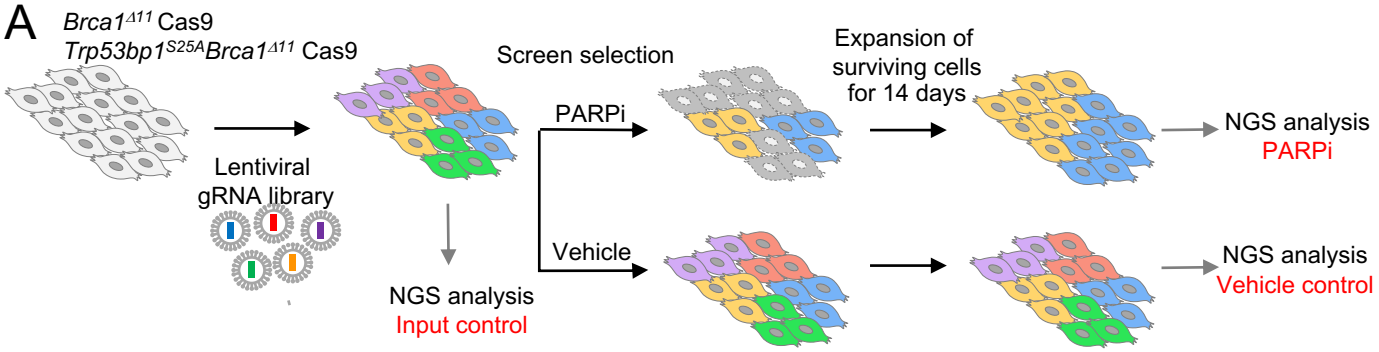
Ghezraoui, H., Oliveira, C., Becker, J.R., Bilham, K., Moralli, D., Anzilotti, C., Fischer, R., Deobagkar-Lele, M., Sanchiz-Calvo, M., Fueyo-Marcos, E.*, et al.* (2018). 53BP1 cooperation with the REV7-shieldin complex underpins DNA structure-specific NHEJ. Nature *560*, 122-127.
Gupta, R., Somyajit, K., Narita, T., Maskey, E., Stanlie, A., Kremer, M., Typas, D., Lammers, M., Mailand, N., Nussenzweig, A.*, et al.* (2018). DNA Repair Network Analysis Reveals Shieldin as a Key Regulator of NHEJ and PARP Inhibitor Sensitivity. Cell *173*, 972-988 e923.
He, Y.J., Meghani, K., Caron, M.C., Yang, C., Ronato, D.A., Bian, J., Sharma, A., Moore, J., Niraj, J., Detappe, A.*, et al.* (2018). DYNLL1 binds to MRE11 to limit DNA end resection in BRCA1-deficient cells. Nature *563*, 522-526.
Hollstein, R., Reiz, B., Kotter, L., Richter, A., Schaake, S., Lohmann, K., and Kaiser, F.J. (2017). Dystonia-causing mutations in the transcription factor THAP1 disrupt HCFC1 cofactor recruitment and alter gene expression. Hum Mol Genet *26*, 2975-2983.
Huang, T.H., Fowler, F., Chen, C.C., Shen, Z.J., Sleckman, B., and Tyler, J.K. (2018). The Histone Chaperones ASF1 and CAF-1 Promote MMS22L-TONSL-Mediated Rad51 Loading onto ssDNA during Homologous Recombination in Human Cells. Mol Cell *69*, 879-892 e875.
Jaspers, J.E., Kersbergen, A., Boon, U., Sol, W., van Deemter, L., Zander, S.A., Drost, R., Wientjens, E., Ji, J., Aly, A.*, et al.* (2013). Loss of 53BP1 causes PARP inhibitor resistance in Brca1-mutated mouse mammary tumors. Cancer Discov *3*, 68-81.
Karanam, K., Kafri, R., Loewer, A., and Lahav, G. (2012). Quantitative live cell imaging reveals a gradual shift between DNA repair mechanisms and a maximal use of HR in mid S phase. Mol Cell *47*, 320-329.
Kuznetsov, S.G., Liu, P., and Sharan, S.K. (2008). Mouse embryonic stem cell-based functional assay to evaluate mutations in BRCA2. Nat Med *14*, 875-881.
Li, W., Xu, H., Xiao, T., Cong, L., Love, M.I., Zhang, F., Irizarry, R.A., Liu, J.S., Brown, M., and Liu, X.S. (2014). MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. Genome Biol *15*, 554.
Madabhushi, R., Gao, F., Pfenning, A.R., Pan, L., Yamakawa, S., Seo, J., Rueda, R., Phan, T.X., Yamakawa, H., Pao, P.C.*, et al.* (2015). Activity-Induced DNA Breaks Govern the Expression of Neuronal Early-Response Genes. Cell *161*, 1592-1605.
Malewicz, M., and Perlmann, T. (2014). Function of transcription factors at DNA lesions in DNA repair. Exp Cell Res *329*, 94-100.
Manis, J.P., Morales, J.C., Xia, Z., Kutok, J.L., Alt, F.W., and Carpenter, P.B. (2004). 53BP1 links DNA damage-response pathways to immunoglobulin heavy chain class-switch recombination. Nat Immunol *5*, 481-487.
Mazars, R., Gonzalez-de-Peredo, A., Cayrol, C., Lavigne, A.C., Vogel, J.L., Ortega, N., Lacroix, C., Gautier, V., Huet, G., Ray, A.*, et al.* (2010). The THAP-zinc finger protein THAP1 associates with coactivator HCF-1 and O-GlcNAc transferase: a link between DYT6 and DYT3 dystonias. J Biol Chem *285*, 13364-13371.
McKinnon, P.J. (2009). DNA repair deficiency and neurological disease. Nat Rev Neurosci *10*, 100-112.
McKinnon, P.J. (2017). Genome integrity and disease prevention in the nervous system. Genes Dev *31*, 1180-1194.

Meissner, W.G., Fernet, M., Couturier, J., Hall, J., Lauge, A., Henry, P., Stoppa-Lyonnet, D., and Tison, F. (2013). Isolated generalized dystonia in biallelic missense mutations of the ATM gene. Mov Disord *28*, 1897-1899.
Melis, C., Beauvais, G., Muntean, B.S., Cinaru, M.-D., Otrimski, G., Creus-Muncunill, J., Martemyanov, K.A., Gonzalez-Alegre, M.D., and Ehrlich, M.E. (2020). Increased striatal dopamine-induced ERK phosphorylation in a mouse model of DYT1 dystonia: a shared mechanism with levodopa-induced dyskinesia? submitted
Michaud, J., Praz, V., James Faresse, N., Jnbaptiste, C.K., Tyagi, S., Schutz, F., and Herr, W. (2013). HCFC1 is a common component of active human CpG-island promoters and coincides with ZNF143, THAP11, YY1, and GABP transcription factor occupancy. Genome Res *23*, 907-916.
Minocha, S., Villeneuve, D., Praz, V., Moret, C., Lopes, M., Pinatel, D., Rib, L., Guex, N., and Herr, W. (2019). Rapid Recapitulation of Nonalcoholic Steatohepatitis upon Loss of Host Cell Factor 1 Function in Mouse Hepatocytes. Mol Cell Biol *39*.
Mirman, Z., and de Lange, T. (2020). 53BP1: a DSB escort. Genes Dev *34*, 7-23.
Mirman, Z., Lottersberger, F., Takai, H., Kibe, T., Gong, Y., Takai, K., Bianchi, A., Zimmermann, M., Durocher, D., and de Lange, T. (2018). 53BP1-RIF1-shieldin counteracts DSB resection through CST- and Polalpha-dependent fill-in. Nature *560*, 112-116.
Mouse, E.C., Stamatoyannopoulos, J.A., Snyder, M., Hardison, R., Ren, B., Gingeras, T., Gilbert, D.M., Groudine, M., Bender, M., Kaul, R.*, et al.* (2012). An encyclopedia of mouse DNA elements (Mouse ENCODE). Genome Biol *13*, 418.
Nandi, S., Liang, G., Sindhava, V., Angireddy, R., Basu, A., Banerjee, S., Hodawadekar, S., Zhang, Y., Avadhani, N.G., Sen, R.*, et al.* (2020). YY1 control of mitochondrial-related genes does not account for regulation of immunoglobulin class switch recombination in mice. Eur J Immunol.
Noordermeer, S.M., Adam, S., Setiaputra, D., Barazas, M., Pettitt, S.J., Ling, A.K., Olivieri, M., Alvarez-Quilon, A., Moatti, N., Zimmermann, M.*, et al.* (2018). The shieldin complex mediates 53BP1-dependent DNA repair. Nature *560*, 117-121.
Ochs, F., Karemore, G., Miron, E., Brown, J., Sedlackova, H., Rask, M.B., Lampe, M., Buckle, V., Schermelleh, L., Lukas, J.*, et al.* (2019). Stabilization of chromatin topology safeguards genome integrity. Nature *574*, 571-574.
Ochs, F., Somyajit, K., Altmeyer, M., Rask, M.B., Lukas, J., and Lukas, C. (2016). 53BP1 fosters fidelity of homology-directed DNA repair. Nat Struct Mol Biol *23*, 714-721.
Paiano, J., Wu, W., Yamada, S., Sciascia, N., Callen, E., Paola Cotrim, A., Deshpande, R.A., Maman, Y., Day, A., Paull, T.T.*, et al.* (2020). ATM and PRDM9 regulate SPO11-bound recombination intermediates during meiosis. Nat Commun *11*, 857.
Pannunzio, N.R., Watanabe, G., and Lieber, M.R. (2018). Nonhomologous DNA end-joining for repair of DNA double-strand breaks. J Biol Chem *293*, 10512-10523.
Petersen, S., Casellas, R., Reina-San-Martin, B., Chen, H.T., Difilippantonio, M.J., Wilson, P.C., Hanitsch, L., Celeste, A., Muramatsuk, M., Pilch, D.R.*, et al.* (2001). AID is required to initiate Nbs1/gamma-H2AX focus formation and mutations at sites of class switching. Nature *414*, 660-665.
Rooney, S., Chaudhuri, J., and Alt, F.W. (2004). The role of the non-homologous end-joining pathway in lymphocyte development. Immunol Rev *200*, 115-131.

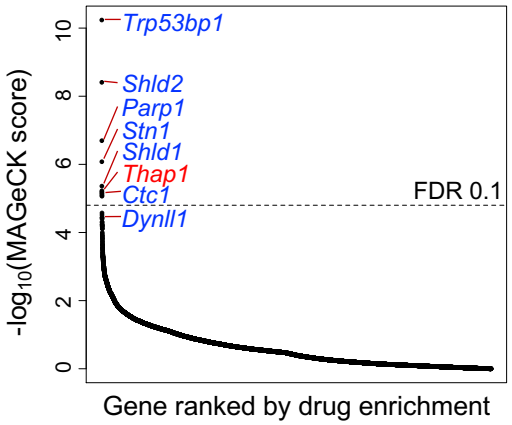
1143 Roussigne, M., Kossida, S., Lavigne, A.C., Clouaire, T., Ecochard, V., Glories, A., Amalric,
1144 F., and Girard, J.P. (2003). The THAP domain: a novel protein motif with similarity to
1145 the DNA-binding domain of P element transposase. Trends Biochem Sci *28*, 66-69.
1146 Ruiz, M., Perez-Garcia, G., Ortiz-Virumbrales, M., Meneret, A., Morant, A., Kottwitz, J.,
1147 Fuchs, T., Bonet, J., Gonzalez-Alegre, P., Hof, P.R.*, et al.* (2015). Abnormalities of motor
1148 function, transcription and cerebellar structure in mouse models of THAP1 dystonia.
1149 Hum Mol Genet *24*, 7159-7170.
1150 Saunders-Pullman, R., Raymond, D., Stoessl, A.J., Hobson, D., Nakamura, K., Pullman,
1151 S., Lefton, D., Okun, M.S., Uitti, R., Sachdev, R.*, et al.* (2012). Variant ataxia-
1152 telangiectasia presenting as primary-appearing dystonia in Canadian Mennonites.
1153 Neurology *78*, 649-657.
1154 Scully, R., Panday, A., Elango, R., and Willis, N.A. (2019). DNA double-strand break
1155 repair-pathway choice in somatic mammalian cells. Nat Rev Mol Cell Biol *20*, 698-
1156 714.
1157 Setiaputra, D., and Durocher, D. (2019). Shieldin - the protector of DNA ends. EMBO
1158 Rep *20*.
1159 Shinoda, K., Maman, Y., Canela, A., Schatz, D.G., Livak, F., and Nussenzweig, A. (2019).
1160 Intra-Vkappa Cluster Recombination Shapes the Ig Kappa Locus Repertoire. Cell Rep
1161 *29*, 4471-4481 e4476.
1162 Sigova, A.A., Abraham, B.J., Ji, X., Molinie, B., Hannett, N.M., Guo, Y.E., Jangi, M.,
1163 Giallourakis, C.C., Sharp, P.A., and Young, R.A. (2015). Transcription factor trapping
1164 by RNA in gene regulatory elements. Science *350*, 978-981.
1165 Slatter, M.A., and Gennery, A.R. (2010). Primary immunodeficiency syndromes. Adv
1166 Exp Med Biol *685*, 146-165.
1167 Tarsounas, M., and Sung, P. (2020). The antitumorigenic roles of BRCA1-BARD1 in
1168 DNA repair and replication. Nat Rev Mol Cell Biol *21*, 284-299.
1169 Tomida, J., Takata, K.I., Bhetawal, S., Person, M.D., Chao, H.P., Tang, D.G., and Wood,
1170 R.D. (2018). FAM35A associates with REV7 and modulates DNA damage responses of
1171 normal and BRCA1-defective cells. EMBO J *37*.
1172 Wang, L., Zhao, H., Li, J., Xu, Y., Lan, Y., Yin, W., Liu, X., Yu, L., Lin, S., Du, M.Y.*, et al.*
1173 (2020). Identifying functions and prognostic biomarkers of network motifs marked
1174 by diverse chromatin states in human cell lines. Oncogene *39*, 677-689.
1175 Wang, Y., Bernhardy, A.J., Cruz, C., Krais, J.J., Nacson, J., Nicolas, E., Peri, S., van der
1176 Gulden, H., van der Heijden, I., O'Brien, S.W.*, et al.* (2016). The BRCA1-Delta11q
1177 Alternative Splice Isoform Bypasses Germline Mutations and Promotes Therapeutic
1178 Resistance to PARP Inhibition and Cisplatin. Cancer Res *76*, 2778-2790.
1179 Ward, I.M., Reina-San-Martin, B., Olaru, A., Minn, K., Tamada, K., Lau, J.S., Cascalho, M.,
1180 Chen, L., Nussenzweig, A., Livak, F.*, et al.* (2004). 53BP1 is required for class switch
1181 recombination. J Cell Biol *165*, 459-464.
1182 Xu, G., Chapman, J.R., Brandsma, I., Yuan, J., Mistrik, M., Bouwman, P., Bartkova, J.,
1183 Gogola, E., Warmerdam, D., Barazas, M.*, et al.* (2015). REV7 counteracts DNA double-
1184 strand break resection and affects PARP inhibition. Nature *521*, 541-544.
1185 Yellajoshyula, D., Liang, C.C., Pappas, S.S., Penati, S., Yang, A., Mecano, R., Kumaran, R.,
1186 Jou, S., Cookson, M.R., and Dauer, W.T. (2017). The DYT6 Dystonia Protein THAP1
1187 Regulates Myelination within the Oligodendrocyte Lineage. Dev Cell *42*, 52-67 e54.

Yue, F., Cheng, Y., Breschi, A., Vierstra, J., Wu, W., Ryba, T., Sandstrom, R., Ma, Z., Davis, C., Pope, B.D.*, et al.* (2014). A comparative encyclopedia of DNA elements in the mouse genome. Nature *515*, 355-364.
Zhao, W., Wiese, C., Kwon, Y., Hromas, R., and Sung, P. (2019). The BRCA Tumor Suppressor Network in Chromosome Damage Repair by Homologous Recombination. Annu Rev Biochem *88*, 221-245.
Zong, D., Adam, S., Wang, Y., Sasanuma, H., Callen, E., Murga, M., Day, A., Kruhlak, M.J., Wong, N., Munro, M.*, et al.* (2019). BRCA1 Haploinsufficiency Is Masked by RNF168-Mediated Chromatin Ubiquitylation. Mol Cell *73*, 1267-1281 e1267.
Zong, D., Callen, E., Pegoraro, G., Lukas, C., Lukas, J., and Nussenzweig, A. (2015). Ectopic expression of RNF168 and 53BP1 increases mutagenic but not physiological non-homologous end joining. Nucleic Acids Res *43*, 4950-4961.

Figure 1

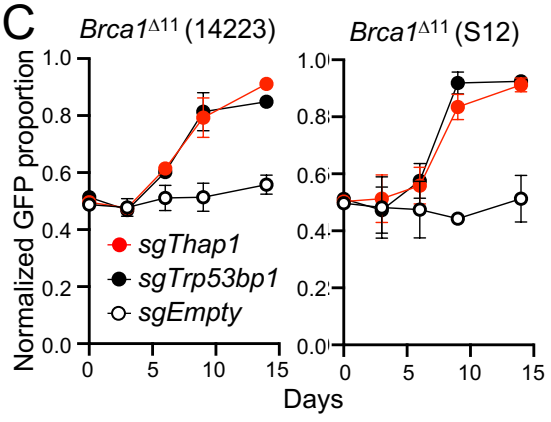


B

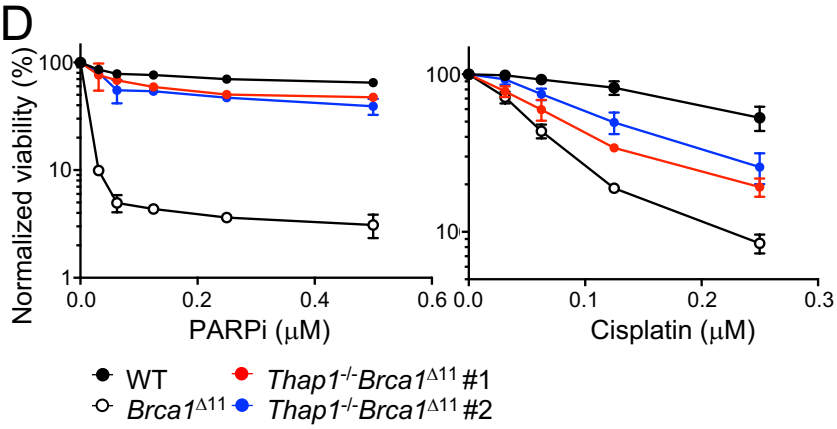


Gene	-log10 (MAGeCK score)	Vehicle vs. PARPi		
		P value	FDR	Rank
Trp53bp1	10.237	2.52E-07	0.00248	1
Shld2	8.408	2.52E-07	0.00248	2
Parp1	6.695	7.55E-07	0.00495	3
Stn1	6.077	2.77E-06	0.01361	4
Shld1	5.362	1.61E-05	0.06076	5
Thap1	5.223	2.34E-05	0.06076	6
Ctc1	5.190	2.44E-05	0.06076	7
Atp6v1b2	5.154	2.74E-05	0.06076	8
Smu1	5.124	3.05E-05	0.06076	9
Hist1h2ac	5.085	3.30E-05	0.06076	10
Zfp706	5.065	3.40E-05	0.06076	11
Dynll1	4.438	0.00014369	0.17791	14

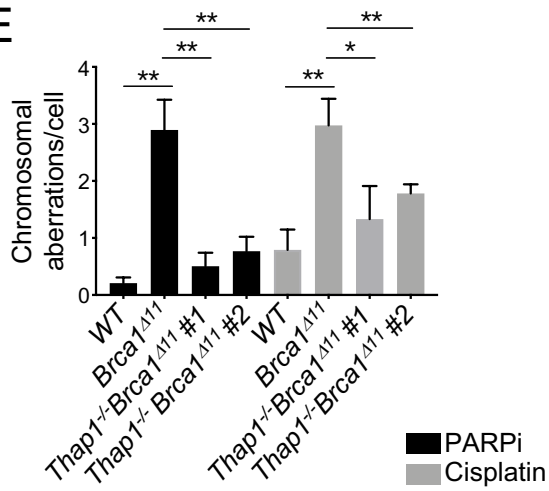
C



D



E



F

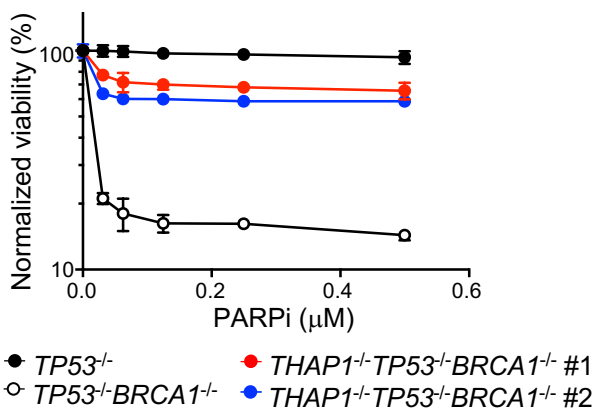


Figure 2

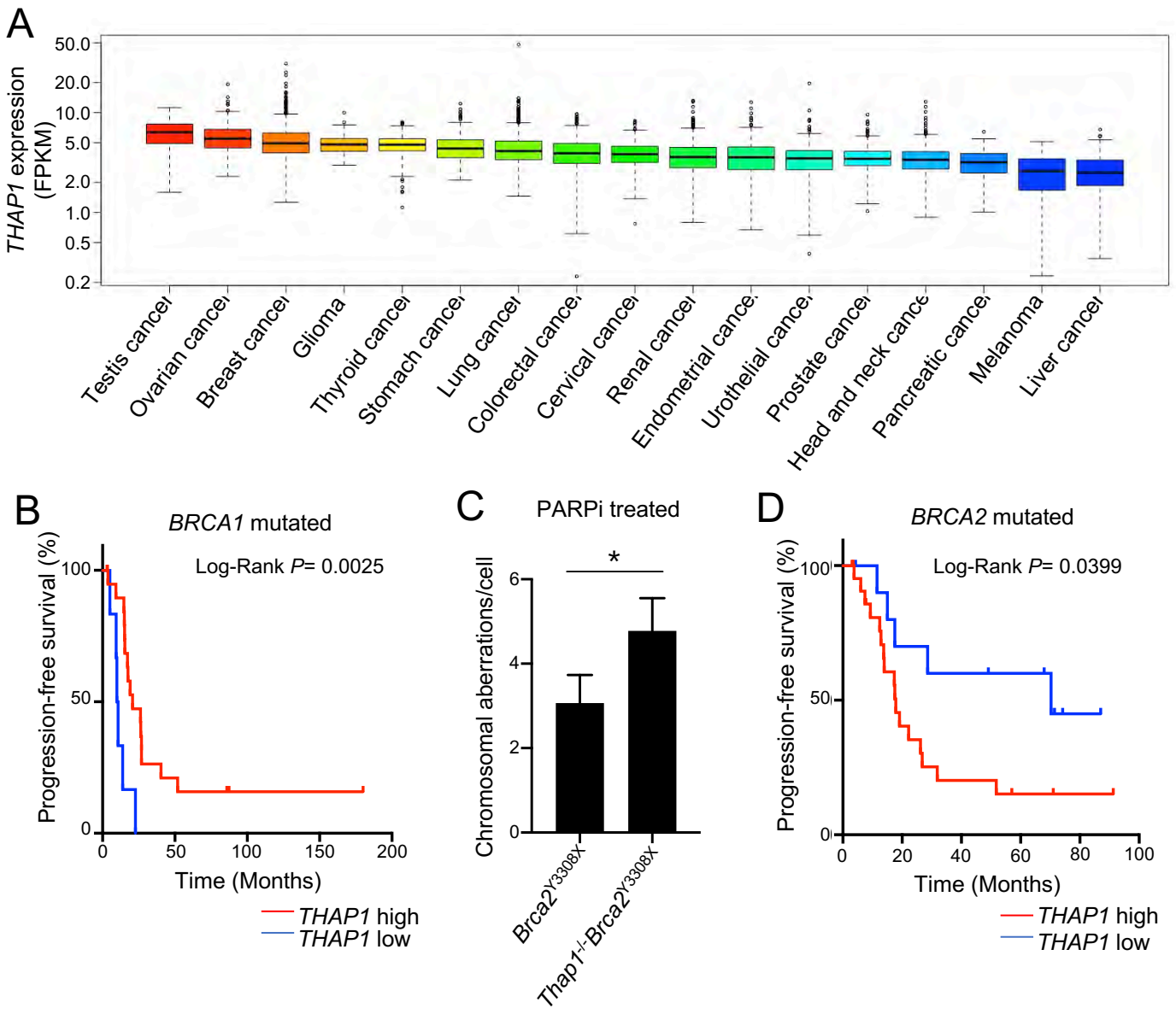

A
THAP1 expression (FPKM)
50.0
20.0
10.0
5.0
2.0
1.0
0.5
0.2
Testis cancer
Ovarian cancer
Breast cancer
Glioma
Thyroid cancer
Stomach cancer
Lung cancer
Colorectal cancer
Cervical cancer
Renal cancer
Endometrial cancer
Urothelial cancer
Prostate cancer
Head and neck cancer
Pancreatic cancer
Melanoma
Liver cancer
B
BRCA1 mutated
Log-Rank P= 0.0025
Progression-free survival (%)
100
50
0
0
50
100
150
200
Time (Months)
THAP1 high
THAP1 low
C
PARPi treated
*
Chromosomal aberrations/cell
6
4
2
0
Brca2Y3308X
Thap1-/-Brca2Y3308X
D
BRCA2 mutated
Log-Rank P= 0.0399
Progression-free survival (%)
100
50
0
0
20
40
60
80
100
Time (Months)
THAP1 high
THAP1 low

Figure 3

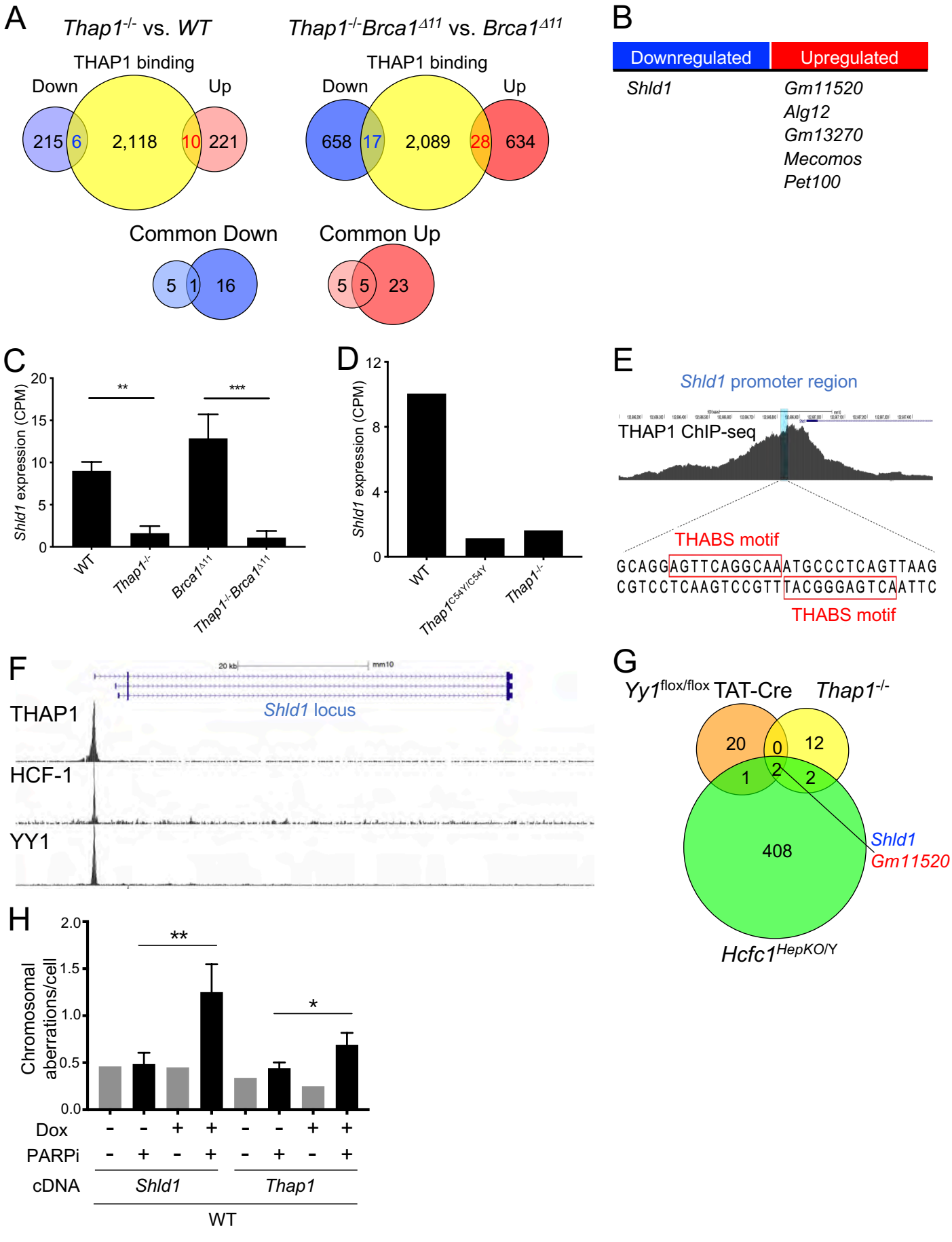

A
Thap1-/- vs. WT
THAP1 binding
Down
Up
215
6
2,118
10
221
Thap1-/-Brca1Δ11 vs. Brca1Δ11
THAP1 binding
Down
Up
658
17
2,089
28
634
Common Down
5
1
16
Common Up
5
5
23
B
Downregulated
Upregulated
Shld1
Gm11520
Alg12
Gm13270
Mecomos
Pet100
C
Shld1 expression (CPM)
20
15
10
5
0
**
***
WT
Thap1-/-
Brca1Δ11
Thap1-/-Brca1Δ11
D
Shld1 expression (CPM)
12
8
4
0
WT
Thap1C54Y/C54Y
Thap1-/-
E
Shld1 promoter region
THAP1 ChIP-seq
THABS motif
GCAGGAGTTCAGGCAAATGCCCTCAGTTAAG
CGTCCTCAAGTCCGTTTACGGGAGTCAATTC
THABS motif
F
20 kb
mm10
Shld1 locus
THAP1
HCF-1
YY1
G
Yy1flox/flox TAT-Cre
Thap1-/-
20
0
12
1
2
2
408
Shld1
Gm11520
Hcfc1HepKO/Y
H
Chromosomal aberrations/cell
2.0
1.5
1.0
0.5
0.0
**
*
Dox
PARPi
cDNA
Shld1
Thap1
WT

Figure 4

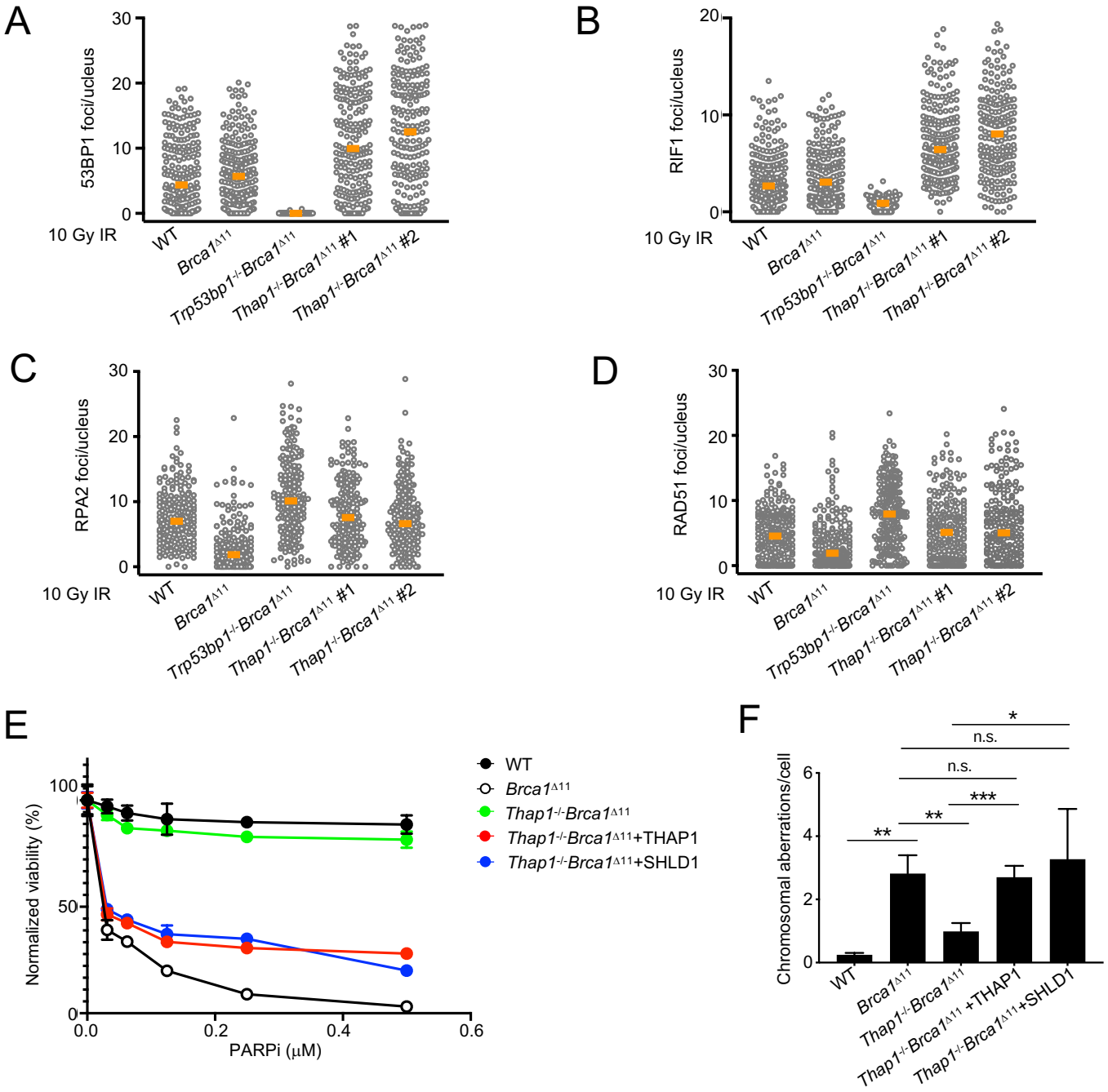
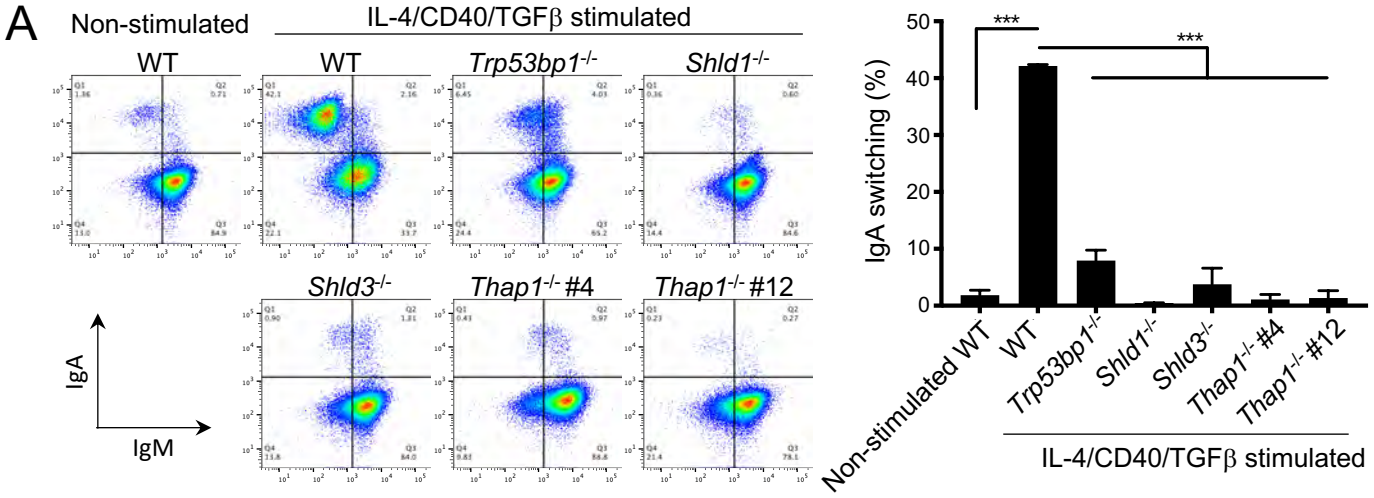

A
53BP1 foci/ucleus
30
20
10
0
10 Gy IR
WT
Brca1Δ11
Trp53bp1-/-Brca1Δ11
Thap1-/-Brca1Δ11 #1
Thap1-/-Brca1Δ11 #2
B
RIF1 foci/ucleus
20
10
0
C
RPA2 foci/ucleus
D
RAD51 foci/ucleus
E
Normalized viability (%)
100
50
0
0.0
0.2
0.4
0.6
PARPi (μM)
WT
Brca1Δ11
Thap1-/-Brca1Δ11
Thap1-/-Brca1Δ11+THAP1
Thap1-/-Brca1Δ11+SHLD1
F
Chromosomal aberrations/cell
6
4
2
0
n.s.
n.s.
***
**
**
*
WT
Brca1Δ11
Thap1-/-Brca1Δ11
Thap1-/-Brca1Δ11 +THAP1
Thap1-/-Brca1Δ11+SHLD1

Figure 5


A
Non-stimulated
IL-4/CD40/TGFβ stimulated
WT
WT
Trp53bp1-/-
Shld1-/-
Shld3-/-
Thap1-/- #4
Thap1-/- #12
IgA
IgM
IgA switching (%)
50
40
30
20
10
0
***
***
Non-stimulated WT
WT
Trp53bp1-/-
Shld1-/-
Shld3-/-
Thap1-/- #4
Thap1-/- #12
IL-4/CD40/TGFβ stimulated

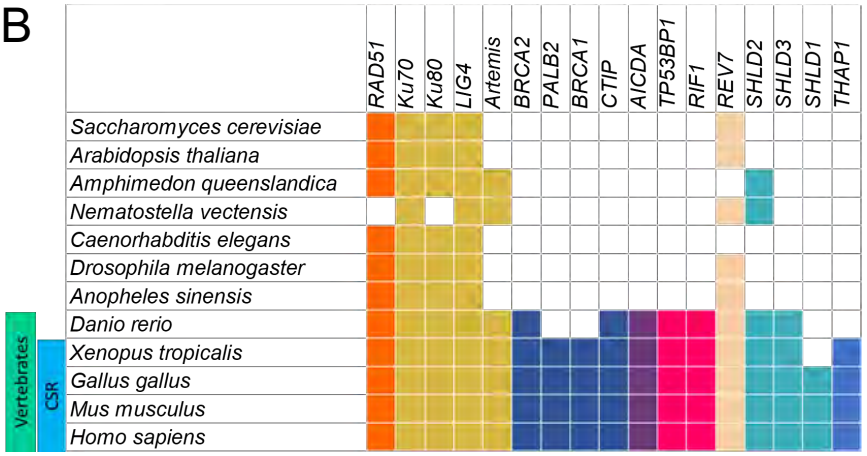

B
RAD51
Ku70
Ku80
LIG4
Artemis
BRCA2
PALB2
BRCA1
CTIP
AICDA
TP53BP1
RIF1
REV7
SHLD2
SHLD3
SHLD1
THAP1
Saccharomyces cerevisiae
Arabidopsis thaliana
Amphimedon queenslandica
Nematostella vectensis
Caenorhabditis elegans
Drosophila melanogaster
Anopheles sinensis
Danio rerio
Xenopus tropicalis
Gallus gallus
Mus musculus
Homo sapiens
Vertebrates
CSR

Figure 6

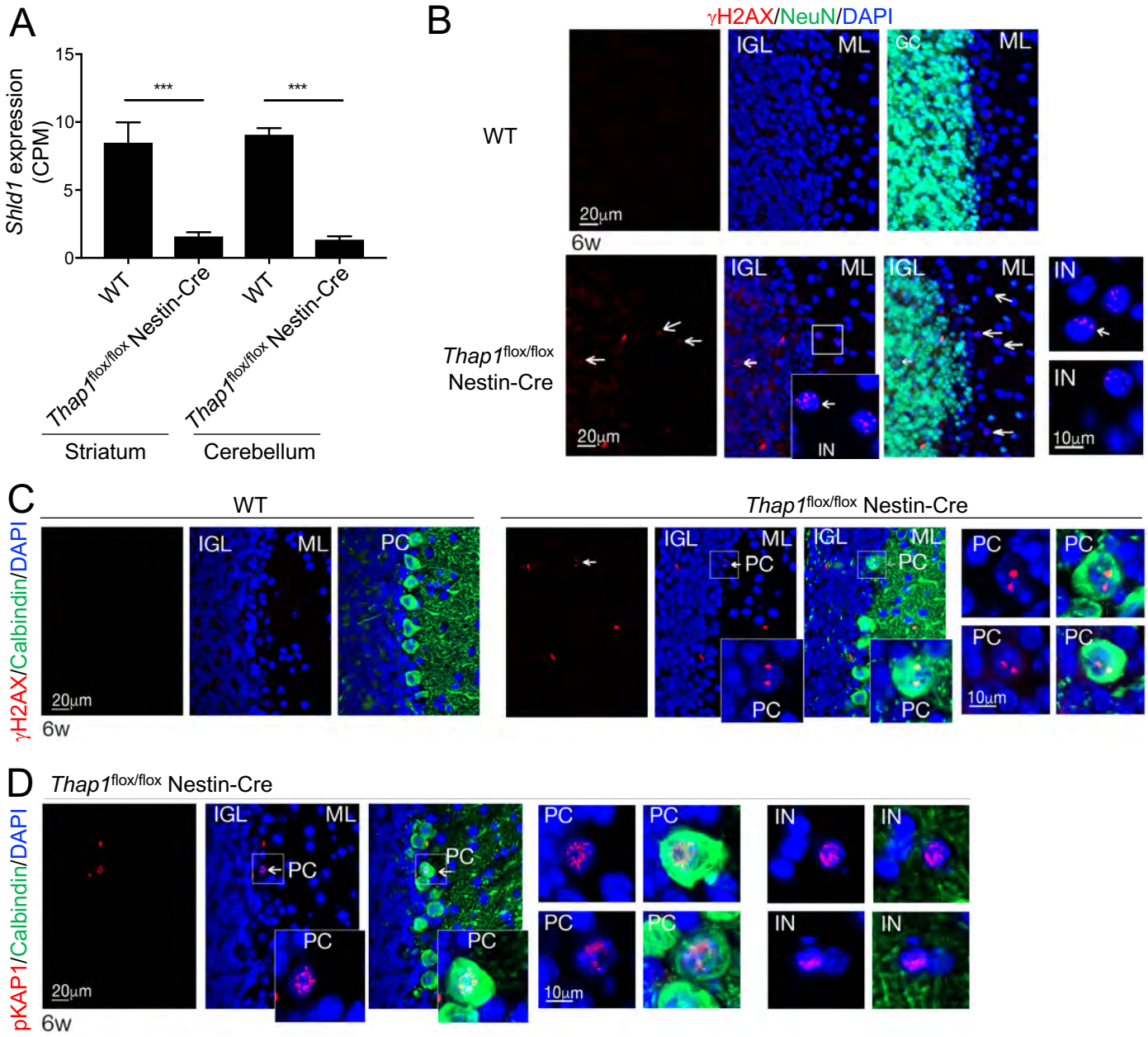

A
Shld1 expression (CPM)
15
10
5
0
***
***
WT
Thap1flox/flox Nestin-Cre
WT
Thap1flox/flox Nestin-Cre
Striatum
Cerebellum
B
γH2AX/NeuN/DAPI
WT
IGL
ML
GC
ML
20μm
6w
Thap1flox/flox
Nestin-Cre
IGL
ML
IGL
ML
IN
IN
20μm
IN
10μm
C
WT
Thap1flox/flox Nestin-Cre
γH2AX/Calbindin/DAPI
IGL
ML
PC
IGL
ML
PC
IGL
ML
PC
PC
PC
PC
PC
20μm
10μm
6w
D
Thap1flox/flox Nestin-Cre
pKAP1/Calbindin/DAPI
IGL
ML
PC
PC
PC
PC
PC
PC
IN
IN
PC
PC
IN
IN
20μm
10μm
6w

Figure S1

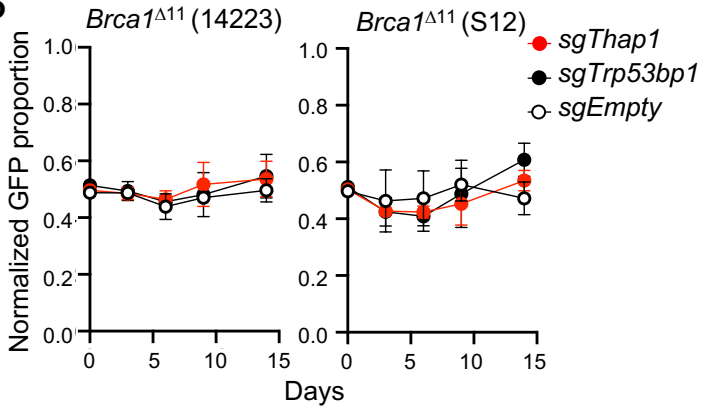
A

Mouse *Thap1* exon2:

5'-TGGATTGTCTCTCCTTAGGTTTCCTCTTACTCGCCCCAG
CCTTTGTAAGCAGTGGGAGGCAGCTGTTAAAAGGAAAAAC
TTCAAGCCCACCAAGTACAGCAGCATCTGCTCGGAGCACT
TCACCCCGGACTGCTTTAAGAGGGAGTGCAACAACAAGCT
ACT-3'

PAM sgRNA1 sgRNA2

B



C

Thap1$^{-/-}$*Brca1*$^{\Delta 11}$ (#1)

WT sequence GTCTCTCCTTAGGTTTCCTCTTACTCGCCCCAG
mut allele 1 (7bp deletion) GTCTCTC_____TTCCTCTTACTCGCCCCAG
mut allele 2 (1bp deletion) GTCTCTCCTTAGGTT_CCTCTTACTCGCCCCAG

Thap1$^{-/-}$*Brca1*$^{\Delta 11}$ (#2)

WT sequence TCACCCCGGACTGCTTTAAGAGGGAGTGCAAC
mut allele 1 (2bp deletion) TCACCCCGGA__GCTTTAAGAGGGAGTGCAAC
mut allele 2 (17bp deletion) TCACCCCGG_____TGCAAC

PAM sgRNA1 sgRNA2

D

Thap1$^{-/-}$

WT sequence AGTCAGAGAACTGGATTGTCTCTCCTTAGGTTTCCTCTTACTCGCCCCAGCCTTTGTAAGCAGTGGG
mut allele 1 (2bp deletion) AGTCAGAGAACTGGATTGTCTCTCCTTAG__TTCCTCTTACTCGCCCCAGCCTTTGTAAGCAGTGGG
mut allele 2 (36bp deletion) AGTCAGAGAACTGGATT_____TTGTAAGCAGTGGG

PAM sgRNA

E

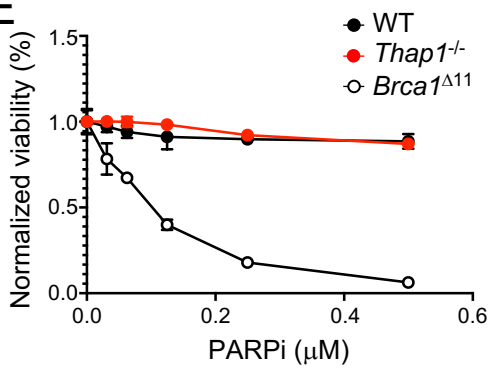


Figure S2

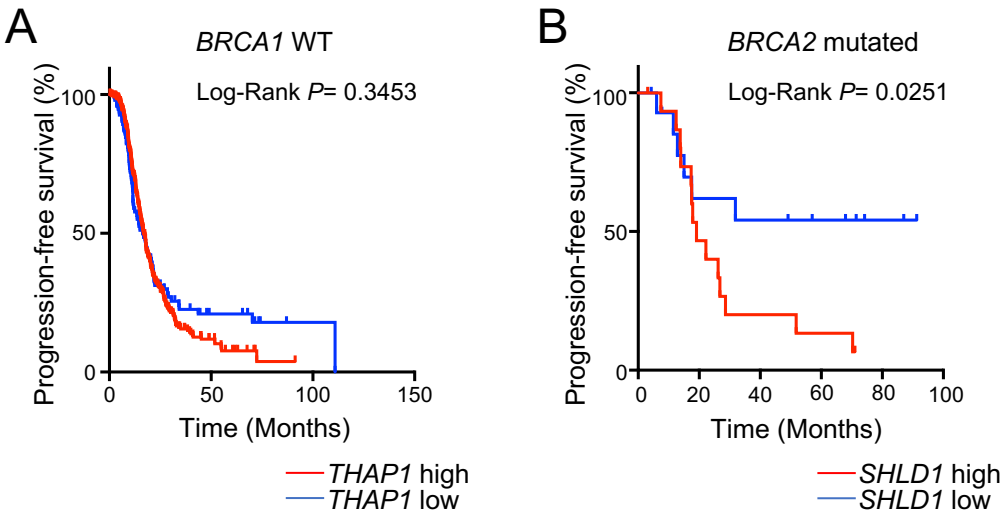

A
BRCA1 WT
Log-Rank P= 0.3453
Progression-free survival (%)
100
50
0
0
50
100
150
Time (Months)
THAP1 high
THAP1 low
B
BRCA2 mutated
Log-Rank P= 0.0251
Progression-free survival (%)
100
50
0
0
20
40
60
80
100
Time (Months)
SHLD1 high
SHLD1 low

Figure S3

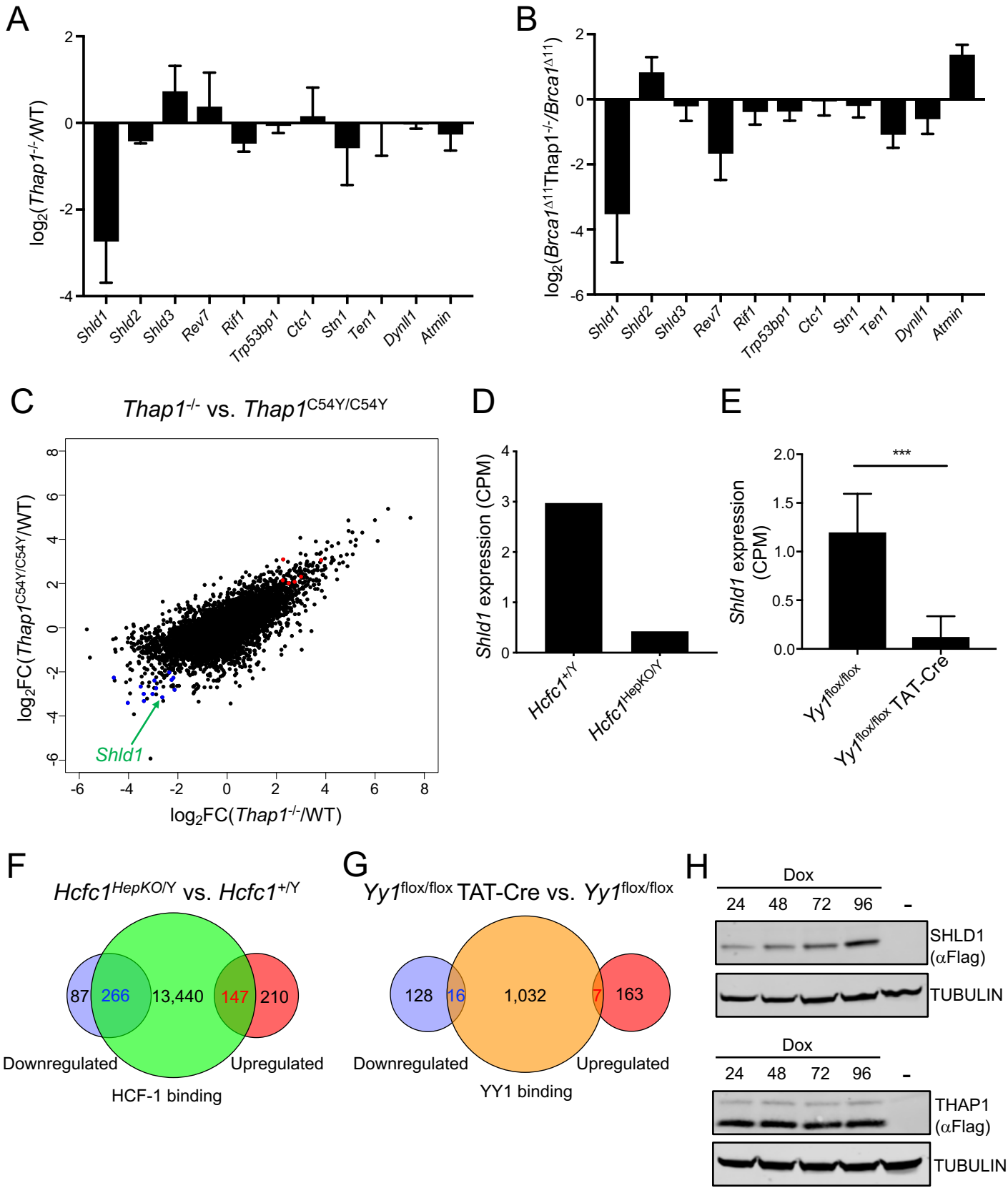

A
log2(Thap1-/-/WT)
Shld1
Shld2
Shld3
Rev7
Rif1
Trp53bp1
Ctc1
Stn1
Ten1
Dynll1
Atmin
B
log2(Brca1Δ11Thap1-/-/Brca1Δ11)
C
Thap1-/- vs. Thap1C54Y/C54Y
log2FC(Thap1C54Y/C54Y/WT)
log2FC(Thap1-/-/WT)
Shld1
D
Shld1 expression (CPM)
Hcfc1+/Y
Hcfc1HepKO/Y
E
Shld1 expression (CPM)
***
Yy1flox/flox
Yy1flox/flox TAT-Cre
F
Hcfc1HepKO/Y vs. Hcfc1+/Y
87
266
13,440
147
210
Downregulated
Upregulated
HCF-1 binding
G
Yy1flox/flox TAT-Cre vs. Yy1flox/flox
128
16
1,032
7
163
Downregulated
Upregulated
YY1 binding
H
Dox
24
48
72
96
-
SHLD1
(αFlag)
TUBULIN
Dox
24
48
72
96
-
THAP1
(αFlag)
TUBULIN

Figure S4

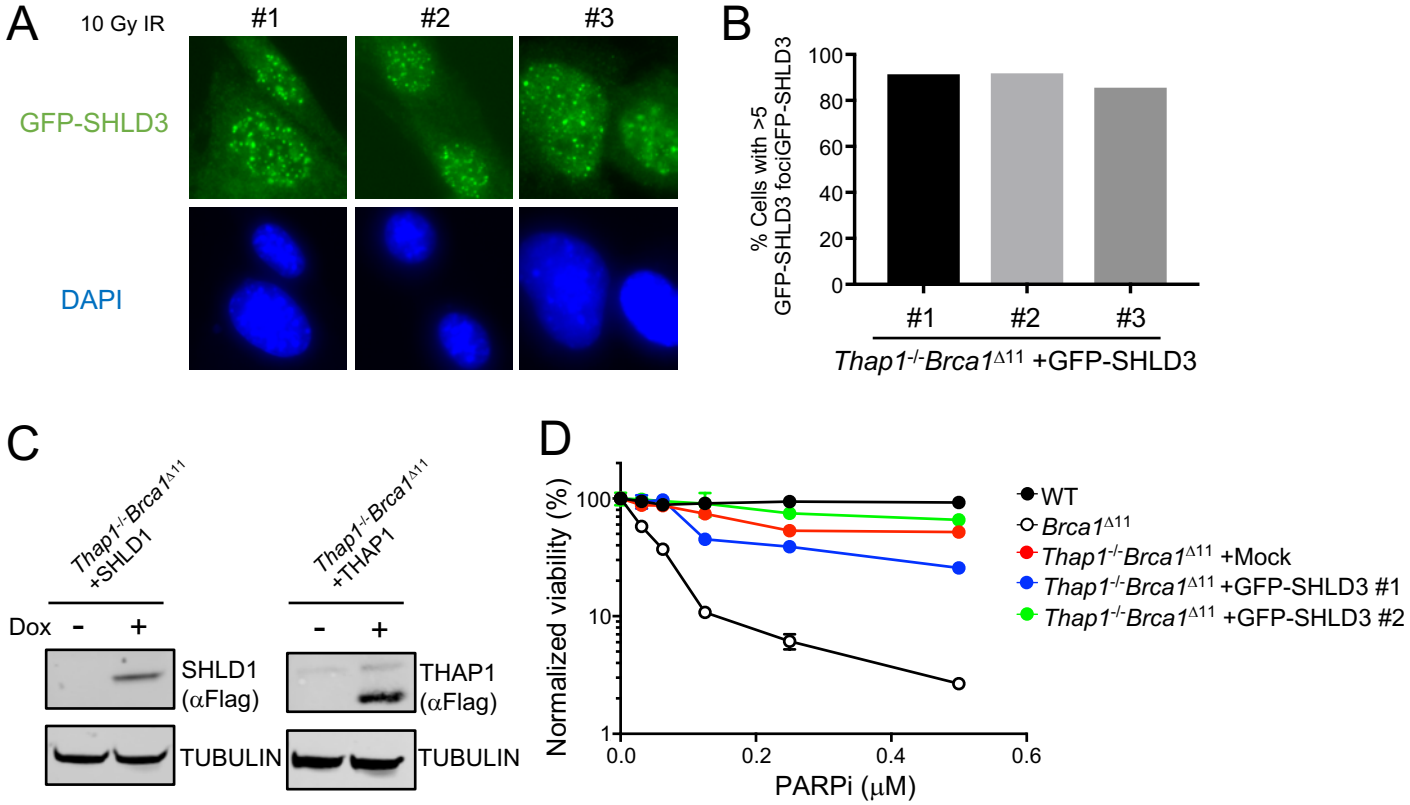
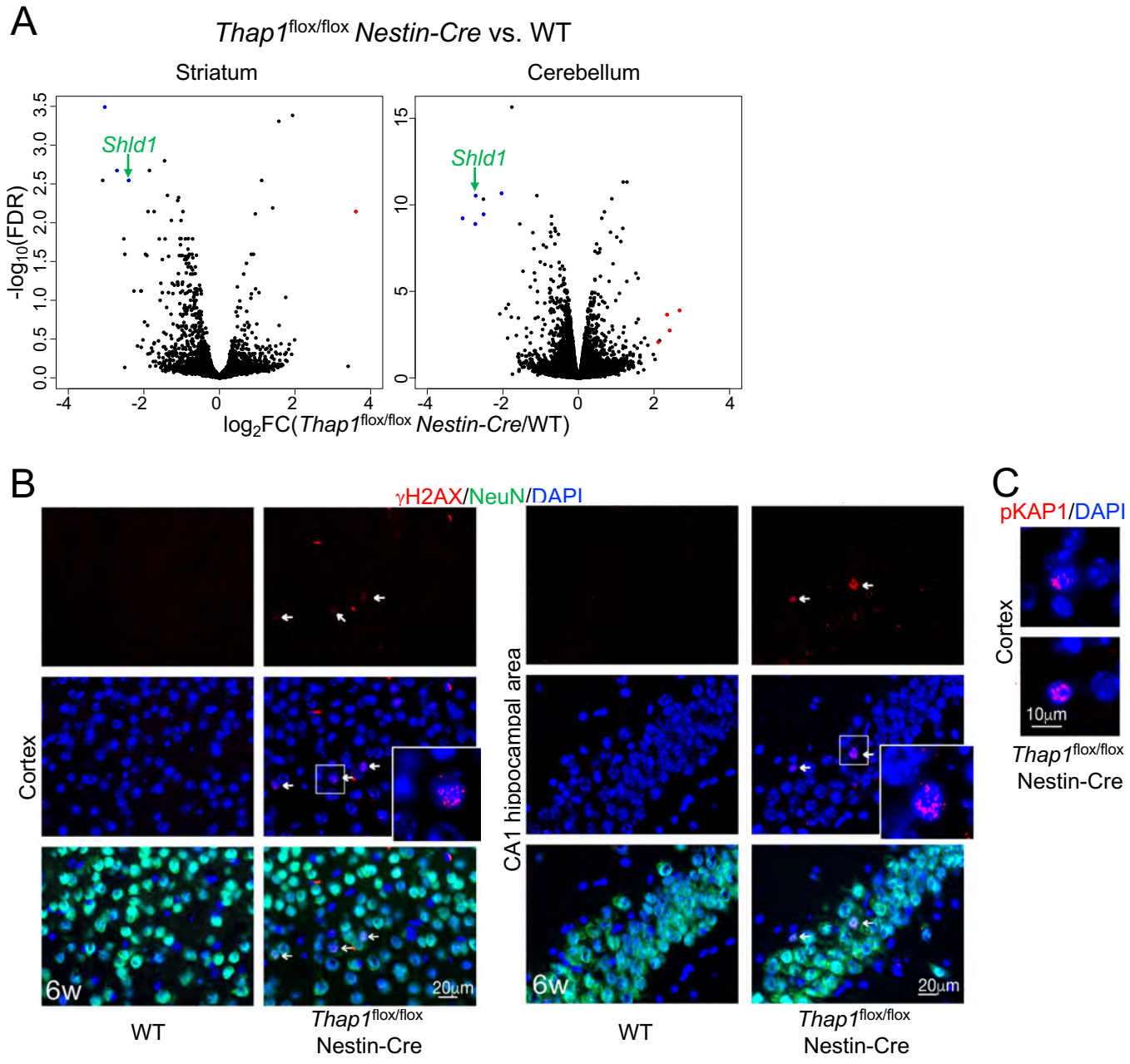

A
10 Gy IR
#1
#2
#3
GFP-SHLD3
DAPI
B
% Cells with >5 GFP-SHLD3 foci
GFP-SHLD3
100
80
60
40
20
0
#1
#2
#3
Thap1-/-Brca1Δ11 +GFP-SHLD3
C
Thap1-/-Brca1Δ11 +SHLD1
Thap1-/-Brca1Δ11 +THAP1
Dox - + - +
SHLD1 (αFlag)
THAP1 (αFlag)
TUBULIN
TUBULIN
D
Normalized viability (%)
100
10
1
0.0
0.2
0.4
0.6
PARPi (μM)
WT
Brca1Δ11
Thap1-/-Brca1Δ11 +Mock
Thap1-/-Brca1Δ11 +GFP-SHLD3 #1
Thap1-/-Brca1Δ11 +GFP-SHLD3 #2

Figure S5


A
Thap1flox/flox Nestin-Cre vs. WT
Striatum
Cerebellum
Shld1
Shld1
-log10(FDR)
log2FC(Thap1flox/flox Nestin-Cre/WT)
B
γH2AX/NeuN/DAPI
Cortex
CA1 hippocampal area
6w
20μm
WT
Thap1flox/flox
Nestin-Cre
C
pKAP1/DAPI
Cortex
10μm
Thap1flox/flox
Nestin-Cre

Repeat expansions confer WRN dependence in microsatellite-unstable cancers

https://doi.org/10.1038/s41586-020-2769-8

Received: 20 April 2020

Accepted: 16 July 2020

Published online: 30 September 2020

Check for updates

Niek van Wietmarschen[1,12], Sriram Sridharan[1,12], William J. Nathan[1,2,12], Anthony Tubbs[1,12], Edmond M. Chan[3,4,12], Elsa Callen[1], Wei Wu[1], Frida Belinky[1], Veenu Tripathi[1], Nancy Wong[1], Kyla Foster[4], Javad Noorbakhsh[4], Kiran Garimella[4], Abimael Cruz-Migoni[2], Joshua A. Sommers[5], Yongqing Huang[4], Ashir A. Borah[4], Jonathan T. Smith[4], Jeremie Kalfon[4], Nikolas Kesten[6], Kasper Fugger[7], Robert L. Walker[8], Egor Dolzhenko[9], Michael A. Eberle[9], Bruce E. Hayward[10], Karen Usdin[10], Catherine H. Freudenreich[11], Robert M. Brosh Jr[5], Stephen C. West[7], Peter J. McHugh[2], Paul S. Meltzer[8], Adam J. Bass[3,4] & André Nussenzweig[1]✉

The RecQ DNA helicase WRN is a synthetic lethal target for cancer cells with microsatellite instability (MSI), a form of genetic hypermutability that arises from impaired mismatch repair[1–4]. Depletion of WRN induces widespread DNA double-strand breaks in MSI cells, leading to cell cycle arrest and/or apoptosis. However, the mechanism by which WRN protects MSI-associated cancers from double-strand breaks remains unclear. Here we show that TA-dinucleotide repeats are highly unstable in MSI cells and undergo large-scale expansions, distinct from previously described insertion or deletion mutations of a few nucleotides[5]. Expanded TA repeats form non-B DNA secondary structures that stall replication forks, activate the ATR checkpoint kinase, and require unwinding by the WRN helicase. In the absence of WRN, the expanded TA-dinucleotide repeats are susceptible to cleavage by the MUS81 nuclease, leading to massive chromosome shattering. These findings identify a distinct biomarker that underlies the synthetic lethal dependence on WRN, and support the development of therapeutic agents that target WRN for MSI-associated cancers.

MSI is characterized by hypermutability of short repetitive DNA sequences that are scattered throughout the genome. MSI arises from deficiency in DNA mismatch repair (MMR) and contributes to the formation of many types of cancer including colorectal cancer (15%), endometrial cancer (20–30%), gastric cancers (15%), and ovarian cancers (12%)[6]. Recent studies have shown that several cancer types with MSI are reliant on activity of the WRN helicase for survival[1–4]. WRN is a member of the RecQ family of DNA helicases that includes WRN, BLM, and RECQL4, which when mutated cause the distinct chromosome instability disorders Werner syndrome, Bloom syndrome, and Rothmund–Thomson syndrome, respectively[7]. RecQ helicases do not display sequence specificity but resolve non-canonical secondary DNA structures such as bubbles, Holliday junctions, and G-quadruplexes that may be encountered during replication and recombination. However, the mechanism by which WRN helicase is required to protect chromosomal integrity of MSI, but not microsatellite stable (MSS), cancers is not understood.

MMR restoration was previously shown to only partially rescue MSI cells from WRN depletion[2]. Recent studies have reported inconsistent results on the effect of acute *MLH1* silencing to sensitize MSS cells to WRN depletion[1,3]. We therefore evaluated *WRN* dependency in human primary stomach epithelial cells (HSEC) after knockout of the MMR genes *MLH1* or *MSH2*. After 4 months of culture, *MLH1* or *MSH2* knockout cells failed to develop a dependency on *WRN* for survival (Extended Data Fig. 1a, b). These data suggest that instead of WRN loss being simply synthetic lethal with impaired MMR, a 'genomic scar' may gradually accumulate in MSI cancers that requires WRN as a structure-specific helicase.

WRN loss causes DSBs and end-resection in MSI cells

Loss of viability in MSI cells after silencing of *WRN* is associated with a decrease in the proliferation and accumulation of DNA double-strand breaks (DSBs)[1–4]. Consistently, we found that WRN depletion using a doxycycline-inducible *WRN* short hairpin RNA (shRNA) expressed in the MSI KM12 cell line[2] resulted in decreased DNA synthesis and high levels of the DSB marker, KAP1 phosphorylation at Ser824 (pKAP1)[8], predominantly in the G2/M phase of the cell cycle (Extended Data Fig. 1c–e). By contrast, expression of a non-targeting control shRNA (WRN.C911)[2] did not substantially induce DSBs in KM12 cells (Extended

[1]Laboratory of Genome Integrity, National Cancer Institute, NIH, Bethesda, MD, USA. [2]Department of Oncology, MRC Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford, UK. [3]Department of Medical Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA. [4]Broad Institute of Harvard and MIT, Cambridge, MA, USA. [5]Laboratory of Molecular Gerontology, National Institute on Aging, NIH, Baltimore, MD, USA. [6]Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Harvard Medical School, Cambridge, MA, USA. [7]DNA Recombination and Repair Laboratory, The Francis Crick Institute, London, UK. [8]Genetics Branch, National Cancer Institute, NIH, Bethesda, MD, USA. [9]Illumina Inc., San Diego, CA, USA. [10]Laboratory of Cell and Molecular Biology, National Institute of Diabetes, Digestive and Kidney Diseases, NIH, Bethesda, MD, USA. [11]Department of Biology, Tufts University, Medford, MA, USA. [12]These authors contributed equally: Niek van Wietmarschen, Sriram Sridharan, William J. Nathan, Anthony Tubbs, Edmond M. Chan.
✉e-mail: andre_nussenzweig@nih.gov

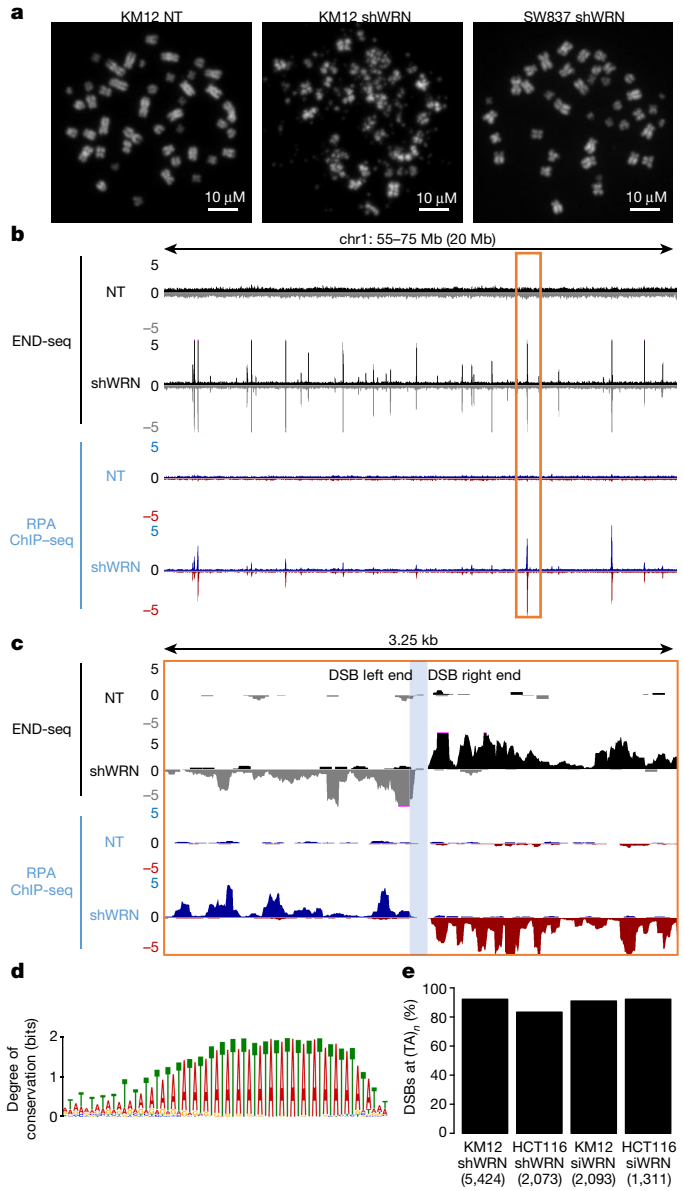


Fig. 1 | WRN depletion in MSI cells induces recurrent DSBs at $(TA)_n$ dinucleotide repeats. a, Representative metaphase spreads from KM12 and SW837 cells containing an inducible *WRN* shRNA (shWRN) cassette. Cells were treated with DMSO (NT) or doxycycline (shWRN) for 48 h. Data are representative of three independent experiments, $n = 100$ metaphases for each condition. **b**, Genome browser screenshots displaying END-seq ($n = 5$) and RPA-bound ssDNA ChIP–seq ($n = 1$) profiles as normalized read density (reads per million, RPM) for KM12-shWRN cells. Positive- and negative-strand END-seq reads are displayed in black and grey, and positive- and negative-strand RPA–ssDNA ChIP–seq reads are in blue and red, respectively. **c**, Genome browser screenshot zoomed in on the highlighted region in **b** (orange box). The light-blue shading indicates the gap region between the left and right ends of the DSB. **d**, Motif analysis for sequence enrichment in the gap between positive and negative END-seq peaks in KM12-shWRN cells. **e**, Fraction of END-seq peaks occurring at sites of $(TA)_n$ repeats for KM12-shWRN and HCT116-shWRN cells treated with either doxycycline (shWRN, $n = 2$ for HCT116) or *WRN* siRNAs (siWRN, $n = 1$) for 72 h. Numbers in parentheses indicate number of peaks identified by peak-calling.

Data Fig. 1f). Analysis of mitotic spreads ($n = 100$) showed that all chromosomes were shattered in approximately 35% of WRN-depleted KM12 cells (Figs. 1a, 2a). By contrast, chromosome shattering was not evident in the microsatellite stable (MSS) colon cancer line SW837 after WRN depletion ($n = 100$ metaphases) (Fig. 1a, Extended Data Fig. 1e).

To determine whether WRN is necessary to unwind specific regions of the genome to prevent DNA breakage, we analysed sites of recurrent DSBs by END-seq[9]. The MSI cell lines KM12 and HCT116 show little endogenous DNA breakage; however, after WRN depletion by either shRNA or small interfering RNA (siRNA), recurrent DSBs were detected at specific locations throughout the genome (Fig. 1b, Extended Data Fig. 2a). Moreover, END-seq peak intensities were highly reproducible among different experiments (Extended Data Fig. 2b). Using the criteria of 20-fold enrichment compared with non-treated cells (WRN proficient), END-seq peaks overlapped significantly between HCT116 and KM12 cells (Extended Data Fig. 2a, c). By contrast, WRN depletion in SW837 cells did not substantially induce DSBs (Extended Data Fig. 2d). Thus, the loss of WRN induces DSBs at recurrent genomic loci that are reproducible across distinct MSI cancer cell lines.

DNA breaks accumulate around $(TA)_n$ repeats

END-seq peaks displayed the characteristic pattern of positive- and negative-strand reads representing the right and left ends of DSBs, respectively (Fig. 1c). END-seq reads spread outwards from DSB sites in a pattern consistent with DNA end resection[9,10]. To confirm this, we mapped sites of single-stranded DNA (ssDNA) bound by replication protein A (RPA)[11,12]. Notably, 79% of ssDNA peaks overlapped with END-seq peaks, with resection lengths averaging 500 bp and extending up to 5 kb (Fig. 1b, c, Extended Data Fig. 2e, f). Moreover, the polarity of RPA binding was indicative of the accumulation of 3′-overhangs (Extended Data Fig. 2f). Thus, the loss of WRN in MSI cells leads to DSBs with extensive 5′–3′ end-processing.

The left- and right-end of the DSBs were separated from each other by a variable distance (Fig. 1c, Extended Data Fig. 2g). Notably, this 'gap' contained a major reduction in sequencing reads, suggesting that DSBs occur at the borders of these regions. We therefore searched for specific DNA motifs in the gap region in the hg19 reference genome, which revealed a dominant TA-dinucleotide repeat motif (Fig. 1d) with a median repeat length of 51 bp (Extended Data Fig. 2h). Nearly all breaks associated with WRN deficiency in both KM12 and HCT116 cells occurred at TA dinucleotide repeats (Fig. 1e). However, only about 8% of all $(TA)_n$ repeats in the reference genome (5,400 out of 66,644) were associated with DSBs (Extended Data Fig. 2h). We conclude that DSBs flank a fraction of $(TA)_n$ repeats. Hereafter, we refer to these sites as 'broken' $(TA)_n$ repeats, and to the DSBs themselves as 'TA breaks'.

Cruciform structures form at $(TA)_n$ repeats in plasmids in *Escherichia coli* and yeast when their length exceeds roughly 20–22 repeat units[13–15]. Long $(TA)_n$ tracts also cause stalling of replication forks and chromosome fragility at late replicating common fragile sites (CFSs)[16–18]. Exogenous replication stress further enhances replication fork collapse at simple repeats[12,19], including CFSs. Accordingly, we found that WRN depletion induces DNA breakage precisely at $(TA)_n$ repeats within several CFSs (Extended Data Fig. 3a) and at palindromic TA-rich repeats (Extended Data Fig. 3b), which have been proposed to form cruciform structures[20]. These data suggest that $(TA)_n$ repeats at these sites might fold into secondary structures that are targeted by WRN.

MUS81 shatters chromosomes in WRN-minus MSI cells

MUS81–EME1 is a structure-specific endonuclease complex that processes late recombination intermediates at CFSs[21]. MUS81–EME1 forms a complex with the scaffolding protein SLX4 that hyperactivates it at the G2/M boundary[22]. Recent studies indicate that the yeast MUS81–EME1 homologue (Mus81–Mms4) causes DSBs at $(TA)_n$ repeats when the tract exceeds a threshold length for forming cruciform structures[17]. To determine whether the DSBs that accumulate in WRN-depleted MSI cells are dependent on MUS81 and SLX4, we depleted these factors

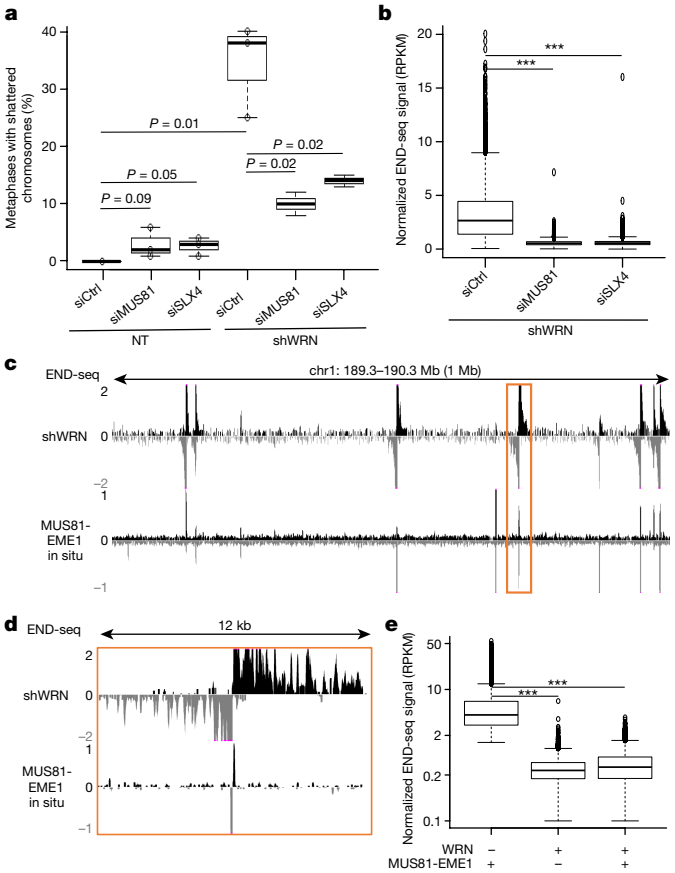


Fig. 2 | TA breaks are dependent on structure-specific endonucleases MUS81–EME1 and SLX4. a, Quantification of metaphases displaying chromosome shattering (defined as the absence of intact chromosomes) in KM12-shWRN cells treated with doxycycline plus siRNAs against *MUS81* (siMUS81) or *SLX4* (siSLX4), or non-targeting control (siCtrl) siRNAs for 72 h ($n = 3$). The top, centre mark, and bottom hinges of the box plots, respectively, indicate the 75th, median, and 25th percentile values. Whiskers represent minimum and maximum values. *P* values were calculated using Student's *t*-test. **b**, Quantification of END-seq signal intensity ($n = 2$) at broken TA repeats in KM12-shWRN cells with treatment as in **a**. $n = 5{,}001$ peaks were examined. Boxplots are as in **a**, with outliers beyond 1.5 times the interquartile range shown. $^{***}P < 2.2 \times 10^{-16}$, one-sided Wilcoxon rank sum. **c**, Genome browser screenshot for KM12-shWRN cells treated with doxycycline (shWRN, top), and DMSO-treated cells processed with purified recombinant MUS81–EME1 enzyme in situ (bottom, $n = 2$). **d**, Genome browser screenshot zoomed in on the highlighted region in **c** (orange box). **e**, Quantification of END-seq peak intensity for DMSO-treated KM12-shWRN cells processed in situ with purified recombinant MUS81–EME1, WRN or WRN followed by MUS81–EME1 ($n = 2$). For the latter, proteinase K digestion was performed between the two enzymatic treatments. $n = 5{,}496$ peaks were examined. Box plots are as in **a** and **b**. $^{***}P < 2.2 \times 10^{-16}$, one-sided Wilcoxon rank sum test.

before WRN depletion (Extended Data Fig. 4a). Depletion of MUS81 or SLX4 markedly reduced chromosome shattering (Fig. 2a, Extended Data Fig. 4b). Consistent with this result, depletion of MUS81 and SLX4 also strongly reduced pKAP1 signalling (Extended Data Fig. 4c) and the formation of DSBs at $(TA)_n$ repeats (Fig. 2b, Extended Data Fig. 4d). Thus, MUS81 and SLX4 induce toxic chromosome breakage when WRN is depleted from MSI cells.

MSI genomes are cleaved by recombinant MUS81–EME1

To test whether MUS81–EME1 acts directly on secondary structures in MSI cells, we treated agarose-embedded genomic DNA from KM12 cells with recombinant MUS81–EME1 in situ before performing END-seq. MUS81–EME1 promotes the resolution of cruciform structures by a nick and counter-nick mechanism (Extended Data Fig. 4e), and we found that MUS81–EME1 generated recurrent and reproducible DSBs (Fig. 2c, d, Extended Data Fig. 4f). These overlapped markedly with DSBs generated by WRN depletion in these cells (Extended Data Fig. 4g, h). These data show that secondary structures accumulate and can be cleaved by MUS81–EME1 even at baseline conditions in MSI cells. Unlike DSBs that were highly resected after WRN depletion, in situ cleavage by MUS81–EME1 led to accumulation of reads precisely adjacent to the border of $(TA)_n$ repeats (Fig. 2d, Extended Data Fig. 4i). Thus, the broad distribution of END-seq reads in the WRN-depleted samples is indeed caused by 5′ end-processing in vivo after MUS81 cleavage.

We hypothesized that WRN is recruited to unwind DNA secondary structures before they can be cleaved by physiologically active MUS81. To test this, we incubated agarose-embedded DNA from KM12 cells with recombinant human WRN before MUS81–EME1 treatment in situ, followed by END-seq detection of DSBs. Incubation with WRN alone did not result in DSB formation, but pre-treatment with WRN substantially decreased cleavage by MUS81–EME1 (Fig. 2e, Extended Data Fig. 4j). These results suggest that WRN melts secondary DNA structures at $(TA)_n$ repeats in MSI cells.

If structure-forming $(TA)_n$ repeats are responsible for massive breakage in MSI lines, there should be fewer such structures in MSS cells. To test this, we performed the in situ MUS81–EME1 cleavage assay in two MSS cell lines (SW837 and RPE-1) and compared these to MSI cell lines KM12 and HCT116. MSI cells displayed an overlapping set of strong MUS81–EME1 cleavage sites (Extended Data Fig. 5a, b); by contrast, manifold fewer substrates for MUS81–EME1 cleavage were detected in the MSS cells. Thus, structure-forming $(TA)_n$ repeats accumulate in much higher abundance in MSI than in MSS cells.

$(TA)_n$ repeats are prone to replication fork stalling

DNA polymerase stalling generates RPA-bound ssDNA, which activates the checkpoint kinase ATR to protect the replication fork. RPA chromatin immunoprecipitation followed by high-throughput sequencing (ChIP–seq) analysis in WRN-proficient KM12 cells demonstrated an enrichment of RPA in the vicinity of broken $(TA)_n$ sites, which suggests that stalled forks spontaneously accumulate at these sites (Fig. 3a). WRN is recruited to stalled replication forks in a manner that requires ATR phosphorylation[23]. Consistent with this, HCT116 cells expressing the WRN mutants WRN(3A) or WRN(6A) that contain alanine substitutions at identified ATR phosphorylation sites[23] showed an increase in KAP1 phosphorylation when endogenous WRN was ablated by an siRNA that targeted the 5′ untranslated region (Extended Data Fig. 5c). We therefore hypothesized that stalled forks at structure-forming $(TA)_n$ repeats would be susceptible to fork collapse after inhibition of the ATR kinase. To test this, we treated MSI and MSS cells with a combination of ATR inhibitor and low-dose aphidicolin to partially inhibit DNA polymerase elongation. As determined by END-seq analysis, replication forks collapsed into DSBs preferentially at $(TA)_n$ repeats, including those within CFSs (Fig. 3b, c, Extended Data Fig. 5d). The frequency of fork collapse at $(TA)_n$ repeats in MSI cells was at least 30-fold higher than in MSS cells (Fig. 3b, c), and these sites largely overlapped with DSBs generated in the absence of WRN (Extended Data Fig. 5e). Thus, secondary structure-forming $(TA)_n$ repeats are associated with replication fork stalling that is overcome by WRN, possibly through activation by ATR.

$(TA)_n$ repeats inhibit DNA synthesis in vitro

In cancers associated with MSI, insertions or deletions of a few nucleotides are commonly found at mononucleotide repeats[5]. Among dinucleotides, $(TA)_n$ repeats are reported to be the least frequently altered[24] (Extended Data Fig. 6a). To determine why $(TA)_n$ repeats are susceptible

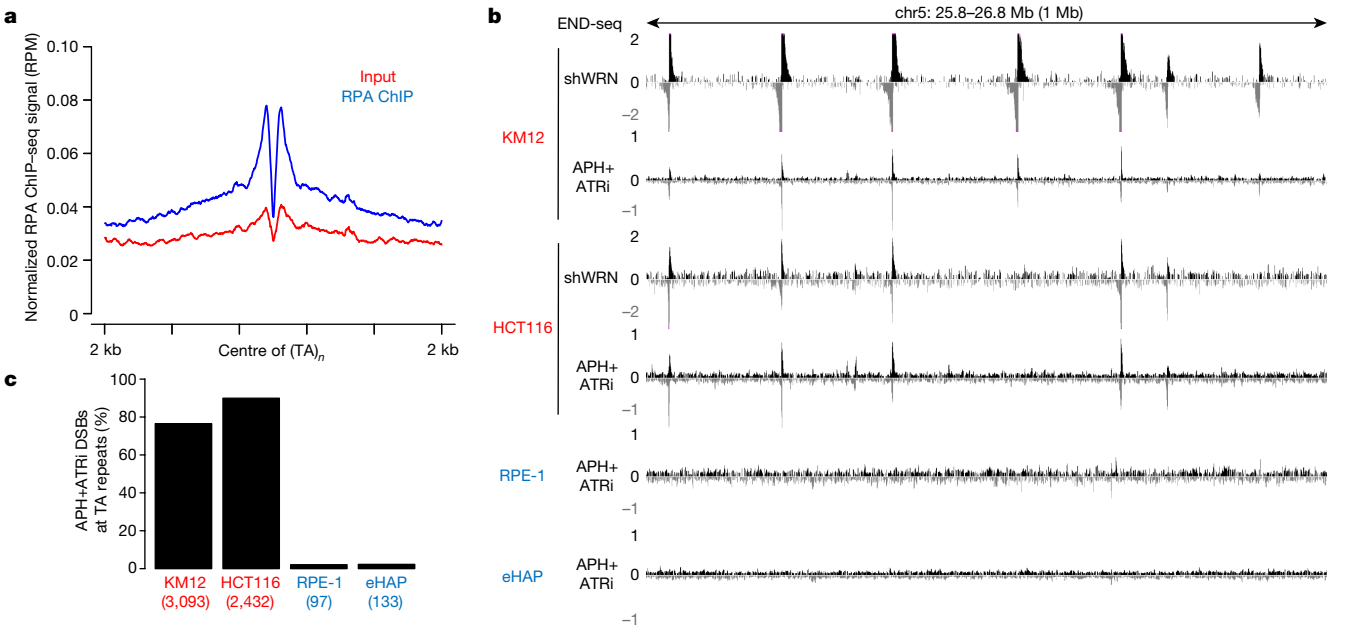


Fig. 3 | Replication stalling and collapse at $(TA)_n$ repeats in MSI cell lines. **a**, Composite plot of RPA-bound ssDNA ChIP–seq signal around broken $(TA)_n$ repeats in KM12 cells treated with DMSO. **b**, Genome browser screenshot displaying END-seq profiles for KM12, HCT116, RPE-1, and eHAP cells containing an inducible shWRN cassette. Cells were treated with doxycycline (shWRN) for 72 h, or with a combination of aphidicolin (APH) and ATR inhibitor (ATRi) ($n = 1$) for 8 h. MSI cells are in red; MSS cell lines are in blue. **c**, Percentage of total DSBs located at $(TA)_n$ repeats after treatment with APH plus ATRi, as shown in **b**. MSI cell are in red; MSS cell lines are in blue.

to breakage, we performed a PCR-based size analysis of several $(TA)_n$ repeats in a panel of MSI and MSS cell lines (Fig. 4a, Extended Data Fig. 6b, Supplementary Table 1). We examined eight $(TA)_n$ repeats that were broken and three that were not broken after WRN depletion in MSI cells. All 11 sites that we analysed contained uninterrupted stretches of more than 20 $(TA)_n$ units (in the hg19 reference genome), and were amplified with primers flanking the $(TA)_n$ repeat.

Although MSS cell lines showed the predicted PCR products at broken and non-broken sites (Fig. 4a, Extended Data Fig. 6b), many recurrently broken sites were not amplified in KM12 cells (Fig. 4a). These same sites frequently could not be amplified using genomic DNA from other MSI cell lines, which highlights that identical sites are affected across distinct cancers (Extended Data Fig. 6b). A notable exception was HCT116 cells, which showed the expected banding pattern. However, as discussed below, this probably reflects the fact that one of the two alleles is of normal size and can therefore be amplified.

The absence of PCR products at several broken $(TA)_n$ repeats using genomic DNA from MSI cell lines was unlikely to be caused by large deletions that included primer-binding sites, as these primer-binding sites were highly covered in END-seq and whole-genome sequencing data (below). We therefore speculated that these sites contain large, structure-forming expansions that might inhibit polymerase extension during PCR amplification. To test this hypothesis, we performed PCR-free whole-genome sequencing on KM12 and HCT116 cells, enabling us to reach an average sequencing depth of greater than 100×. By visual inspection, it was immediately clear that broken $(TA)_n$ repeats had a markedly lower read depth than flanking regions and non-broken $(TA)_n$ repeats (Extended Data Fig. 6c). Broken $(TA)_n$ repeats (mapped in WRN-depleted MSI cells) displayed the most significant drop in coverage compared with other classes of mono- and di-nucleotide repeats (Fig. 4b, Extended Data Fig. 6d). Finally, as assessed by CRISPR–Cas9 and RNA interference fitness screens[2], the degree of dependence on WRN for the survival of human cancer cells was inversely correlated with sequencing coverage at broken $(TA)_n$ sites (Fig. 4c). Thus, structure-forming $(TA)_n$ repeats are highly recalcitrant to sequencing. Although not previously documented as mutated in MSI-associated cancers, $(TA)_n$ repeats are nevertheless predictive of WRN dependency.

MSI cells accumulate large $(TA)_n$ repeat expansions

Because the identification of large-scale changes in short tandem repeats is challenging with short-read sequencing, we used the ExpansionHunter algorithm[25], which has been used to detect various large pathogenic repeat expansions. This analysis showed that broken $(TA)_n$ repeats exhibit large-scale expansions compared to non-broken $(TA)_n$ repeats (Extended Data Fig. 7a, b). Consistent results were obtained with the expanded short tandem repeat algorithm (exSTRa)[26] (Extended Data Fig. 7c).

As large expansions should increase the size of $(TA)_n$ repeats relative to the reference genome, we performed Southern blotting with probes that spanned two broken (B1 and B2) and two non-broken (NB1 and NB3) sites previously analysed by PCR (Fig. 4d, Extended Data Fig. 7d, e). For this analysis, we compared two MSI (KM12 and HCT116) and two MSS (RPE-1 and eHAP) cell lines. At the non-broken $(TA)_n$ repeats, we detected bands that corresponded to the expected sizes in all cell lines (Extended Data Fig. 7d). In addition, neither of the broken $(TA)_n$ regions was altered in control RPE-1 or eHAP cell lines (Extended Data Fig. 7d). For the B1 broken site, we detected one allele of the expected size, but also one expanded allele in both KM12 and HCT116 cells (Fig. 4d, Extended Data Fig. 6b). The shorter, non-expanded allele probably corresponds to the PCR product detected at B1 in both cell lines. At the B2 broken site, we detected one expanded allele in HCT116, whereas the other allele was the expected size (Fig. 4d), corresponding to the product detected by PCR (Extended Data Fig. 6b). In KM12 cells, broken site B2 exhibited distinct expansions on the two alleles (Fig. 4d, Extended Data Fig. 7e), and no PCR product was detectable at this site (Fig. 4a). Together, our data suggest that $(TA)_n$ repeats undergo large-scale expansion, and that expanded alleles inhibit DNA synthesis during PCR and short read whole-genome sequencing.

Long-read sequencing technology has successfully characterized expansions of simple tandem repeats that are responsible for various diseases[27]. To determine whether a similar approach could uncover variations in secondary structure-forming $(TA)_n$ repeats in MSI cells,

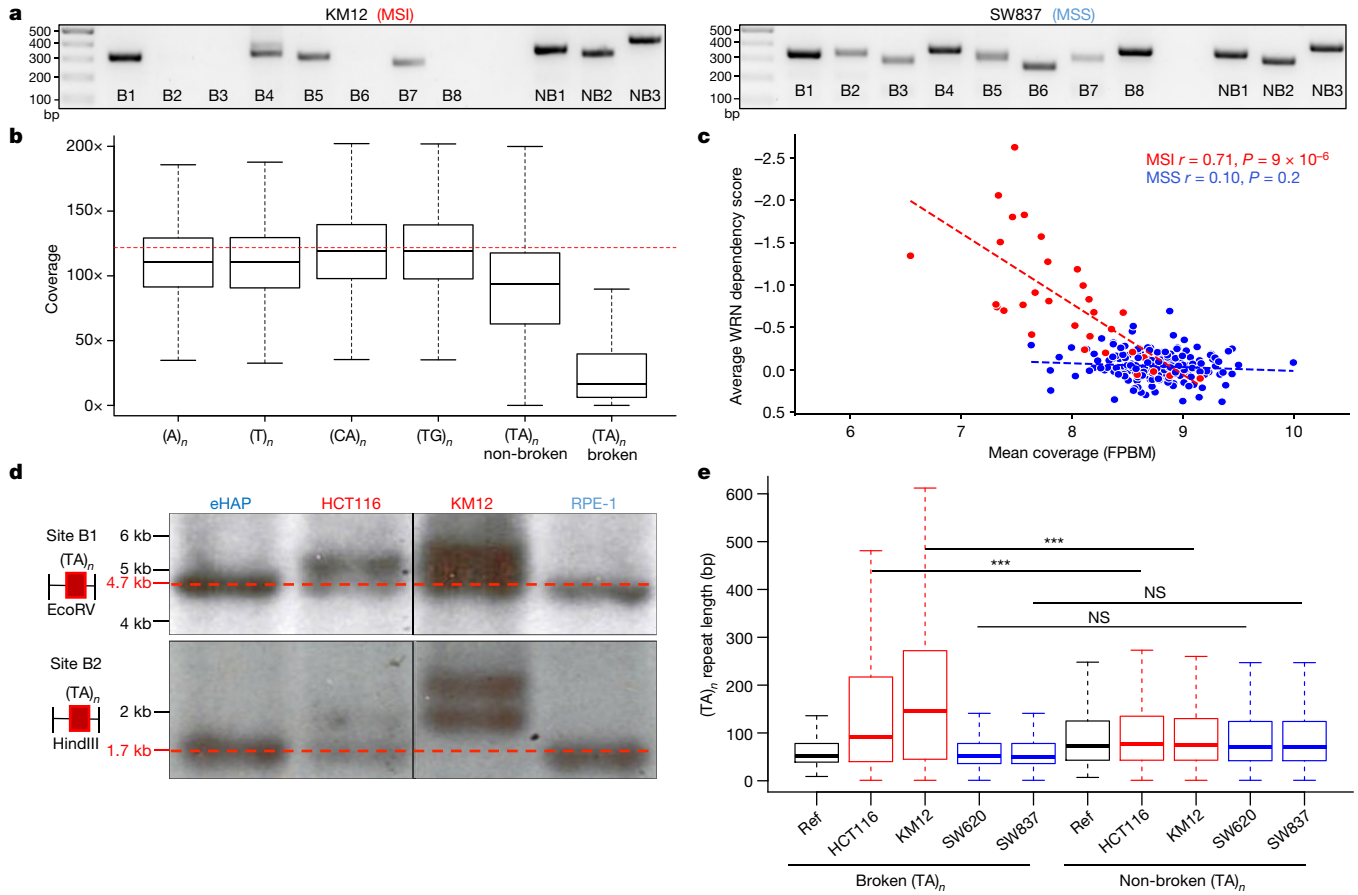


Fig. 4 | $(TA)_n$ repeats undergo large-scale expansion in MSI cell lines. **a**, Agarose gels showing PCR fragments (or lack thereof) at $(TA)_n$ repeats in KM12 (MSI) and SW837 (MSS) cells. B1–B8 were chosen based on the presence of END-seq peaks after WRN depletion in KM12 cells. NB1–NB3 were chosen for similar $(TA)_n$ repeat lengths as broken sites without breakage after WRN depletion in KM12 cells. Ladder fragment sizes (in bp) are displayed. Data are representative of three independent experiments. For gel source data, see Supplementary Fig. 1. **b**, Box plots displaying coverage at different classes of repeats in WGS data from KM12 cells ($n = 1$). Dotted red lines indicate the average coverage over the genome. **c**, Cell lines plotted by their average WRN dependency score and sequencing coverage of broken $(TA)_n$ loci. FPBM, fragments per base per million. Pearson correlation coefficients were used to calculate r values and two-sided P values. **d**, Southern blots ($n = 2$) for two different genomic regions containing broken $(TA)_n$ repeats corresponding to the same sites in **a**. Red markers and dotted lines represent expected fragment sizes. For gel source data, see Supplementary Fig. 1. **e**, Box plot of long-read sequencing data ($n = 1$) demonstrating total length of broken and non-broken $(TA)_n$ in indicated cell lines compared with the hg19 reference genome. MSI (red), MSS (blue). $n = 5{,}400$ (broken), $n = 61{,}244$ (non-broken). Box plots are as in Fig. 2a, b. $^{***}P < 2.2 \times 10^{-16}$, one-sided Wilcoxon rank sum test. NS, not significant. Wilcox testing for the alternative hypothesis that the broken $(TA)_n$ has a greater group mean than the non-broken $(TA)_n$ was used. P values were corrected for multiple testing using the Benjamini–Hochberg method.

we used the Pacific Biosciences (PacBio) continuous long-read (CLR) sequencing platform[27]. With this platform, MSI (HCT116 and KM12) and MSS (SW620 and SW837) colon cancer cell lines displayed similar coverage at broken $(TA)_n$ and other simple repeats (Extended Data Fig. 7f). We detected a large range of expansions at different broken $(TA)_n$ sites in MSI cells, with median repeat lengths expanding from 54 bp in the hg19 reference genome to 91 bp and 125 bp in HCT116 and KM12 cells, respectively (Fig. 4e). By contrast, the median lengths of non-broken $(TA)_n$ repeats in HCT116 (75 bp) and KM12 (74 bp) cells were much more similar to the reference genome (72 bp). MSS cell lines did not show substantial expansions at either broken or non-broken $(TA)_n$ sites (Fig. 4e). Motif analysis within the broken sites showed that the expanded alleles consisted of almost pure $(TA)_n$ repeats (Extended Data Fig. 7g). Thus, MMR deficiency predisposes $(TA)_n$ repeat tracts to undergo large-scale expansions.

$(TA)_n$ repeat length and purity contribute to breakage

Of all the annotated $(TA)_n$ repeats in the reference genome, only 8% display DSBs after WRN depletion in MSI cells (Extended Data Fig. 2h). Thus, it remains unclear what characterizes the underlying expansion and breakage specifically at these sites and why susceptible sites are recurrent across several distinct MSI cell lines. Changes in the length of microsatellite DNA are thought to arise from replication slippage caused by the transient dissociation of the polymerase from replicating DNA strands followed by self-annealing and misaligned reassociation[28]. Because the probability of self-annealing increases with the purity and length of the repeat sequence and decreases with repeat interruptions[28], we speculated that predisposition to large-scale expansions of $(TA)_n$ repeats, and therefore breakage, would be influenced by the exact sequence composition of the repeat.

We found that broken sites have a higher TA content, fewer interruptions in the $(TA)_n$ repeat, and longer uninterrupted $(TA)_n$ sequences compared with non-broken sites as assessed both in the reference genome (Extended Data Fig. 8a–c) and in long-read sequencing reads from MSI cells (Extended Data Fig. 8d–f). Quantitative modelling showed that three features of $(TA)_n$ repeats were predictive of breakage in WRN-deficient cells: the probability that they form secondary structures as measured by MUS81–EME1 cleavage in situ; the sizes of expansions determined by long-read

sequencing; and the likelihood that they occur in late replicating regions (Extended Data Fig. 8g–i). Although these features cannot fully predict the propensity of $(TA)_n$ repeats to break after WRN deficiency, the analysis demonstrates that longer, uninterrupted $(TA)_n$ repeats are more likely to undergo expansion in MSI cell lines, where they form secondary structures and disrupt DNA replication. In the absence of WRN, $(TA)_n$ repeats that are replicated in late S phase are less likely to be resolved before mitosis, when they are recognized and cleaved by MUS81 to generate DSBs.

Expandable repeats and genome stability

Secondary structure-forming $(TA)_n$ tracts within CFSs cause replication fork stalling and chromosome breakage. CFSs are also associated with deletions in several tumour types[29], probably owing to their susceptibility to replication fork collapse. Because broken $(TA)_n$ sites in MSI cells are also susceptible to replication fork stalling and collapse (Fig. 3), we hypothesized they could be hotspots for deletions in MSI-associated cancers. Although genomic instability in MSI cancers is most frequently associated with small insertions and deletions, large kilobase-to-megabase scale deletions of unknown aetiology have been detected[30]. We therefore analysed data from pan-cancer whole genome sequencing cohorts in uterine corpus endometrial carcinoma (UCEC), colon adenocarcinoma (COAD) and gastric adenocarcinoma (STAD)[31]. Breakpoints associated with deletions (ranging in size from 459 bp to 176 Mb) in MSI (24) and MSS (93) cancers were identified[5]. In several cases, one or both deletion breakpoints mapped precisely to broken $(TA)_n$ repeats (Extended Data Fig. 9a, b). We then calculated the enrichment of different annotated repeats at the tumour breakpoints compared to their enrichment at random breakpoints with comparable chromosome and size distributions, and found that MSI tumour breakpoints showed the greatest enrichment at broken $(TA)_n$ repeats (Extended Data Fig. 9c). By contrast, these sites were not enriched for deletion breakpoints in MSS tumours (Extended Data Fig. 9c). Thus, in cancers associated with MSI, large-scale deletions are frequently associated with broken $(TA)_n$ sites, which suggest that they are inherently fragile.

On the basis of our results, we propose the following model (Extended Data Fig. 8j): MMR-deficient cells undergo microsatellite instability, which gradually manifests as large-scale expansions at $(TA)_n$ repeats. The most susceptible sites are those that contain pure, uninterrupted TA-dinucleotide repeats. Over the course of months to years, $(TA)_n$ repeats reach a threshold length above which they extrude into cruciform structures, perhaps as a result of negative supercoiling during nucleosome removal ahead of the replication fork. These structures would stall replication forks and trigger ATR-dependent WRN phosphorylation, which promotes unwinding of the secondary structure to complete replication. In the absence of WRN, the structure-specific MUS81–EME1 endonuclease, in conjunction with its scaffold SLX4, cleave these structures in an attempt to salvage the replication fork. However, thousands of concerted MUS81 cleavage events lead to extensive DNA end-resection, RPA exhaustion[8] (Fig. 1b, c), chromosomal fragmentation, and cell death.

Thirty years ago, Vogelstein and colleagues discovered the *DCC* gene, which has vastly reduced expression in colon cancer[32]. By Southern blot analysis, several MSI tumour cell lines were shown to contain 'insertions' of up to 300 bp at a locus containing an uninterrupted $(TA)_{22}$ repeat just downstream of *DCC* exon 7. However, numerous attempts to clone or amplify alleles with insertions failed, leading the authors to conclude that the inserted sequence might form an unusual DNA structure. Notably, we detected a MUS81–EME1-sensitive DNA structure at precisely the same $(TA)_n$ repeat (Extended Data Fig. 10a), and long-read sequencing confirmed a $(TA)_n$ expansion at that locus (Extended Data Fig. 10b, right highlighted sequence). Thus, our data demonstrate that rather than a non-templated insertion, DCC contains a structure-forming, MSI-expanded $(TA)_n$ repeat.

Anti-PD-1 antibodies have been approved for use in patients with cancers with mismatch repair deficiency or MSI, independent of the cancer lineage. The therapeutic response is correlated with the number of indels in coding regions, which can generate immunogenic neoantigens[33,34]. $(TA)_n$ repeat expansions are mostly localized outside of coding regions, and therefore might not generate immunogenic neoantigens. However, DSBs at $(TA)_n$ motifs can trigger innate cytosolic DNA- and RNA-dependent sensing and signalling pathways[35,36], which indicates that WRN inhibition has the potential for further stimulating immune responses. Although indel mutations of a few nucleotides and large-scale $(TA)_n$ expansions are both features of MSI, further studies will be necessary to determine whether these defects arise through different mechanisms, in distinct MSI tumours, or within different subclones. In summary, our findings provide a mechanistic explanation for WRN dependence with MSI and identify a new biomarker to guide the selection of patients in which WRN inhibition may be effective in combination with immune checkpoint blockade or as an independent line of therapy.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2769-8.

1. Behan, F. M. et al. Prioritization of cancer therapeutic targets using CRISPR–Cas9 screens. *Nature* **568**, 511–516 (2019).
2. Chan, E. M. et al. WRN helicase is a synthetic lethal target in microsatellite unstable cancers. *Nature* **568**, 551–556 (2019).
3. Kategaya, L., Perumal, S. K., Hager, J. H. & Belmont, L. D. Werner syndrome helicase is required for the survival of cancer cells with microsatellite instability. *iScience* **13**, 488–497 (2019).
4. Lieb, S. et al. Werner syndrome helicase is a selective vulnerability of microsatellite instability-high tumor cells. *eLife* **8**, e43333 (2019).
5. Fujimoto, A. et al. Comprehensive analysis of indels in whole-genome microsatellite regions and microsatellite instability across 21 cancer types. *Genome Res*. (2020).
6. Dudley, J. C., Lin, M. T., Le, D. T. & Eshleman, J. R. Microsatellite instability as a biomarker for PD-1 blockade. *Clin. Cancer Res*. **22**, 813–820 (2016).
7. Chu, W. K. & Hickson, I. D. RecQ helicases: multifunctional genome caretakers. *Nat. Rev. Cancer* **9**, 644–654 (2009).
8. Toledo, L. I. et al. ATR prohibits replication catastrophe by preventing global exhaustion of RPA. *Cell* **155**, 1088–1103 (2013).
9. Canela, A. et al. DNA breaks and end resection measured genome-wide by end sequencing. *Mol. Cell* **63**, 898–911 (2016).
10. Paiano, J. et al. ATM and PRDM9 regulate SPO11-bound recombination intermediates during meiosis. *Nat. Commun*. **11**, 857 (2020).
11. Khil, P. P., Smagulova, F., Brick, K. M., Camerini-Otero, R. D. & Petukhova, G. V. Sensitive mapping of recombination hotspots using sequencing-based detection of ssDNA. *Genome Res*. **22**, 957–965 (2012).
12. Tubbs, A. et al. Dual roles of Poly(dA:dT) tracts in replication initiation and fork collapse. *Cell* **174**, 1127–1142 (2018).
13. Bowater, R., Aboul-ela, F. & Lilley, D. M. Large-scale stable opening of supercoiled DNA in response to temperature and supercoiling in (A + T)-rich regions that promote low-salt cruciform extrusion. *Biochemistry* **30**, 11495–11506 (1991).
14. Dayn, A. et al. Formation of (dA-dT)n cruciforms in *Escherichia coli* cells under different environmental conditions. *J. Bacteriol*. **173**, 2658–2664 (1991).
15. McClellan, J. A., Boublíková, P., Palecek, E. & Lilley, D. M. Superhelical torsion in cellular DNA responds directly to environmental and genetic factors. *Proc. Natl Acad. Sci. USA* **87**, 8373–8377 (1990).
16. Zlotorynski, E. et al. Molecular basis for expression of common and rare fragile sites. *Mol. Cell. Biol*. **23**, 7143–7151 (2003).
17. Kaushal, S. et al. Sequence and nuclease requirements for breakage and healing of a structure-forming (AT)n sequence within fragile site FRA16D. *Cell Rep*. **27**, 1151–1164 (2019).
18. Wang, H. et al. CtIP maintains stability at common fragile sites and inverted repeats by end resection-independent endonuclease activity. *Mol. Cell* **54**, 1012–1021 (2014).
19. Shastri, N. et al. Genome-wide identification of structure-forming repeats as principal sites of fork collapse upon ATR inhibition. *Mol. Cell* **72**, 222–238 (2018).
20. Inagaki, H. et al. Chromosomal instability mediated by non-B DNA: cruciform conformation and not DNA sequence is responsible for recurrent translocation in humans. *Genome Res*. **19**, 191–198 (2009).
21. Minocherhomji, S. & Hickson, I. D. Structure-specific endonucleases: guardians of fragile site stability. *Trends Cell Biol*. **24**, 321–327 (2014).
22. Wyatt, H. D., Laister, R. C., Martin, S. R., Arrowsmith, C. H. & West, S. C. The SMX DNA repair tri-nuclease. *Mol. Cell* **65**, 848–860 (2017).

23. Ammazzalorso, F., Pirzio, L. M., Bignami, M., Franchitto, A. & Pichierri, P. ATR and ATM differently regulate WRN to prevent DSBs at stalled replication forks and promote replication fork recovery. *EMBO J.* **29**, 3156–3169 (2010).
24. Cortes-Ciriano, I., Lee, S., Park, W. Y., Kim, T. M. & Park, P. J. A molecular portrait of microsatellite instability across multiple cancers. *Nat. Commun.* **8**, 15180 (2017).
25. Dolzhenko, E. et al. Detection of long repeat expansions from PCR-free whole-genome sequence data. *Genome Res.* **27**, 1895–1903 (2017).
26. Tankard, R. M. et al. Detecting expansions of tandem repeats in cohorts sequenced with short-read sequencing data. *Am. J. Hum. Genet.* **103**, 858–873 (2018).
27. Mitsuhashi, S. & Matsumoto, N. Long-read sequencing for rare human genetic diseases. *J. Hum. Genet.* **65**, 11–19 (2020).
28. Khristich, A. N. & Mirkin, S. M. On the wrong DNA track: molecular mechanisms of repeat-mediated genome instability. *J. Biol. Chem.* **295**, 4134–4170 (2020).
29. Glover, T. W., Wilson, T. E. & Arlt, M. F. Fragile sites in cancer: more than meets the eye. *Nat. Rev. Cancer* **17**, 489–501 (2017).
30. The Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337 (2012).
31. The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. *Nature* **578**, 82–93 (2020).
32. Fearon, E. R. et al. Identification of a chromosome 18q gene that is altered in colorectal cancers. *Science* **247**, 49–56 (1990).
33. Ding, L. & Chen, F. Predicting tumor response to PD-1 blockade. *N. Engl. J. Med.* **381**, 477–479 (2019).
34. Mandal, R. et al. Genetic diversity of tumors with mismatch repair deficiency influences anti-PD-1 immunotherapy response. *Science* **364**, 485–491 (2019).
35. Feng, X. et al. ATR inhibition potentiates ionizing radiation-induced interferon response via cytosolic nucleic acid-sensing pathways. *EMBO J.* **39**, e104036 (2020).
36. Harding, S. M. et al. Mitotic progression following DNA damage enables pattern recognition within micronuclei. *Nature* **548**, 466–470 (2017).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2020

Methods

Cell lines and cell culture

Cell lines containing doxycycline-inducible shWRN cassette (KM12, HCT116, SW837, RPE-1) were generated as previously described[2]. Cell lines were grown in medium supplemented with 10% fetal bovine serum (FBS), penicillin (100 μg ml^{-1}), streptomycin (100 μg ml^{-1}), and L-glutamine (Gibco, 292 μg ml^{-1}) unless stated otherwise. KM12, SW837, and SW48 cells were grown in RPMI1640 (Gibco), HCT116 in McCoy's 5A (Gibco), RPE-1 in DMEM/F12 (Gibco), SW620 in Leibovitz's L-15 (Gibco), OVK18 and LS180 in MEMα (Gibco) supplemented with 15% FBS. HSECs were obtained from Cell Biologics (H-6039) and cultured with complete human epithelial cell medium (CellBiologics H6621). Independent HCT116 clones were used for whole genome sequencing and continuous long-read sequencing. Cell lines were tested for mycoplama contamination.

Generation of MMR-deficient HSEC cell lines

Transduction of HSEC with Cas9 was performed with pLX311-Cas9 (Addgene 118018) followed by blasticidin (4 μg ml^{-1}) selection. Subsequently, we transduced pXPR_BRD003 carrying sgRNAs targeting *MLH1* (target sequence TTTGGCCAGCATAAGCCATG), *MSH2* (CCGGTCGAAAAGGCGCACTG), or luciferase (ACAACTTTACCGA CCGCGCC) to generate stable cell lines after puromycin selection (2 μg ml^{-1}).

Cell viability assay

Cas9-expressing HSEC with sgRNAs targeting *MLH1*, *MSH2*, or luciferase were transduced with pXPR_BRD003 harbouring the following sgRNAs. sgRNAs targeting *WRN* include sg*WRN*2 (ATCCT-GTGGAACATACCATG) and sg*WRN*3 (GTAGCAGTAAGTGCAACGAT). Two negative controls targeting intergenic sites on chromosome 2 were used: Chr.2-2 sgRNA (GGTGTGCGTATGAAGCAGTG) and Chr.2-4 sgRNA (GCAGTGCTAACCTTGCATTG). sgRNA targeting *POLR2D* (AGAGACTGCTGAGGAGTCCA) was used as a pan-essential control. All sgRNAs were inserted in the pXPR_BRD003 lentiviral vector and inserts were verified by Sanger sequencing. Viability was determined using CellTiter-Glo (Promega G7572) 7 days after lentiviral transduction.

Protein depletion and exogenous replication stress

Expression of shWRN was induced by adding 1 μg ml^{-1} doxycycline to cell culture medium for indicated time. The following siRNAs were used for experiments: siGenome Human siRNA Smartpool targeting WRN, MUS81, and SLX4, as well as non-targeting control pool (Horizon Discovery). WRN 5′ UTR (AAACCCGAGAAGAUCCAGUCCAACA) was described previously[3] and ordered from ThermoFisher. Cells were transfected using RNAiMax Transfection Reagent (ThermoFisher) according to manufacturer's instructions. To induce exogenous replication fork collapse, cells were treated with aphidicolin (0.2 μg ml^{-1}, Sigma Aldrich) for 24 h, and ATR inhibitor AZ20 (10μM, SelleckChem) was added for the final 8 h.

RNA isolation and qRT–PCR

Total RNA was extracted from cells using Trizol Reagent (Invitrogen), and cDNA was made using SuperScript II Reverse Transcriptase (ThermoFisher), according to manufacturer's instructions. qPCR was performed using iTaq Universal SyBR Green (BioRad), samples were run and analysed on a BioRad CFX96 Real-Time PCR detection system. Primer sequences were as follows.

MUS81 forward: 5′ GCTGCTCCGAGAGCTACAG 3′ MUS81 reverse: 5′ CAGGGTTTGCTGGGTCTCTA 3′ *SLX4* forward: 5′AGTGTGCTG TGAAGATGGAG 3′ SLX4 reverse: 5′ CCGTTTCAGACCTCTACTGTG 3′ *ACTB* forward: 5′ CGTCACCAACTGGGACGACA 3′ *ACTB* reverse: 5′ CTTCTCGCGGTTGGCCTTGG 3′

Metaphase analysis

Cells were arrested at mitosis with 0.04 μg ml^{-1} colcemid (Roche) for 16 h and metaphase chromosome spreads were prepared as previously described[37].

Western blotting

Cells were lysed in a buffer containing 50 mM Tris-HCl (pH 7.5), 200 mM NaCl, 5% Tween-20, 2% Igepal CA-630, 2 mM PMSF, 50 mM β-glycerophosphate (Merck) and protease inhibitor cocktail tablet (cOmplete Mini, Roche Diagnostics). Equal amounts of lysates were loaded into precast mini-gels (Invitrogen) and resolved by SDS–PAGE. Transfer of proteins onto nitrocellulose membranes and incubation with primary/secondary antibodies were performed according to standard procedures. Visualization of protein bands was achieved by fluorescence imaging on the Odyssey Clx system (LI-COR Biosciences). Antibodies and dilutions used were as follows: anti-WRN (1:5,000, Novus), anti-pKAP1 (Bethyl Laboratories, 1:100), anti-tubulin (1:5,000, sigma), anti-MLH1 (Cell Signaling Technology, 3515, 1:1,000), anti-MSH2 (Abcam, ab52266, 1:1,000), anti-GAPDH (Cell Signaling Technology, 5174, 1:5,000), IRDye 680RD Goat anti-Mouse IgG (1:2,000, LI-COR Biosciences), IRDye 800CW Goat anti-Rabbit IgG (1:2,000, LI-COR Biosciences), Goat anti-Rabbit Alexa Fluor 647 (ThermoFisher, A27040, 1:5,000).

Flow cytometry

For cell cycle analysis, exponentially growing cells were incubated with 10 μM (5-ethynyl-2′-deoxyuridine) for 30 min at 37 °C and stained using the Click-IT EdU Alexa Fluor 488 Flow Cytometry Assay Kit (ThermoFisher) according to the manufacturer's instructions. DNA content was measured by DAPI (4′,6-diamidino-2-phenylindole, 0.5 μg ml^{-1}). For pKAP1 staining, fixed and permeabilized cells were washed in PBS plus 2% FBS, incubated for 1 h with pKAP1 antibody (Bethyl Laboratories, 1:100), washed, stained with anti-rabbit Alexa Fluor 647 antibody (ThermoFisher, 1:2,000). Data were analysed using FlowJo v10 software.

Protein purification

Expression and purification of MUS81–$_{FLAG}$EME1[22] and WRN[38] were performed as previously described. Human MUS81 (residues 246–551) and double 8×His-EME1 (residues 246–570) were expressed in a bicistronic expression vector and purified as previously described[39], minus the cation exchange chromatography step.

Enzymatic reactions

For enzymatic reactions, END-seq plugs were incubated in the presence of 50 nM of the indicated enzyme for 1.5 h at 37 °C. MUS81-EME1 enzyme reactions were done in buffer consisting of 25 mM Tris-HCl pH 8.0, 30 mM NaCl, 3 mM $MgCl_2$, 100 ng μl^{-1} BSA, 5% glycerol, and 1 mM DTT. WRN enzyme reactions were done in buffer consisting of 30 mM HEPES pH 7.4, 40 mM KCl, 8 mM $MgCl_2$, 100 ng μl^{-1} BSA, 5% glycerol and 2 mM ATP. For in situ experiments in which plugs were treated with both WRN and MUS81–EME1, after WRN incubation, agarose plugs were treated with proteinase K in lysis buffer at 50 °C for 1 h and then washed in 10 mM Tris-HCl pH 8.0 solution, before incubation with MUS81–EME1.

END-seq

For END-seq, 7–8 million cells were collected, embedded in 1% agarose plugs, and processed as previously reported[9,12].

RPA ChIP–seq

Approximately 20 million cells were collected and ChIP–seq was performed as previously reported using anti-RPA32/RPA2 (Abcam, ab10359)[12].

PCR-free whole-genome sequencing

Genomic DNA was isolated from cultured cells by phenol–chloroform extraction. To provide sufficient mass and maximize library diversity, two identical library preparations were performed in parallel and combined for each sample. DNA (1,000 ng) was fragmented on a Covaris S2 sonicator, using 50 µl microTUBE AFA Snap-Caps (intensity = 5, duty cycle = 5%, cycles per burst = 200, time = 35 s). Sequencing libraries were prepared using PCR-free KAPA HyperPrep Kit (Roche) according to the manufacturer's instructions. During ligation, standard adaptors were substituted for xGen Dual Index UMI Adapters (IDT). Following ligation, libraries were size selected using SPRIselect Beads (Beckman Coulter) twice, at concentrations of 0.7× and 0.5× by volume.

Sequencing of END-seq and ChIP–seq libraries

END-seq and ChIP–seq libraries were sequenced on the Nextseq 500 or Nextseq 550 platform (Illumina), using 75-bp single-end read kits. WGS libraries were sequenced on the Novaseq 6000 platform (Illumina), using 250-bp paired-end kits. Data were processed using bcl2fastq software.

CLR sequencing

High molecular mass genomic DNA was purified using the MagAttract HMW DNA kit (Qiagen). For CLR library preparation, at least 5 µg of high molecular mass genomic DNA (more than 50% of fragments ≥40 kb) was sheared to approximately 40 kb using the Megaruptor 3 (Diagenode B06010003), followed by DNA repair and ligation of PacBio adapters using the PacBio SMRTbell Express Template Prep Kit 2.0 (100-938-900). Libraries were then size-selected for >30 kb using a BluePippin instrument with 0.75% agarose cassettes (Sage Science). After quantification with the Qubit dsDNA High Sensitivity assay (Thermo Q32854), libraries were diluted to 50 pM per SMRT cell, hybridized with PacBio V2 sequencing primer, and bound with SMRT seq polymerase using Sequel II Binding Kit 2.0 (PacBio 101-842-900). CLR sequencing was performed on the Sequel II instrument using 8M SMRT Cells (101-389-001) and Sequel II Sequencing 2.0 Kit (101-820-200), with a 15 h movie time per SMRT cell. Initial quality filtering, basecalling, and adaptor marking were done automatically on board the Sequel II.

PCR

PCR reactions consisted of 1× KAPA LongRange Buffer, 0.5 U KAPA LongRange HotStart Polymerase (Roche), 1.75 mM $MgCl_2$, 300 µM dNTPs, 0.4 µM of each primer, and 75 ng of genomic DNA used as template. Samples were initially denatured at 95 °C for 3 min, underwent 32 cycles of: 1) denaturation at 95 °C for 30 s and 2) annealing/extension at 60 °C for 3 min, followed by a final extension at 60 °C for 10 min. PCR products were separated on a 2% agarose gel and visualized by 0.5 µg ml^{-1} ethidium bromide staining.

PCR primer sequences are as follows: B1 forward: 5′-GCAACCA GCTGTTTTTGTGA-3′, B1 reverse: 5′-GCAATAGTATGCAGCTTGCCC-3′; B2 forward: 5′-TTTGCATCCTGCTTTTCTCATCT-3′, B2 reverse: 5′-GA AGAGGTGCCTGGTAGCTG-3′; B3 forward: 5′-TTTGGCTTAGGG GAAGTGTGG-3′, B3 reverse: 5′-GTTTTGAGCATGCTGACCTGA-3′; B4 forward: 5′-GCAAGAACCAAATGCTGCAC-3′, B4 reverse: 5′-ACTCCT GTTGCTCAGGCAAT-3′; B5 forward: 5′-TGTCCGTGTCTCGAGGAGT-3′, B5 reverse: 5′-GTGTCTCCATCCATTGTTCTGC-3′; B6 forward: 5′-TG CTTTCAACCTGCCCAAAC-3′, B6 reverse: 5′-GCACTTGAGCCTTG CTGGTA-3′; B7 forward: 5′-TGTGGTTGTCTTCTCCACCC-3′, B7 reverse: 5′-AGCTGGGTGTTAAGGGATGAA-3′; B8 forward: 5′-ATG GGATGGCCACACTGAAG-3′, B8 reverse: 5′-AACTTGCCTTTCAC CTGCCT-3′; NB1 forward: 5′-ATAGTCTGTCTCCCGCAGTCT-3′, NB1 reverse: 5′-GAGACCGCCGATTAGCATTC-3′; NB2 forward: 5′-TACC TGACAGAACCACTGGC-3′, NB2 reverse: 5′-GACAAGGATTCC CCTCCTGC-3′; NB3 forward: 5′-GTGGTGTGGTAAAGGGACCA-3′, NB3 reverse: 5′-CCTCTCCCTGTTAAGTCATTACC-3′.

Genomic location and reference genome size for TA loci amplified by PCR

Genomic location and reference genome size of sites selected for PCR amplification are shown below. Each site amplified contains an uninterrupted stretch of $(TA)_n$ repeats and was amplified with primers localized in flanking areas outside of the $(TA)_n$ repeat. Broken sites B1–B8 were chosen based on the presence of END-seq peaks after WRN depletion in KM12 cells. Sites NB1–NB3 were chosen with similar TA repeat lengths as broken sites but were not broken after WRN depletion in KM12 cells.

Southern blotting

Native (1% agarose) Southern blot analyses of broken and non-broken $(TA)_n$ dinucleotide lengths were carried out using standard Southern blot techniques. Genomic DNA (10 µg) from indicated cell lines were digested with the indicated restriction enzymes: B1: EcoRV; B2: HindIII; NB1: ApaLI; NB3: HindIII. Probes were generated by PCR amplification using genomic DNA from the RPE-1 cell line using the following oligonucleotides:

B1 forward: 5′-GTCAAGGAAAGCCAAGAATTGGAA-3′, B1 reverse: 5′-AGAGTTGGAATTTGAACCAAAGC-3′; B2 forward: 5′-GAAACTG CCAAACACAGGGC-3′, B2 reverse: 5′-TCTTCAGCACAGGAGCAA GG-3′; NB1 forward: 5′-GTAAGGAGTGTGTGTGGGGG-3′, NB1 reverse: 5′-ACTCATGGAGAATAGACTACGACT-3′; NB3 forward: 5′-AGCCATG GTGTGTTTCTGGT-3′, NB3 reverse: 5′-GCCGTCTTCGAACCTGTAGA-3′.

Quantification and statistical analysis

Genome alignment. For single-end sequencing, NextSeq 500/550 was used for single end sequencing using 75-bp reads. The sequenced tags were aligned using Bowtie2[40] with the parameters -N 0 -k 1 -q--local--fast-local. For paired end sequencing: NovoSeq 6000 was used for paired-end sequencing using 2 × 250-bp reads. The sequenced tags were aligned using Bowtie2 with the parameters -N 0--local.

Peak calling. Peaks were called for single end sequencing experiments using MACS 1.4.3[41] using the parameters -p 1e-5--nolambda--nomodel--keep-dup = all (keep all redundant reads). For each case the experimental sample was doxycycline (shWRN) treated sample compared with DMSO (NT) sample as control. There were some additional filtering criteria as explained below:

For shWRN END-seq and RPA ChIP–seq, the output of the peak calling was filtered by a 20-fold enrichment over background and a minimum size of 1 kb. The resulting regions were merged using bedtools[42] merge -d 1000 function to define the list of regions.

For siWRN and MUS81–EME1 END-seq, the output of the peak calling was filtered by a 20-fold enrichment over background. The resulting regions were merged using bedtools merge -d 1000 function to define the list of regions. For each region defined above the summit in positive and negative strand reads was evaluated using bedtools coverage -d function. The bedtools intersect function was used to filter the regions for which the positive summit was downstream of the negative summit and the distance between the two summits was set to be between 30 and 2,500 bp. This defined the final list of peaks for each experiment.

Motif finding. For each peak the taller of the positive or negative summit was identified. From this taller summit a 1-kb window was generated downstream (upstream) if the summit was on negative (positive) strand. MEME suite[43] was used to identify the common sequence motif in these genomic windows.

For motif analysis from pacbio data, the nucleotide sequences (greater than equal to 5) corresponding to the broken TA sites were used. For each annotated $(TA)_n$ repeat sequence extracted from the hg19 genome sequence, three parameters were calculated: (a) TA proportion—that is, the cumulative TA length divided by the length of the full annotated $(TA)_n$ repeat; (b) uninterrupted $(TA)_n$ proportion—that is,

the length of the longest continuous TA sequence divided by the length of the full annotated $(TA)_n$ repeat; (c) uninterrupted $(TA)_n$ length—that is, the length of the longest continuous TA sequence within the annotated $(TA)_n$ repeat. These three parameters were plotted separately to compare between $(TA)_n$ repeats at broken and non-broken sites of shWRN-KM12 cells. Data are plotted in Extended Fig. Data 8a–c. The same was done to calculate the proportion and length of $(TA)_n$ repeats at annotated $(TA)_n$ sequences in the KM12 cell line using long-read sequencing data (Extended Data Fig. 8d–f).

Visualization of sequencing data. The alignment output sam files were converted and sorted into bam files using samtools[44] view function. For WGS paired-end data, the bam files are directly visualized using IGV[45] from the Broad Institute. The following steps were carried out for single end sequencing experiments: The bamtobed function in bedtools package was used to convert the bam files to bed files. The genomecov function in bedtools package was used to convert the bedfiles into bedGraph. This file was converted to bigwig file using the UCSC tools package and visualized on the UCSC Genome Browser[46].

Expansion detection by ExpansionHunter and exSTRa. Expansions are identified for broken and non-broken $(TA)_n$ from the HCT116 and KM12 whole genome sequencing data by ExpansionHunter v.3.2.2[25,47]. The supporting reads for expansions are visualized by GraphAlignmentViewer (https://github.com/Illumina/GraphAlignmentViewer). Empirical cumulative distribution function for the TA repeat located at hg19 coordinates chr8:106950919–106950985 was generated using exSTRa v.0.89.0[26] and Bio-STR-exSTRa v.1.1.0 using the default parameter settings.

Long-read sequencing data processing. We aligned CLR reads to the hg19 reference sequence using minimap2[48] (with parameters -ayYL--MD--eqx -x map-pb) and called indel/structural variants with the pbsv tool using the -tandem-repeats argument for increased sensitivity to repeat expansion/contraction variants in accordance with the software's documentation (Pacific Biosciences, pbsv https://github.com/PacificBiosciences/pbsv). We extracted indel variants passing built-in filters and overlapping with repetitive loci of interest using the bedtools closest utility[42]. We computed coverage over each interval using the mosdepth[49] package and determined the repetitive motifs present using the tandem repeat finder (trf) tool[50], lowering the score threshold to see all possible motifs (that is, with parameters 2 5 7 80 10 0 2000 -l 6 -ngs -h). We quantified the copy number of each motif in the reference haplotype and the alternate haplotype (computed by permuting the reference haplotype with the variants detected in each locus). We report the copy number change for the highest copy number motif per locus.

Cancer Cell Line Encyclopedia whole-genome sequencing analysis. For 321 cancer cell lines from the Cancer Cell Line Encyclopedia (CCLE) 2 project[51] that had whole-genome sequencing data, we determined the number of reads at each AT-rich interval using 'samtools bedcov'[44]. For each interval, we then defined: fragments per base (FPB) = number of reads covering the interval/(length of the interval in bases × 2). We then normalized this value across all intervals within the genome as: fragments per base per million (FPBM) = (FPB/total FPB) × 1 million. This latter quantity (FPBM) was used to determine the read coverage at broken/non-broken loci. We determined overall read counts at loci for with high AT repeats.

MSI deletion analysis. MSI deletions in UCEC, COAD, and STAD[5] were identified from ICGC (https://dcc.icgc.org/releases/PCAWG/msi and https://dcc.icgc.org/releases/PCAWG/consensus_sv). In total, 396 deletions from 24 MSI samples and 4,501 deletions from 93 MSS samples were analysed. 1,000 random sets were generated using the bedtools[42] shuffle command in order to estimate enrichments for different repeat annotations relative to random.

Statistical analysis

For Venn diagrams, 1,000 random sites were generated for one of the peak lists (while keeping others the same) using bedtools shuffle command. The significance was generated using the fisher.test function in R. For box plots, the statistical significance of the data represented was calculated using the wilcox.test function in R. No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

END-seq, ChIP–seq, whole-genome sequencing and Pacbio CLR data have been deposited in the Gene Expression Omnibus (GEO) database under the accession number GSE149709. Source data are provided with this paper.

37. Callen, E. et al. 53BP1 enforces distinct pre- and post-resection blocks on homologous recombination. *Mol. Cell* **77**, 26–38 (2020).
38. Palermo, V. et al. CDK1 phosphorylates WRN at collapsed replication forks. *Nat. Commun.* **7**, 12880 (2016).
39. Chang, J. H., Kim, J. J., Choi, J. M., Lee, J. H. & Cho, Y. Crystal structure of the Mus81-Eme1 complex. *Genes Dev.* **22**, 1093–1106 (2008).
40. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
41. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
42. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
43. Bailey, T. L. et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202-8 (2009).
44. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
45. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
46. Kent, W. J. et al. The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
47. Dolzhenko, E. et al. ExpansionHunter: a sequence-graph-based tool to analyze variation in short tandem repeat regions. *Bioinformatics* **35**, 4754–4756 (2019).
48. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
49. Pedersen, B. S. & Quinlan, A. R. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics* **34**, 867–868 (2018).
50. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
51. Ghandi, M. et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* **569**, 503–508 (2019).

Acknowledgements We thank R. Awasthi for assistance with Southern blotting; D. Goldstein, B. Tran and the CCR Genomics core for sequencing support; M. Lawrence for computational assistance; and F. Alt, B. Vogelstein and J. Haber for helpful discussions. Work in the S.C.W. laboratory is supported by the Francis Crick Institute (FC10212) and the European Research Council (ERC-ADG-666400). The Francis Crick Institute receives core funding from Cancer Research UK, the Medical Research Council, and the Wellcome Trust. K. Fugger is the recipient of fellowships from the Benzon Foundation and the Lundbeck Foundation. The P.J.M. laboratory is funded by the MRC MR/R009368/1; A.C.-M. is the recipient of a fellowship from AstraZeneca; E.M.C. is supported by the Damon Runyon Cancer Research Foundation, and E.M.C. and A.J.B. are supported by a pilot grant from the Dana-Farber Department of Medical Oncology. The A.N. laboratory is supported by the Intramural Research Program of the NIH, an Ellison Medical Foundation Senior Scholar in Aging Award (AG-SS- 2633-11), the Department of Defense Idea Expansion (W81XWH-15-2-006) and Breakthrough (W81XWH-16-1-599) Awards, the Alex's Lemonade Stand Foundation Award, and an NIH Intramural FLEX Award.

Author contributions N.v.W. set up the project, performed END-seq and flow cytometry experiments upon WRN, MUS81 and SLX4 depletion, and performed preliminary analysis of END-seq data; W.J.N. performed MUS81–EME1 in situ END-seq and PCR; A.T. performed END-seq, Southern blotting and designed ATR-mutant WRN cDNA; E.M.C. generated the inducible *WRN* shRNA in KM12 and HCT116 cells, performed and analysed the HSEC western blot and viability experiments, long-read sequencing, and analysed the CCLE and WRN dependency data; E.C. performed ATRi END-seq experiments, western blotting, and metaphase analysis. V.T. performed RPA ChIP–seq; K. Foster performed the HSEC and long-read sequencing experiments; N.W. performed western blotting and helped to generate WRN(3A) and WRN(6A) cells; J.N. and J.K. analysed the CCLE and WRN dependency data; S.S.

analysed END-seq, RPA ChIP–seq experiments; W.W. analysed WGS, PacBio coverage across repeats, deletion breakpoints in MSI cancers, and performed quantitative modeling; F.B. analysed nucleotide composition of broken versus non-broken repeats and replication timing; E.D. performed ExpansionHunter and exSTRa bioinformatic analysis; M.A.E. supervised computational work; K.G., Y.H., A.A.B., J.T.S. and N.K. analysed the data and designed bioinformatic pipelines; R.L.W. prepared WGS libraries; A.C.-M. and K. Fugger provided recombinant MUS81–EME1; J.A.S. provided recombinant WRN; B.E.H. provided advice about PCR across repeats; K.U. provided advice about repeat expansion biology; C.H.F. provided advice about secondary structure biology; R.M.B. provided advice about WRN helicase; S.C.W. provided advice about structure specific nucleases and recombination intermediates; P.J.M. helped design in situ experiments with recombinant proteins; P.S.M. provided advice on WGS experiments and analyses; A.J.B. and A.N. supervised the project; N.v.W., W.J.N., A.T., E.M.C., A.J.B. and A.N. wrote the manuscript with comments from the other authors. N.v.W., S.S., W.J.N., A.T. and E.M.C. contributed equally; E.C. and W.W. contributed equally as second authors.

Competing interests A.J.B. has research support from Bayer, Merck and Novartis.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2769-8.

Correspondence and requests for materials should be addressed to A.N.

Peer review information *Nature* thanks Sergei Mirkin and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at http://www.nature.com/reprints.

RESEARCH ARTICLE

Suppressing proteasome mediated processing of topoisomerase II DNA-protein complexes preserves genome integrity

Nicholas Sciascia[1,2], Wei Wu[1†], Dali Zong[1†], Yilun Sun[3], Nancy Wong[1], Sam John[1], Darawalee Wangsa[4], Thomas Ried[4], Samuel F Bunting[5], Yves Pommier[3], André Nussenzweig[1]*

[1]Laboratory of Genome Integrity, National Institutes of Health, Bethesda, United States; [2]Institute for Biomedical Sciences, George Washington University, Washington, United States; [3]Developmental Therapeutics Branch, National Institutes of Health, Bethesda, United States; [4]Genetics Branch National Cancer Institute, National Institutes of Health, Bethesda, United States; [5]Department of Molecular Biology and Biochemistry, Rutgers University, Piscataway, United States

Abstract Topoisomerase II (TOP2) relieves topological stress in DNA by introducing double-strand breaks (DSBs) via a transient, covalently linked TOP2 DNA-protein intermediate, termed TOP2 cleavage complex (TOP2cc). TOP2ccs are normally rapidly reversible, but can be stabilized by TOP2 poisons, such as the chemotherapeutic agent etoposide (ETO). TOP2 poisons have shown significant variability in their therapeutic effectiveness across different cancers for reasons that remain to be determined. One potential explanation for the differential cellular response to these drugs is in the manner by which cells process TOP2ccs. Cells are thought to remove TOP2ccs primarily by proteolytic degradation followed by DNA DSB repair. Here, we show that proteasome-mediated repair of TOP2cc is highly error-prone. Pre-treating primary splenic mouse B-cells with proteasome inhibitors prevented the proteolytic processing of trapped TOP2ccs, suppressed the DNA damage response (DDR) and completely protected cells from ETO-induced genome instability, thereby preserving cellular viability. When degradation of TOP2cc was suppressed, the TOP2 enzyme uncoupled itself from the DNA following ETO washout, in an error-free manner. This suggests a potential mechanism of developing resistance to topoisomerase poisons by ensuring rapid TOP2cc reversal.

***For correspondence:** andre_nussenzweig@nih.gov

[†]These authors contributed equally to this work

Competing interests: The authors declare that no competing interests exist.

Funding: See page 23

Received: 08 November 2019
Accepted: 12 February 2020
Published: 14 February 2020

Reviewing editor: Maureen Murphy, The Wistar Institute, United States

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

Introduction

Topoisomerase II (TOP2) is an evolutionarily conserved enzyme capable of generating reversible double strand breaks (DSBs) in DNA, which enables the resolution of problematic DNA topological structures that arise during normal cellular processes, such as transcription, replication, and mitosis. This process is mediated by the topoisomerase II cleavage complex (TOP2cc), a transient DNA-protein complex that is formed when each monomer of the dimeric TOP2 anchors itself to the DNA backbone via a covalent 5'-phosphotyrosyl bond. A second DNA strand can then be passed through the enzyme-bridged DSB, after which the break is re-ligated. Under physiological conditions, topoisomerases reliably execute this reaction without error, as TOP2ccs are highly reversible and short lived (***Pommier et al., 2016***). However, when they become trapped, the abortive TOP2ccs fail to re-ligate the transient DSB intermediate, presenting a severe threat to genome integrity which must be

eLife digest Molecules of DNA contain the archive of a cell's genetic information and identity. DNA comprises two strands that twist together into a structure known as a double helix. Physical tension tends to build up in the double helix that can cause it to break apart. To avoid this, cells have an enzyme called Topoisomerase II (TOP2) that relieves the tension by attaching itself to DNA and breaking it in a controlled way before re-sealing the break.

Drugs known as TOP2 poisons stop TOP2 from working and trap it on the DNA, which may lead to cells accumulating DNA breaks and eventually dying. Cancer cells are particularly prone to acquiring breaks in their DNA, and TOP2 poisons are therefore often used as part of chemotherapy treatments for cancer. However, it remains unclear why TOP2 poisons are more effective at killing some types of cancer cells than others.

It is thought that a molecular machine, known as the proteasome, helps cells repair the damage caused by TOP2 poisons by removing the trapped TOP2 proteins and allowing DNA repair proteins access to the broken DNA underneath. Now, Sciascia et al. have used a genetic approach to study the relationship between the proteasome and DNA repair in mouse cells exposed to TOP2 poisons.

The experiments found that when the proteasome removed TOP2 proteins that had become trapped on DNA, the subsequent DNA repair was prone to errors. Pre-treating mouse cells with another drug that inhibited the proteasome protected the cells from the effects of the TOP2 poison. Once the TOP2 poison had left the cells, the previously trapped TOP2 proteins correctly fixed the DNA and detached as they would normally. As a result, cells that had been treated with a proteasome inhibitor were more likely to survive treatment with TOP2 poisons.

Since both TOP2 poisons and proteasome inhibitors are clinically approved drugs for treating cancer they can be, and already have been, tested for use together in combination drug therapies. However, these findings suggest that caution should be taken when using these drugs together, because instead of harming the cancer cells, the proteasome inhibitors may protect the cells from the toxic effects of TOP2 poisons.

resolved quickly via the DNA repair machinery. The resolution of TOP2ccs is, therefore, crucial for cell survival.

TOP2 poisons, such as the anti-cancer agent etoposide (ETO), efficiently trap and stabilize TOP2ccs (***Muslimović et al., 2009***; ***Nitiss, 2009a***; ***Pommier et al., 2016***). The accumulation of these TOP2 DNA-protein complexes can create significant problems for cells by stalling replication forks and the transcription apparatus, generating torsional and genotoxic stress, which leads to the accumulation of single and double strand DNA breaks (***Cowell and Austin, 2012***; ***Muslimović et al., 2009***). How these abortive TOP2ccs are recognized and targeted by the DNA damage response (DDR) remains unclear.

It has been suggested that a robust DNA damage response (DDR) is elicited only after ETO-trapped TOP2ccs have been converted to 'clean' protein-free DSB ends (***Mårtensson et al., 2003***; ***Sunter et al., 2010***; ***Zhang et al., 2006***). This suggests that a DSB shielded by a covalently bound protein can evade detection. Several mechanisms exist for removal of trapped TOP2ccs. However, the factors that determine how a cell chooses to deploy them are complex and remain poorly understood. The covalently-linked TOP2 protein is partially or completely degraded by the 26S proteasome, after which the residual DNA-peptide adduct is hydrolyzed by tyrosyl-DNA phosphodiesterase 2 (TDP2) to release the remnants of the entrapped protein (***Ledesma et al., 2009***; ***Gómez-Herreros et al., 2013***; ***Lee et al., 2018***; ***Sunter et al., 2010***; ***Zagnoli-Vieiral and Caldecott, 2017***). Current evidence suggests that transcription increases the rate at which TOP2 is processed by the proteasome (***Canela et al., 2019***). It has also been reported that TDP2, with the co-operation of secondary cofactors, can resolve unprocessed TOP2ccs in a proteasome-independent manner (***Schellenberg et al., 2017***).

An alternate, more error-prone mechanism for TOP2cc removal is mediated by MRE11 endonucleolytic cleavage in the vicinity of a trapped TOP2cc, which releases the entire TOP2 DNA-protein complex that results in the loss of small stretches of DNA (***Hoa et al., 2016***; ***Lee et al., 2012***). Either of these mechanisms are able to produce protein-free DNA breaks that can then be recognized and

repaired by the major cellular DSB repair pathways: non-homologous end joining (NHEJ) and homologous recombination (HR).

It has long been recognized that trapped TOP2ccs are intrinsically reversible upon ETO washout (**Hsiang and Liu, 1989**; **Long et al., 1985**). Consistently, we observed that inhibiting the proteasome prior to ETO treatment enhanced the number of rapidly reversible ETO-stabilized TOP2ccs across the genome (**Canela et al., 2019**). Other studies have shown that inhibiting the proteasome not only preserves the reversibility of TOP2ccs but also is able to suppress DDR signaling (**Mao et al., 2001**; **Zhang et al., 2006**). However, the implications of suppressing the DDR in response to ETO using proteasome inhibitors, on both the long-term genome integrity and the overall viability of the cell, have not been fully explored. Considering that both proteasome inhibitors and topoisomerase poisons are used, sometimes in tandem, as frontline chemotherapeutic agents to treat a variety of cancers (**Cowell and Austin, 2012**; **Dittus et al., 2018**; **Manasanch and Orlowski, 2017**; **Thomas et al., 2017**), understanding how TOP2cc reversibility may impact the effectiveness of these drugs is clinically relevant.

Here, we utilize high-resolution genome-wide mapping of DSBs by END-seq (**Canela et al., 2016**) to examine the impact of proteasome inhibition on the fate of TOP2ccs. We found that once ETO-stabilized TOP2ccs had been proteolytically processed throughout the genome, and a robust DDR had been initiated, the large number of newly generated protein-free DSBs were repaired in a highly error-prone manner, resulting in toxic chromosomal translocations and rearrangements that led to cell death. Notably, the major DSB repair pathways (NHEJ and HR) contribute to the mis-repair of ETO-induced DSBs following proteasomal processing. By inhibiting proteasome-mediated degradation of TOP2ccs either through chemical inhibition or by ablation of RNF4-mediated ubiquitination of TOP2ccs, an intact and enzymatically competent TOP2 was able to re-seal the protein-linked DSBs without invoking a significant DDR. As a result, cells were protected from genomic instability, which ultimately led to enhanced cell survival after treatment with topoisomerase poisons.

Results

Proteasomal activity is necessary for triggering a DNA damage response following ETO treatment

Building on previous observations that proteasome inhibition can specifically attenuate the induction of γ-H2AX by ETO (**Mao et al., 2001**; **Zhang et al., 2006**), we sought to first confirm these findings in mouse primary splenic B-cells. Since ETO stabilizes both TOP2A and TOP2B isoforms, resting B-cells were activated with cytokines for 12 hr (**Figure 1A**). This short-term treatment ensures that B-cells remain in G1 where only the TOP2B isoform is expressed (**Canela et al., 2017**). This treatment regime, therefore, allows us to minimize any differential effect of ETO on each isoform (**Errington et al., 2004**; **Willmore et al., 1998**) and to bypass essential TOP2-dependent processes, such as DNA replication.

In agreement with other studies (**Mao et al., 2001**; **Zhang et al., 2006**), we found that treatment with a high (50 μM) dose of ETO for 2 hr elicited a strong γ-H2AX response in B-cells (**Figure 1B**). Pre-incubating cells with proteasome inhibitors (10 μM MG132, 10 μM Bortezomib (BTZ), 2 μM Epoxomicin (EPN), or 10 μM Ixazomib (IXA)) for 1 hr prior to ETO treatment almost completely abolished the γ-H2AX signal (**Figure 1B**). The suppressed γ-H2AX signal occurred specifically in response to ETO, as pre-treating cells with the same dose of proteasome inhibitor did not affect the γ-H2AX response to ionizing γ-irradiation (5 Gy IR) (**Figure 1B and C**). Thus, proteasomal activity is required for triggering a robust DDR specifically in response to ETO treatment.

Proteasome inhibition promotes the accumulation of reversible TOP2ccs

As trapped TOP2ccs are subjected to proteasomal degradation, unprocessed TOP2ccs accumulate on DNA in cells pre-treated with a proteasome inhibitor compared to cells treated with ETO alone (**Lee et al., 2016**). To confirm that proteasome inhibitors were having the same effect in our experimental system, we quantified TOP2ccs using the ICE assay (**Anand et al., 2018**) in ETO-treated primary B-cells pre-incubated with or without BTZ. In the presence of ETO, TOP2ccs were readily detected on DNA (**Figure 1D**). Immediately after ETO was washed out, however, the amount of

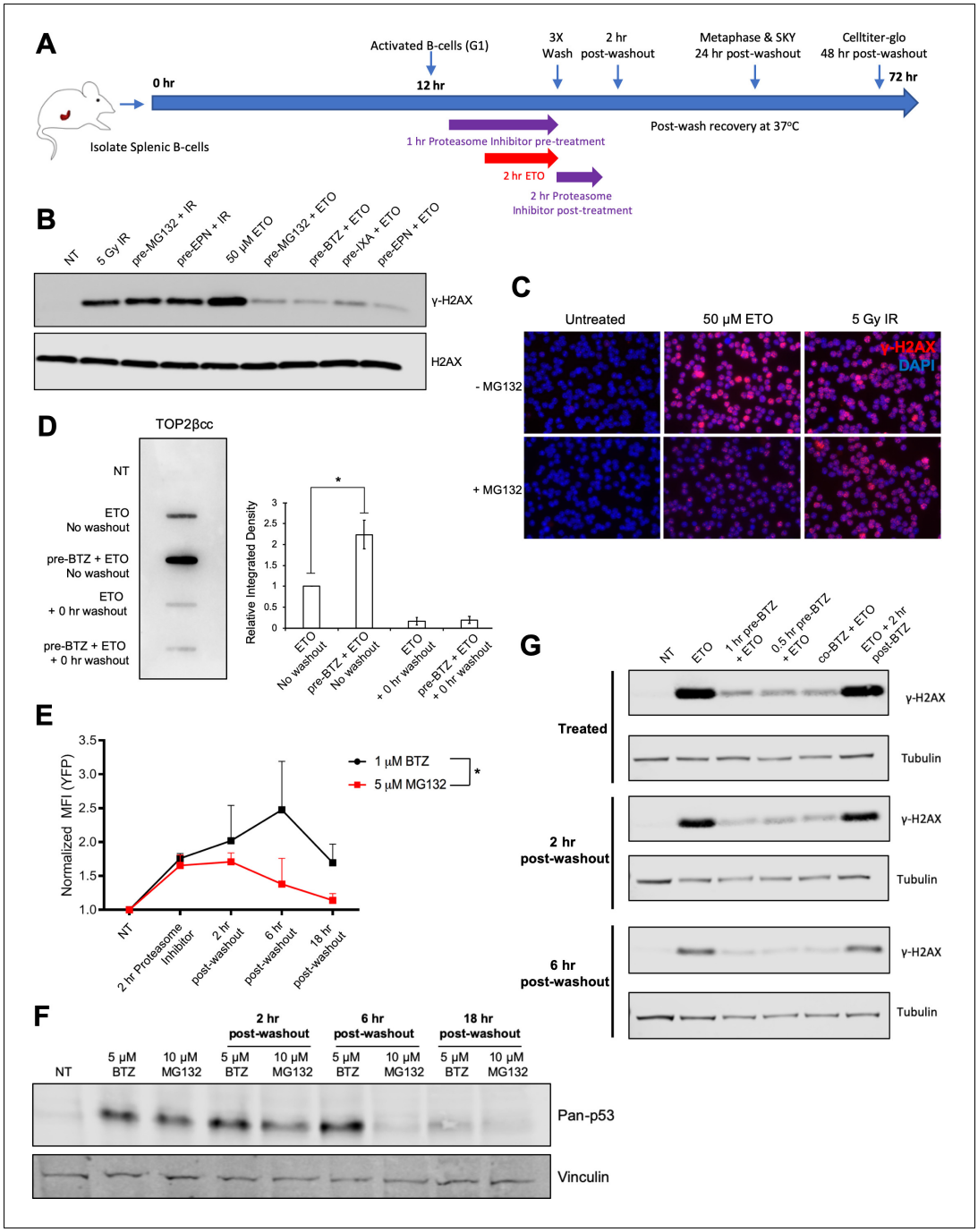


Figure 1. Proteasome machinery is essential for triggering DNA damage response to ETO treatment. (**A**) As indicated in the schematic, primary splenic B-cells were isolated from mouse spleens and activated with a cocktail of Il-4, LPS, and RP105. 12 hr post-activation, while the B-cells were still in the G1 phase of the cell cycle, they were pre-treated with proteasome inhibitor (10 μM MG132 or 5 μM Bortezomib) for 1 hr prior to an additional 2 hr co-treatment with 50 μM ETO. Following the ETO pulse, the cells were washed with ice cold drug-free media (3 × 5 min spin at 1500 rpm and 4˚C to pellet B-cells between washes; 15–20 min total time to complete washout), then returned to fresh drug free media and allowed to recover at 37˚C for up to 48 hr. In the case of post-treatment, proteasome inhibitor was added to the wash media and then the cells were incubated for an additional 2 hr at 37˚C with proteasome inhibitor only. For western blot and ICE assay analysis, cells were harvested immediately following drug treatment or washout (0 hr washout). For metaphase and SKY analysis, cells were harvested 24 hr following washout. For CellTiter-Glo viability, cells were harvested 48 hr following washout. For END-seq analysis, cells were harvested before drug washout, immediately following washout (0 hr washout), or after a 2 hr recovery

Figure 1 continued on next page

Figure 1 continued

in drug free media (2 hr washout). (**B**) γ-H2AX western blot in G1 WT primary splenic B-cells following 2 hr 50 μM ETO treatment ±1 hr proteasome inhibitor pre-treatment or 30 min following 5Gy IR ±1 hr proteasome inhibitor pre-treatment. Proteasome inhibitors tested: 10 μM MG132, 10 μM BTZ (Bortezomib), 2 μM EPN (Epoxomicin), 10 μM IXA (Ixazomib). (**C**) γ-H2AX immunofluorescence staining (red) in G1 arrested WT pre B cells following 2 hr, 50 μM ETO treatment ±1 hr 10 μM MG132 pre-treatment or 30 min after 5 Gy IR treatment ±1 hr 10 μM MG132 pre-treatment. (**D**) WT primary splenic B-cells (N = 3) in G1 were treated for 2 hr with 50 μM ETO ±1 hr, 5 μM BTZ pre-treatment. DNA was then isolated from cells following the ICE assay protocol and probed with anti-TOP2β antibody to quantify levels of TOP2βcc. Relative band intensity was measured for each sample and averaged for all three mice (ETO vs. pre-BTZ + ETO p=0.0168; ETO vs. ETO + 0 hr washout p=0.0045; pre-BTZ + ETO vs. pre-BTZ + ETO 0 hr washout p=0.0046; statistical significance calculated using student T-test). (**E**) eHAP cells were transfected for 48 hr with a YFP-tagged degron construct, which is degraded in a proteasome-dependent manner (***Bence et al., 2001***). Cells were then treated with 5 μM MG132 or 1 μM BTZ for 2 hr and then the drugs were washed out (3X wash in cold drug free media). Cells were fixed before washout with drugs still present or at 2 hr, 6 hr, and 18 hr post-washout. YFP signal intensity was quantified by FACS. (BTZ (before washout) vs. NT p=0.025; 2 hr post-washout MG132 vs NT p=0.049; statistical significance calculated using student T-test. MG132 vs BTZ p=0.038; statistical significance calculated using two column ANOVA). (**F**) pan-p53 levels in primary splenic B-cells treated for 3 hr with 10 μM MG132 or 5 μM BTZ. Drugs were then washed out and samples collected before washout or at 2 hr, 6 hr and 18 hr post-washout. (**G**) γ-H2AX levels in primary splenic B-cells as determined by Western blotting. Cells were treated with 50 μM ETO for 2 hr ±1 hr or 30 min pre-treatment, co-treatment and post-treatment with 5 μM BTZ at three different timepoints: before washout (treated), 2 hr and 6 hr post-washout.

DNA-bound TOP2 diminished strongly (***Figure 1D***). The proteasome continuously degraded ETO-trapped TOP2ccs, as evidenced by the markedly enhanced accumulation of TOP2ccs when cells were pre-incubated with BTZ (***Figure 1D***). The majority of these TOP2ccs also dissociated from DNA immediately following ETO washout (***Figure 1D***). Thus, inhibiting proteasomal degradation appears to preserve the integrity of TOP2ccs enabling reversal of complexes by completion of the enzymes' catalytic cycle upon drug removal.

Alternatively, proteasome activity could recover immediately after the removal of ETO and the proteasome inhibitor and would similarly lead to a loss of TOP2cc signal in the ICE assay. To determine the duration and reversibility of proteasomal inhibition, we directly monitored proteasome activity in eHAP cells transfected with a YFP-Degron reporter that has been demonstrated to be a *bona fide* proteasomal substrate (***Bence et al., 2001***). As expected, a 2 hr treatment with MG132 or BTZ significantly increased the YFP signal from baseline (***Figure 1E***). Following washout of MG132 or BTZ, the elevated YFP signal persisted for several hours before it began to decrease and BTZ appeared to be a more potent and persistent inhibitor of the proteasome than MG132 (***Kisselev and Goldberg, 2001***). As an additional measure of proteasome activity, we quantified the protein levels of p53, as it is known to be stabilized upon proteasome inhibition (***An et al., 2000***; ***Halasi et al., 2014***). Consistent with the YFP-degron results in eHAP cells, we observed that p53 protein remained stabilized in primary B-cells for several hours after proteasome inhibitors were washed out, with BTZ again being more potent than MG132 (***Figure 1F***).

Thus, proteasome activity is not readily recovered even after the removal of proteasome inhibitor, suggesting that the rapid loss of ETO-induced TOP2ccs in MG132 pre-treated cells upon washout most likely reflects the reversal of TOP2ccs by completion of the enzymes' catalytic cycle upon drug removal. Accordingly, we did not observe a delayed γ-H2AX induction at either 2 hr or 6 hr post-ETO and BTZ washout, suggesting that proteasomal activity remain suppressed for at least several hours post-washout (***Figure 1G***). These data imply that persistent proteasome inhibition allows for TOP2cc reversal and prevents trapped TOP2ccs from being converted into protein-free DSB ends that are capable of eliciting a robust DNA damage response (DDR).

Timing of proteasome inhibition determines its impact on TOP2 metabolism

Contrary to our observations, previous studies have shown that proteasome inhibitors synergize with topoisomerase poisons like ETO in mediating cell killing (***Aras and Yerlikaya, 2016***; ***Ceruti et al., 2006***; ***Destanovic et al., 2018***; ***Dittus et al., 2018***; ***Lee et al., 2016***; ***von Metzler et al., 2009***). Interestingly, we found that the addition of BTZ prior to or concurrent with ETO suppressed DDR

signaling, but incubating B-cells with BTZ post-ETO treatment did not (***Figure 1G***). These results showed that the timing of proteasome inhibitor treatment relative to ETO treatment is critical to its effects on TOP2cc metabolism and subsequent DDR signaling.

Proteasome inhibition decreases the persistence of ETO-induced TOP2-mediated DSBs by promoting TOP2cc reversibility

To further analyze the influence of the proteasome on TOP2ccs, we employed genome-wide DSB mapping by END-seq (***Canela et al., 2019***; ***Canela et al., 2016***). While ETO can generate high levels of both SSBs and DSBs (***Baranello et al., 2014***; ***Gittens et al., 2019***), END-seq only detects DSBs generated by ETO (***Canela et al., 2017***; ***Canela et al., 2016***). However, this protocol allows us to capture and distinguish both TOP2ccs and protein-free DSBs generated by ETO (***Canela et al., 2019***). First, we assessed whether proteasome inhibition blocked TOP2 from making incisions in DNA. To this end, we used a cocktail of Exonuclease VII (ExoVII) and Exonuclease T (ExoT) during sample preparation, allowing us to capture all sites of TOP2 activity (***Canela et al., 2019***). As expected, END-seq analysis revealed that TOP2-mediated DSBs were readily detectable in ETO-treated B-cells (***Figure 2A–C***). The number and intensity of TOP2-mediated DSBs were reproducible across experiments (***Figure 2—figure supplement 1A–B***). Consistent with the results shown in ***Figure 1D***, we found that BTZ pre-treatment did not dramatically affect the overall number or location of TOP2-mediated DSBs (***Figure 2A–C***; ***Figure 2—figure supplement 1C–D***); TOP2-mediated DSBs were still strongly induced in the BTZ pre-treated sample, despite eliciting only a weak γ-H2AX response (***Figure 1B***). Thus, while proteasome inhibition does not affect the ability of TOP2 to cleave DNA, it helps maintain the resultant DSBs in a protein-linked form (persistent TOP2cc) that conceals them from the DDR. Nevertheless, we did observe a small decrease in the overall intensity of the TOP2-mediated DSB signal (***Figure 2A and C***) upon BTZ pre-treatment. This could be due to reversal of TOP2ccs in the BTZ pre-treated sample by completion of the enzymes' catalytic cycle upon drug removal.

Since inhibiting proteasomal degradation did not affect the initial formation of TOP2-mediated DSBs, we wanted to understand if proteasome inhibition affected their fate over time. To directly evaluate the impact of proteasome inhibition on repair of TOP2-mediated DSBs, we used END-seq to assess their longevity by harvesting primary B-cells both immediately following ETO/BTZ washout (0 hr washout) and after a 2 hr recovery in drug-free media at 37˚C (2 hr washout). Notably, BTZ exerted a substantial impact on the rate at which TOP2-mediated DSBs resolved over the course of the 2 hr washout, with very few detectable breaks remaining after recovery relative to cells treated with ETO alone (***Figure 2A and D***; ***Figure 2—figure supplement 1E***). We plotted the fraction of TOP2-mediated DSBs at each washout timepoint relative to the initial level of breakage following ETO treatment (***Figure 2D***). Using this analysis, we found that immediately following drug washout, 30% of the initial TOP2-mediated DSBs were still detectable in cells treated with ETO only, while just 15% were still present in the corresponding BTZ pre-treated sample (***Figure 2A and D***). After a 2 hr recovery, B-cells treated with ETO alone still had a detectable break signal (17%), whereas in BTZ-pretreated cells, it returned almost to baseline (4%, ***Figure 2A and D***). This suggested that inhibiting the proteasome enhanced the amount of TOP2ccs which rapidly reverse following ETO washout.

As a second form of analysis, we normalized the TOP2-mediated DSB intensity at each post-washout timepoint. We introduced a new parameter, termed TOP2-mediated DSB persistence, which was calculated as the ratio of the peak intensity (RPKM) for individual TOP2-mediated DSBs at each timepoint (either 0 hr washout or 2 hr washout) relative to its initial intensity in the corresponding pre-washout sample (***Figure 2E***; ***Figure 2—figure supplement 1F***). Through this analysis, we found that approximately 20% of the initial TOP2-mediated DSBs were classified as persistent immediately following ETO washout (0 hr washout), and 13% of initial TOP2-mediated DSBs were still persistent 2 hr post-washout (2 hr washout) (***Figure 2E***). By contrast, BTZ pre-treatment more significantly reduced the fraction of persistent TOP2-mediated DSBs both immediately following washout (~13%), as well as at 2 hr post-washout (~3.5%) (***Figure 2E***). These results, in conjunction with the ICE assay (***Figure 1D***), suggested that inhibiting the proteasome enhances TOP2cc reversal following ETO washout.

To directly measure the reversibility of TOP2ccs, we took advantage of the fact that ExoT lacks the ability to process protein-linked DSBs (persistent TOP2ccs that still have an intact 5'-

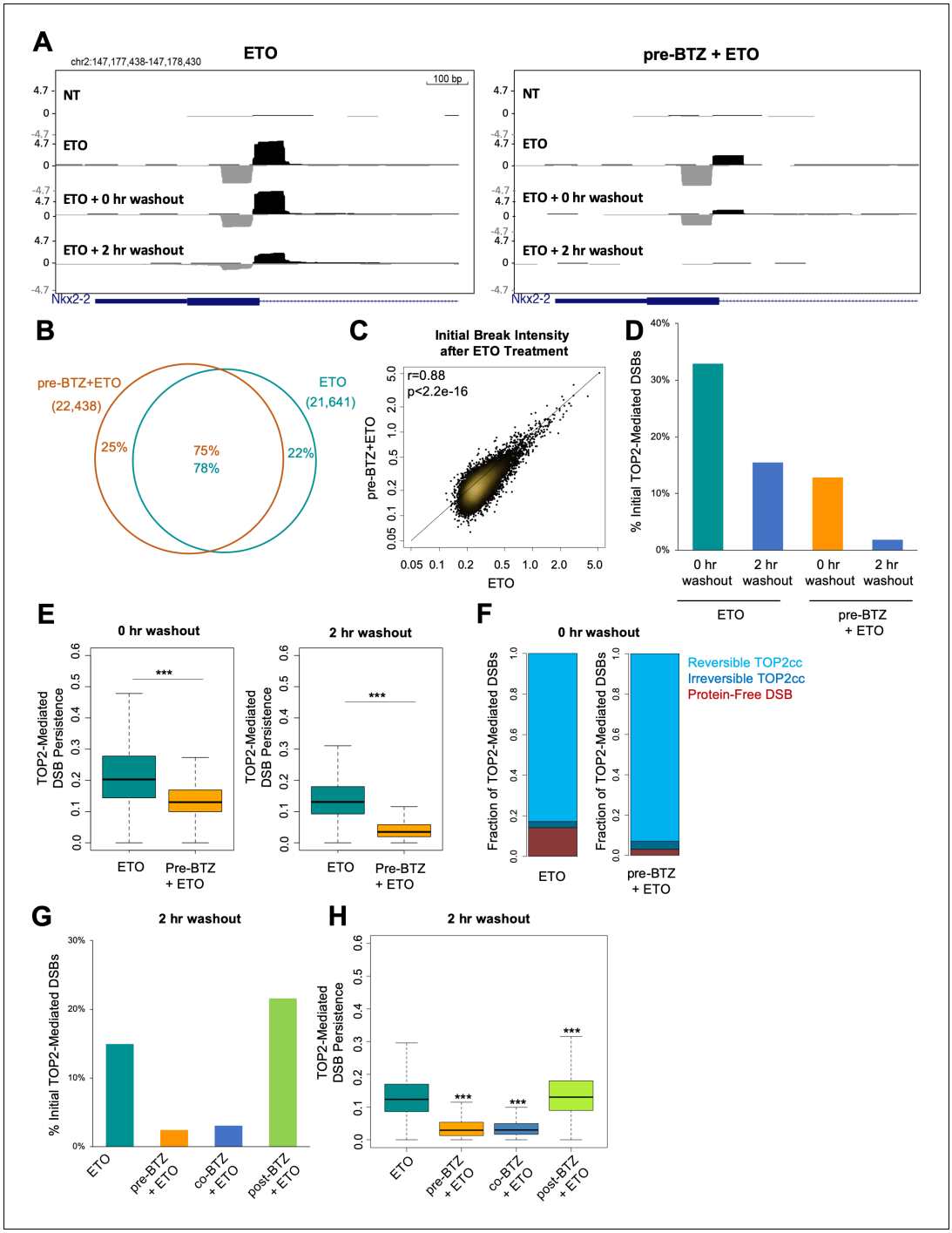


Figure 2. Proteasome inhibition decreases TOP2-mediated DSB persistence following ETO washout. (**A**) UCSC genome browser snapshot of a TOP2-mediated DSB captured within the Nkx2-2 gene locus by END-seq (ExoVII +ExoT) at three timepoints: in primary splenic B-cells treated for 2 hr with 50 μM ETO ± 1 hr 5 μM BTZ pre-treatment, immediately following drug washout (0 hr washout), and at 2 hr post-washout (2 hr washout). (**B**) Venn diagram depicting the overlap between TOP2-mediated DSBs generated by ETO genome-wide in the 50 μM ETO only sample (green) and the 5 μM BTZ pre-treatment sample (orange). WT primary splenic B-cells were treated for 2 hr with 50 μM ETO ± 1 hr 5 μM BTZ pre-treatment and processed by END-seq using ExoVII+ExoT. (**C**) Scatterplot depicting initial intensity of TOP2-mediated DSBs induced by ETO genome-wide in the 50 μM ETO only sample vs. the 5 μM BTZ pre-treatment sample (ETO vs. pre-BTZ + ETO $p<2.2e^{-16}$; statistical significance calculated by Pearson's correlation coefficient). WT primary splenic B-cells were treated for 2 hr with 50 μM ETO ± 1 hr 5 μM BTZ pre-treatment. (**D**) Bar graph measuring the number of TOP2-mediated DSBs (END-seq with

Figure 2 continued on next page

Figure 2 continued

ExoVII+ExoT) at 0 hr and 2 hr post-washout timepoints, represented as a fraction of the initial number of TOP2-mediated DSBs generated genome-wide (% initial breaks after ETO treatment determined by peak calling). WT primary splenic B-cells were treated for 2 hr with 50 μM ETO treatment ±1 hr 5 μM BTZ pre-treatment. (**E**) Box-plot of TOP2-mediated DSB persistence (relative intensity of initial breakage) genome-wide at 0 hr and 2 hr post-washout timepoints (left and right panels respectively). Green box 50 μM ETO only sample and orange box 5 μM BTZ pre-treatment sample (ETO vs. pre-BTZ + ETO $p<2.2e^{-16}$, statistical significance calculated by student T-test). WT primary splenic B-cells were treated for 2 hr with 50 μM ETO treatment ±1 hr 5 μM BTZ pre-treatment prior to washout. (**F**) Bar graph depicting the percentages of total TOP2-mediated DSBs (measured by ExoVII+ExoT), protein-free DSBs (measured by ExoT), and TOP2ccs (inferred by subtracting protein free DSBs from the levels of total TOP2-mediated DSBs). Reversible TOP2ccs (light blue), irreversible TOP2ccs (dark blue), and protein-free DSBs (red). WT primary splenic B-cells were treated for 2 hr with 50 μM ETO treatment ±1 hr 5 μM BTZ pre-treatment. (**G**) Bar graph depicting the number of TOP2-mediated DSBs at 2 hr post-washout (END-seq with ExoVII+ExoT), represented as a fraction of the initial number of TOP2-mediated DSBs measured genome-wide (% initial breaks after ETO treatment determined by peak calling). WT primary splenic B-cells were treated for 2 hr with 50 μM ETO treatment ±pre treatment, co-treatment, and post-treatment with 5 μM BTZ. Post treatment with BTZ was for 2 hr. (**H**) Box-plot depicting TOP2-mediated DSB persistence genome-wide at 2 hr post-washout ($p<2.2e^{-16}$ for ETO vs. pre-BTZ + ETO, ETO vs co-BTZ + ETO, and ETO vs post-BTZ + ETO; statistical significance calculated by student T-test). WT primary splenic B-cells were treated for 2 hr with 50 μM ETO treatment ±pre treatment, co-treatment, and post-treatment with 5 μM BTZ.

The online version of this article includes the following figure supplement(s) for figure 2:

Figure supplement 1. TOP2-Medated DSBs are reproducibly captured by END-seq.

phospotyrosyl bond) and therefore allows for the detection of only protein-free (i.e. proteolytically processed) DSBs (***Canela et al., 2019***). As such, the levels of TOP2ccs can be estimated by the difference between total TOP2-mediated DSBs detected by ExoVII+ExoT and protein-free DSBs detected by ExoT alone (***Canela et al., 2019***). Our analyses revealed that 76% of TOP2-mediated DSBs reversed immediately following washout (0 hr washout), with the majority of remaining lesions being converted into protein-free DSBs (***Figure 2F***). The fraction of reversible TOP2ccs was even higher in the BTZ pre-treated cells, as 90% of the initial TOP2-mediated DSBs were reversible immediately following washout (***Figure 2F***). Thus, suppressing proteasomal degradation of TOP2ccs led to increased TOP2cc self-reversal.

Finally, we assessed how the timing of proteasome inhibition impacted the rate at which TOP2-mediated DSBs are resolved. We found that compared to ETO alone, both BTZ pre-treatment and co-treatment significantly reduced the number and persistence of TOP2-mediated DSBs after a 2 hr washout (***Figure 2G and H***). By contrast, BTZ post-treatment slightly increased the number, as well as persistence, of TOP2-mediated DSBs at the same timepoint (***Figure 2G and H***). These results confirmed that the timing of proteasome inhibition relative to ETO treatment determined its impact on the resolution of TOP2-mediated DSBs.

Inhibiting proteolytic degradation of TOP2ccs suppresses DNA end resection of TOP2-mediated DSBs

Since protein-free DSBs persist for a significant period of time following ETO washout, it enabled the study of DNA processing at these sites. We observed clear evidence of 5' to 3' DNA end-resection of TOP2-mediated DSBs immediately after drug washout (***Figure 3A***). After 2 hr post ETO treatment, the extent and intensity of resection had increased (***Figure 3A***), indicating ongoing end-processing at persistent TOP2-mediated DSBs. In ETO-treated cells pre-incubated with BTZ, while there were still low levels of protein-free DSBs present immediately following drug washout (0 hr washout), we could barely detect end-resection at individual breaks (***Figure 3A***). Genome-wide, we were able to identify 1289 resected breaks in ETO-treated cells immediately following washout, while only 496 resected breaks were detected in BTZ pre-treated cells, a 2.5-fold decrease (***Figure 3B***). The end-resection detected in ETO-treated cells genome-wide was reproducible across replicates (***Figure 3—figure supplement 1***). Therefore, inhibiting the proteolytic degradation of TOP2ccs prevented DNA repair-associated nucleases from processing TOP2-mediated DSBs. Moreover, the extent of resection, defined as the maximum distance away from the break summit, was significantly shorter in BTZ-pretreated cells compared to cells treated with ETO alone (***Figure 3C***;

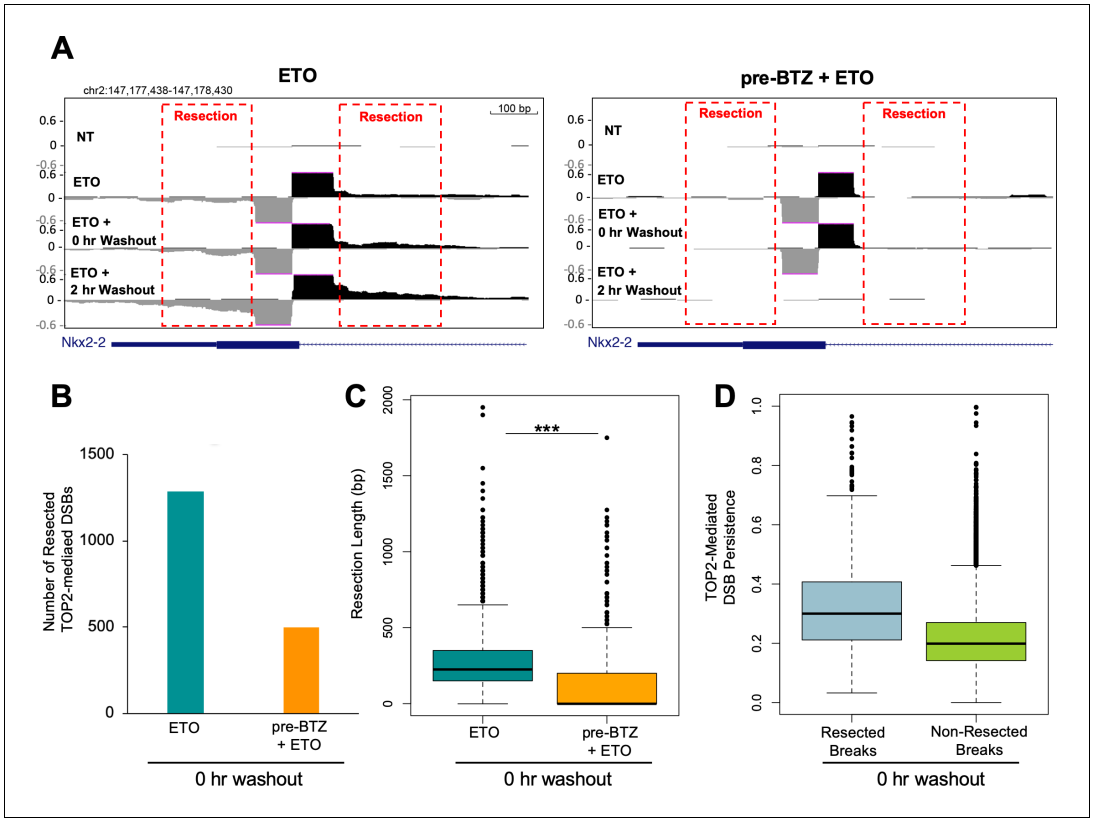


Figure 3. Proteasomal degradation of the TOP2cc is associated with DSB resection. (**A**) Zoomed in UCSC genome browser snapshot of the same TOP2-mediated DSB as in ***Figure 2A*** at three timepoints: 2 hr with 50 μM ETO (ETO), immediately following drug washout (0 hr WO), and at 2 hr post-washout (2 hr WO) (left panel). In right panel, WT primary splenic B-cells were pre-treated for 1 hr with 5 μM BTZ pre-treatment prior to washout. Red box indicates areas with resection signal. (**B**) Bar graph depicting the number of TOP2-mediated DSBs that had a resection signal immediately following drug washout (0 hr washout) in WT primary splenic B-cells treated for 2 hr with 50 μM ETO ±5 μM BTZ pre-treatment. Green bar- 50 μM ETO only sample and orange bar- 5 μM BTZ pre-treatment sample. (**C**) Box-plot quantifying maximum resection distance for all persistent TOP2-mediated DSBs genome-wide immediately following drug washout (0 hr washout) in WT primary splenic B-cells treated for 2 hr with 50 μM ETO ±5 μM BTZ pre-treatment. Green box 50 μM ETO only sample and orange box 5 μM BTZ pre-treatment sample (ETO vs. pre-BTZ + ETO $p<2.2e^{-16}$; statistical significance calculated by student T-test). (**D**) Box-plot of resected (grey) and non-resected (green) TOP2-mediated DSBs genome-wide immediately following drug washout (0 hr washout) in WT primary splenic B-cells treated for 2 hr with 50 μM ETO ± 5 μM BTZ pre-treatment. TOP2-mediated DSB resection values are plotted as a function of their persistence at the 0 hr washout timepoint.
The online version of this article includes the following figure supplement(s) for figure 3:

Figure supplement 1. TOP2-Medated DSB resection is reproducibly captured by END-seq.

Figure 3—figure supplement 1B). Finally, we found that after ETO removal, TOP2-mediated DSBs undergoing nucleolytic processing (resected breaks) tended to be more persistent than non-resected breaks (***Figure 3D***; ***Figure 3—figure supplement 1C***). Thus, the longer a TOP2-mediated DSB persists in the genome, the more likely it will be processed and repaired in a potentially error-prone manner.

Proteasome inhibitor pre-treatment prevents ETO-induced genome instability and cytotoxicity

Enzymatic reversal of TOP2ccs is predicted to be an error-free process. By contrast, the degradation of TOP2ccs followed by DSB repair is potentially more susceptible to error. Nevertheless, repair of ETO-induced DSBs by TDP2-dependent NHEJ might be performed with fidelity (***Gómez-Herreros et al., 2013***) since it involves simple re-ligation of compatible 4 bp overhangs. To directly

address this question, we exposed primary B-cells in G1 to ETO (2 hr, 50 uM) with or without BTZ pre-treatment, or post-treatment. After washing out the drugs, cells were allowed to recover at 37˚C in fresh, drug-free media for 24 hr before they were harvested for mitotic chromosome analysis (***Figure 1A***). Notably, we found that even a short pulse of ETO induced a substantial number and variety of chromosomal aberrations in WT B-cells (~14 aberrations per cell) (***Figure 4A and B***). The majority of aberrations that we observed were dicentric chromosomal fusions and chromosome breaks (***Figure 4A and B***). Post-treatment with MG132 did not mitigate the genotoxic effects of ETO (***Figure 4B***), consistent with the lack of effect it had on ETO-induced γ-H2AX response (***Figure 1G***). However, pre-treating B-cells with MG132 or BTZ completely protected cells from accruing chromosomal aberrations induced by ETO, with less than one aberration detected per cell (***Figure 4A and B***). Therefore, the repair of DSBs that arise from proteasome-mediated degradation of TOP2ccs is error-prone, while preventing TOP2cc processing preserves genome integrity.

Mitotic chromosome analysis is limited in its ability to determine the extent of chromosomal fusion events, as pieces of two or more chromosomes could be ligated together and still appear as a normal chromosome when stained with only DAPI. We consistently detected the presence of elongated chromosomes in mitotic spreads prepared from ETO-treated B-cells, suggesting that these might result from complex fusions involving multiple chromosomal fragments. To further characterize these seemingly intact long chromosomes, we performed spectral karyotyping (SKY) (***Liyanage et al., 1996***). As shown in ***Figure 4C and D***, SKY analysis revealed that ETO-induced chromosomal fusions were extensive and complex. Indeed, these events resembled chromothripsis, in which fragments originating from three or more chromosomes are fused together (***Zhang et al., 2013***). Strikingly, pre-treating B-cells with MG132 before ETO completely protected them from all types of chromosomal fusions (***Figure 4C and D***). These results highlight the severe genotoxic consequences of a transient high dose pulse of ETO and demonstrate that such ETO-induced genome instability is proteasome-dependent.

Consistent with the impact of the proteasome on genome integrity, we found that pre-treatment with MG132, but not post-treatment, mitigated the induction of cell cycle arrest by ETO evidenced by the increased cellular incorporation of EdU after the start of ETO treatment (***Figure 4E***). While proteasome inhibition itself caused a reversible block on S-phase entry, as previously described (***Rastogi and Mishra, 2012***), proteasome inhibitor pre-treatment but not post-treatment, significantly diminished the cytotoxicity of ETO (***Figure 4F***). Taken together, our data indicate that proteasome inhibitor pre-treatment maintains cellular proliferative capacity and long-term viability by blocking ETO-induced genome instability.

Misrepair of DSBs following TOP2 degradation is mediated by multiple DNA repair pathways

To determine if DDR signaling per se is responsible for the error-prone repair of TOP2-associated DSBs, we conducted mitotic spread analyses in $ATM^{-/-}$ and $H2AX^{-/-}$ primary B-cells. We found that these DDR signaling deficient mutants accumulated somewhat higher levels of chromosomal aberrations following ETO treatment compared to WT (***Figure 5A***). Importantly, proteasome inhibitor pre-treatment almost completely abolished ETO-induced damage in WT, as well as in $ATM^{-/-}$ and $H2AX^{-/-}$ B-cells (***Figure 5A***). Thus, proteasome inhibition but not ATM-γ-H2AX driven DDR signaling protects against ETO-induced genome instability.

Both NHEJ and HR are thought to contribute to the repair of ETO-induced DSBs (***Ledesma et al., 2009***; ***Gómez-Herreros et al., 2013***; ***Gómez-Herreros et al., 2017***; ***Hoa et al., 2016***; ***Pommier et al., 2016***). To determine whether classical NHEJ or HR is primarily responsible for the misrepair of DSBs, we tested primary B-cells deficient in the key end joining factor DNA ligase IV ($Lig4^{-/-}$) or the key HR factors BRCA1 ($BRCA1^{\Delta 11}$) and BRCA2 ($BRCA2^{-/-}$). To this end, primary resting (G0) $Lig4^{-/-}$, $BRCA1^{\Delta 11}$ and $BRCA2^{-/-}$ B-cells were first activated with cytokines for 12 hr and 24 hr to drive G1- or S-phase entry, respectively, at which point they were exposed to ETO. After ETO was washed out, B-cells were given an additional 24 hr to recover before mitotic spread analysis. We found that loss of Lig4, BRCA1 or BRCA2 failed to mitigate aberrant chromosomal rearrangements in ETO-treated G1 and cycling B-cells, respectively (***Figure 5B and C***; ***Figure 5—figure supplement 1***). Furthermore, MG132 pre-treatment completely prevented ETO-induced genome instability in $Lig4^{-/-}$ and $BRCA1^{\Delta 11}$ cells (***Figure 5B and C***). Thus, the protective effects of

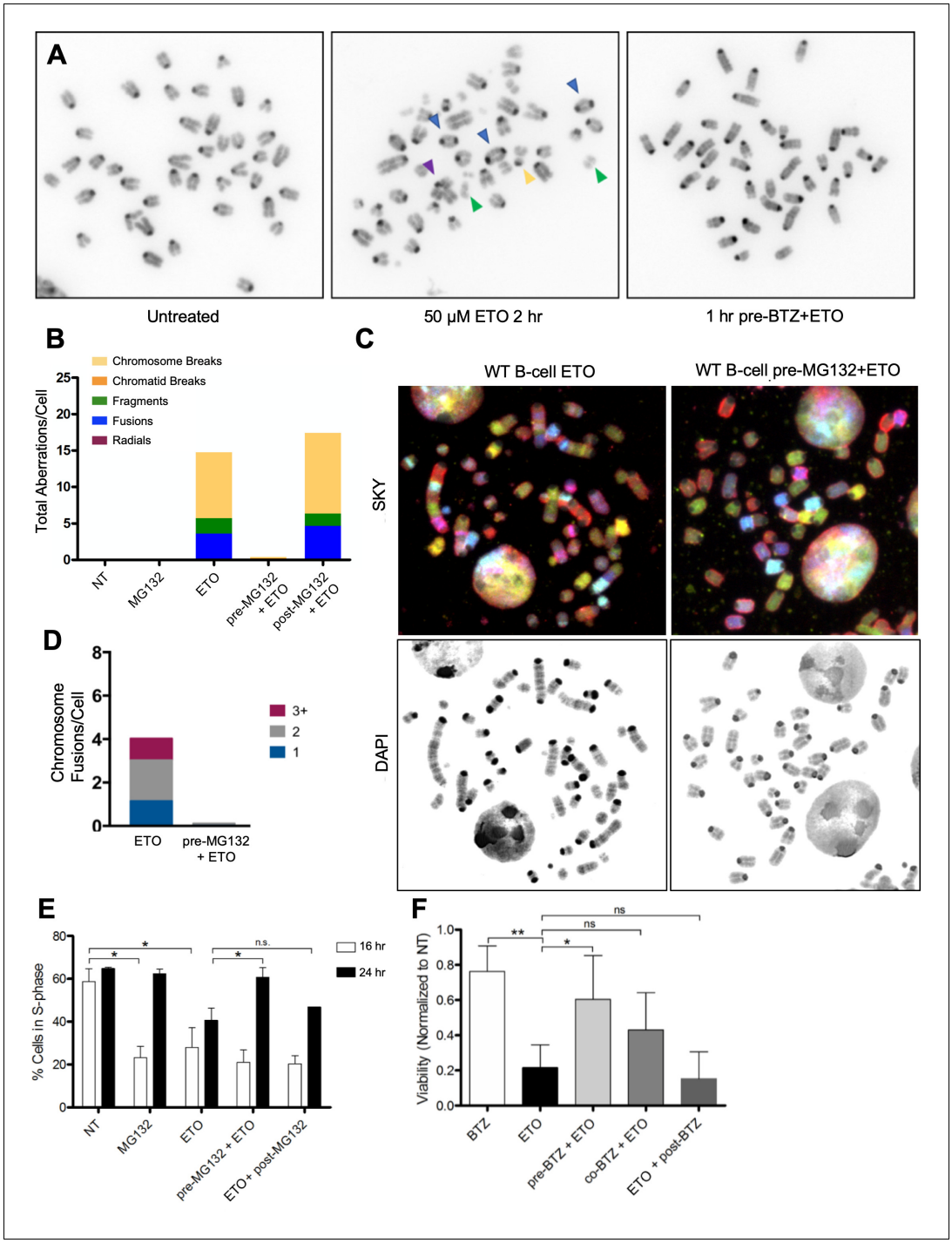


Figure 4. Proteasome inhibition suppresses ETO-induced genome instability and improves cell survival. (**A**) Example of mitotic spreads from WT primary splenic B-cells treated for 2 hr in G1 with 50 µM ETO ± 1 hr 5 µM BTZ pre-treatment, followed by 24 hr recovery in drug free medium prior to harvesting metaphases. Colored arrows indicate the type of aberrations as quantified and described in **B**. (**B**) Analysis of chromosomal aberrations (Chromosome Breaks (yellow), Chromatid Breaks (orange), Fusions (blue), Radials (purple) and Fragments (green)). 50 metaphases were counted for each condition. (**C**) Example of mitotic spreads from WT primary splenic B-cells treated for 2 hr in G1 with 50 µM ETO ± 1 hr 10 µM MG132 pre-treatment, and then harvested 24 hr after washout for Spectral Karyotyping Analysis (SKY) (Left and right top panels). SKY reveals more complex fusions involving multiple chromosomes, compared to the fusions observable by DAPI staining only (Left and right bottom panels). (**D**) Analysis of chromosomal fusion determined by SKY analysis. Chromosome fusions were counted in 35 mitotic

Figure 4 continued on next page

Figure 4 continued

spreads per condition and were broken down by how many chromosomes were involved (1, two or greater than 3 (+)) per fusion event per cell. (**E**) Quantification of percent cells in S phase determined by FACS. WT primary splenic B-cells were treated in G1 for 2 hr with 50 μM ETO ± 1 hr 10 μM MG132 pre-treatment. 16 or 24 hr after washout, cells were pulsed for 30 min with EdU pulse an analyzed by FACS (16 hr NT vs MG132 p=0.012; 16 hr NT v ETO p=0.0485; 24 hr ETO vs pre-MG132 + ETO p=0.049; 24 hr ETO vs. post-MG132 p=0.46; statistical significance calculated by student T-test). (**F**) Viability of WT primary splenic B-cells as determined using the CellTiter-Glo luminescence assay 48 hr following 2 hr of 50 μM ETO treatment ±pre treatment, co-treatment, or post-treatment with 5 μM BTZ (BTZ v ETO p=0.0006, ETO v pre-BTZ + ETO p=0.015, statistical significance calculated by student T-test). During treatment cells were in the G1 phase of the cell cycle.

proteasome inhibition on ETO-induced genome instability is not cell cycle dependent, and neither

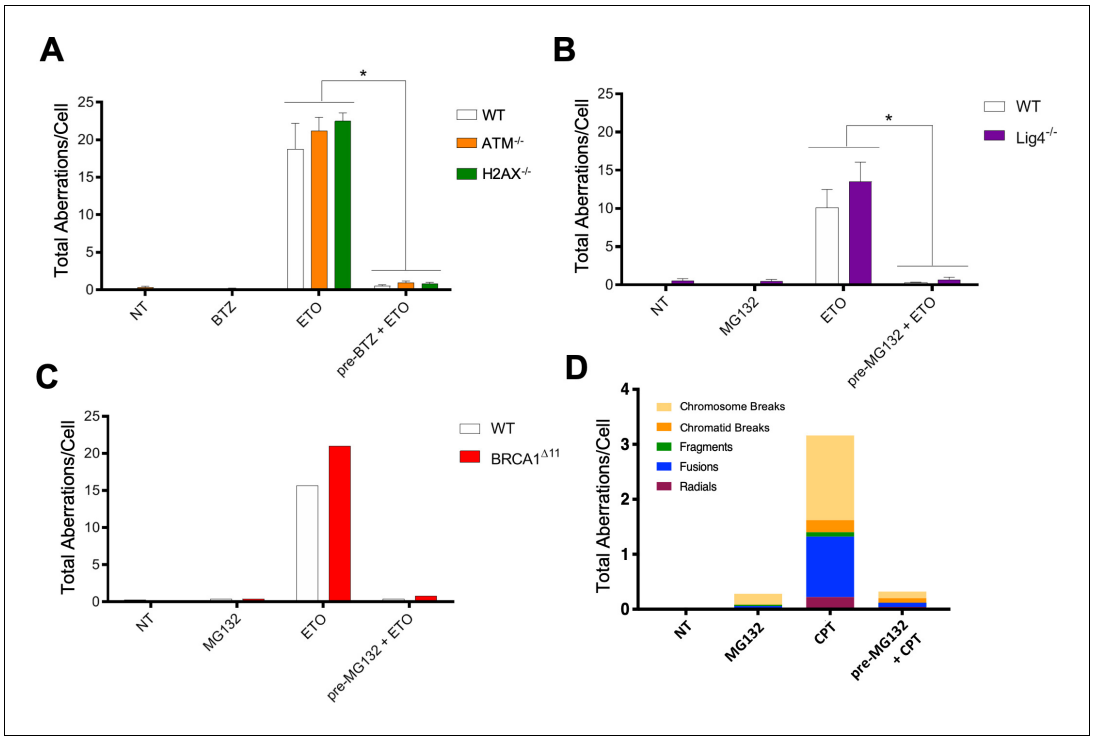


Figure 5. Multiple DNA repair pathways contribute to ETO-induced genome instability. (**A**) Mitotic spread analysis of WT (White), ATM$^{-/-}$ (Orange), and H2AX$^{-/-}$ (Green) primary splenic B-cells treated in G1 for 2 hr with 50 μM ETO ± 1 hr 5 μM BTZ pre-treatment. B-cells were fixed for mitotic spread analysis following a 24 hr recovery at 37˚ C in drug free media. Total chromosomal aberrations were counted in 50 metaphases (N = 3) and averaged for each genotype. (ETO vs. pre-BTZ + ETO: WT, p=0.006; ATM$^{-/-}$, p=0.0004; H2AX$^{-/-}$, p<0.0001; statistical significance calculated by student T-test). (**B**) Mitotic spread analysis of WT (White) and Lig4$^{-/-}$ (Purple) primary splenic B-cells treated in G1 for 2 hr with 50 μM ETO ±1 hr 10 μM MG132 pre-treatment. B-cells were fixed for mitotic spread analysis following a 24 hr recovery at 37˚C in drug free media. Total chromosomal aberrations were counted in 50 metaphases (N = 3) and averaged for each genotype. (ETO vs pre-MG132 + ETO: WT, p=0.0154; Lig4$^{-/-}$, p=0.0075; statistical significance calculated by student T-test). (**C**) Mitotic spread analysis of cycling (24 hr post-activation) WT (White) and BRCA1$^{\Delta 11}$ (Red) primary splenic B-cells treated for 2 hr with 50 μM ETO ± 1 hr 10 μM MG132 pre-treatment. B-cells were fixed for mitotic spread analysis following a 24 hr recovery at 37˚C in drug free media. Total chromosomal aberrations were counted in 50 metaphases per condition. (**D**) Mitotic spread analysis of cycling primary splenic B-cells treated for 2 hr with 12.5 μM CPT ± 1 hr 10 μM MG132 pre-treatment. B-cells were fixed for mitotic spread analysis following a 24 hr recovery at 37˚C in drug free media. Total chromosomal aberrations were counted in 50 metaphases per condition.

The online version of this article includes the following figure supplement(s) for figure 5:

Figure supplement 1. Multiple DNA repair mechanisms contribute to error-prone repair of ETO-induced damage.

classical NHEJ nor HR appears to be strictly required to generate ETO-induced chromosomal aberrations.

Given the lack of clear dependency on canonical DSB repair pathways, we next explored whether Polθ-mediated alternative end joining could be a major contributor of ETO-induced genome instability. Similar to Lig4- and BRCA1/2-deficiency, however, loss of Polθ ($POLQ^{-/-}$) did not significantly attenuate the formation of complex chromosomal rearrangements in ETO-treated cells (***Figure 5—figure supplement 1***). Therefore, multiple DNA repair mechanisms likely contribute to the error-prone repair of ETO-induced DSBs.

Proteasome inhibitor pre-treatment prevents degradation of TOP1 cleavage complexes and camptothecin-induced genome instability

In addition to preventing the degradation of TOP2ccs, previous studies have shown that proteasome inhibitors are able to suppress ATM-γ-H2AX signaling in response to damage induced by TOP1 poisons like Camptothecin (CPT) (***Lin et al., 2008***), which stabilizes TOP1ccs in the same mechanistic way ETO stabilizes TOP2ccs (***Pommier et al., 2016***). We, therefore, assessed if cells pre-treated with proteasome inhibitors were protected from chromosomal aberrations caused by CPT. Our results showed that MG132 pre-treatment largely suppressed CPT-induced chromosomal aberrations (***Figure 5D***), similar to its effects in ETO-treated cells (***Figure 5A–C***). Thus, inhibiting the proteolytic degradation of either TOP1ccs or TOP2ccs prevents subsequent DNA mis-repair that leads to genome instability.

The SUMO-targeted ubiquitin ligase RNF4 promotes proteasome-mediated processing of TOP2ccs

Based on the results presented above, it is clear that the degradation of ETO-stabilized TOP2ccs leads to genome instability, while chemical suppression of proteasomal TOP2cc degradation prevents the accumulation and mis-repair of genotoxic DSBs. Recent work has described a SUMO-ubiquitin (Ub) pathway that recognizes DNA-bound topoisomerase-DNA complexes, wherein the SUMO-targeted E3 ubiquitin ligase (STUbL) RNF4 ubiquitinates TOP2 leading to its degradation (***Sun et al., 2019***). We therefore assessed whether interfering with the cell's ability to sense and recruit the proteasome to trapped TOP2ccs would confer a chemo-protective effect similar to cells treated with BTZ and ETO (***Figure 1B***, ***Figure 4A–F***). To this end, we treated WT and $RNF4^{-/-}$ MEFs (***Hu et al., 2010***) with 10 μM ETO for 1 hr with or without BTZ pre-incubation (1 μM, 1 hr), and probed for γ-H2AX. ETO produced a weakened γ-H2AX response in $RNF4^{-/-}$ MEFs compared to WT MEFs (***Figure 6A***), suggesting that RNF4 facilitates proteasomal degradation of TOP2ccs, which in turn generates γ-H2AX. Notably, while BTZ pre-treatment suppressed the ETO-induced γ-H2AX response in WT MEFs (***Figure 6A***), it did not further reduce the γ-H2AX response in $RNF4^{-/-}$ MEFs (***Figure 6A***), suggesting that RNF4 is epistatic with the proteasome with regard to TOP2cc degradation. To explore this further, we used END-seq to quantify the number and intensity of TOP2-mediated DSBs in ETO treated $RNF4^{-/-}$ and WT primary B-cells. Our results showed that there was a 6-fold reduction of protein-free DSBs detected by END-seq in $RNF4^{-/-}$ B-cells compared to WT (***Figure 6B***, top right), while the initial levels of total TOP2-mediated DNA cleavage were similar between the genotypes (***Figure 6B***, top left). This effect was comparable to what we observed by chemically inhibiting the proteasome with BTZ, which reduced the number of protein-free DSBs by ~5 fold (***Figure 6B***, bottom right). These results indicate that RNF4 ubiquitination is critical for proteasome-mediated processing of TOP2ccs into genotoxic protein-free DSBs.

To assess if the loss of RNF4 conferred resistance to ETO, we determined how RNF4 deletion affected short-term cellular viability and long-term colony formation potential in ETO-treated B-cells and MEFs, respectively. We found that $RNF4^{-/-}$ cells were significantly more viable and retained higher colony formation capacity after a 1 hr pulse of ETO compared to WT counterparts (***Figure 6C and D***). Consistent with increased viability, $RNF4^{-/-}$ B-cells accumulated 60% less chromosomal aberrations after ETO treatment compared to WT B-cells (***Figure 6E***). Taken together, these results indicated that impairing the proteasome response to trapped TOP2ccs reduces the cytotoxicity and genotoxicity of ETO. Notably, pre-treating $RNF4^{-/-}$ B-cells with BTZ further reduced aberrant chromosomal repair (***Figure 6E***), suggesting that RNF4-mediated ubiquitination is likely not the only mechanism by which the proteasome is recruited to TOP2ccs.

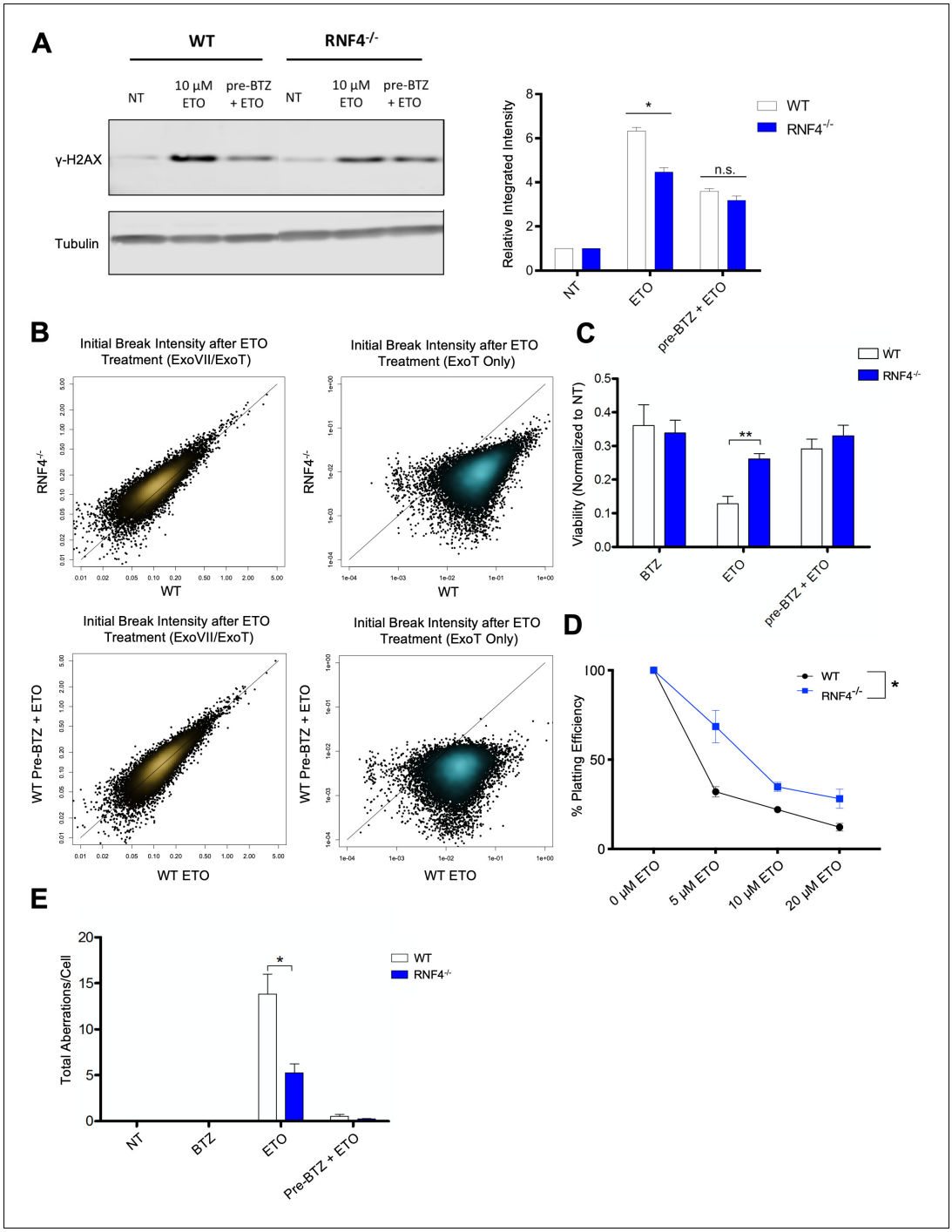


Figure 6. The SUMO-targeted Ubiquitin-Ligase RNF4 functions in proteasome mediated processing of TOP2cc. (**A**) Example γ-H2AX western blot in WT and RNF4$^{-/-}$ MEFs treated with 10 µM ETO for 1 hr ±1 hr 1 µM BTZ pre-treatment (Left Panel). Bar graph quantifying the relative band intensity from γ-H2AX western blots (N = 3) of WT (White) and RNF4$^{-/-}$ (Blue) MEFs treated with 10 µM ETO (Right Panel). (WT ETO vs RNF4$^{-/-}$ ETO p=0.0018; statistical significance calculated by student T-test). (**B**) Scatterplots depicting initial intensity of all TOP2-mediated DSBs (ExoVII/ExoT, Top Left) and initial intensity of protein-free DSBs genome-wide (ExoT, Top Right) induced by 2 hr treatment of WT and RNF4$^{-/-}$ B cells with 50 µM ETO. Scatterplots of all TOP2-mediated DSBs (ExoVII/ExoT, Bottom Left) and protein-free DSBs genome-wide (ExoT, Bottom Right) in WT primary splenic B-cells treated with 50 µM ETO only vs. 50 µM ETO + 5 µM BTZ pre-treatment. (**C**) Viability of WT (White) and RNF4$^{-/-}$ (Blue) primary splenic B-cells, as determined by the CellTiter-Glo luminescence assay 48 hr following 2 hr 50 µM ETO treatment ±1 hr 5 µM BTZ pre-treatment. (**D**) Colony formation assay in WT (White) and RNF4$^{-/-}$ (Blue) MEFs. Cells

Figure 6 continued on next page

Figure 6 continued

were seeded in 6 cm plates at 100, 1000 and 10000 cells/plate and pulsed with 5–20 μM ETO for 1 hr and then left to recover and form colonies for 7 days. The number of colonies from duplicate plates were averaged and plotted at each concentration (N = 3) (WT vs RNF4 p=0.0034; statistical significance calculated by ANOVA). (**E**) Mitotic spread analysis of WT (White) and RNF4$^{-/-}$ (Blue) primary splenic B-cells treated in G1 for 2 hr with 50 μM ETO ± 1 hr 5 μM BTZ pre-treatment. B-cells were fixed for mitotic spread analysis following a 24 hr recovery at 37˚C in drug free media. Total chromosomal aberrations were counted in 50 metaphases (N = 3) and averaged for each genotype. (WT vs. RNF4$^{-/-}$ ETO: p=0.048, statistical significance calculated by student T-test).

Discussion

In this study, we describe a mechanism by which ETO-induced genotoxicity can be abolished by inhibiting proteasome mediated degradation of TOP2 cleavage complexes (TOP2ccs). In the absence of proteolytic degradation, TOP2 remains enzymatically competent and is able to reseal the protein-linked DSB without invoking a DNA damage response (DDR). However, once degradation commences, a TOP2cc is no longer reversible, and the previously hidden DSB become unmasked, triggering a potent DDR signaling response. Following degradation of a TOP2cc, the resultant protein free DSBs appear to engage multiple repair pathways, including error-free NHEJ (***Gómez-Herreros et al., 2013***; ***Gómez-Herreros et al., 2017***; ***Figure 7A***). However, due to the large number of lesions induced by ETO, many DSBs remain unrepaired or mis-repaired which can destabilize the genome. In contrast, by preventing proteasome-mediated unmasking of TOP2-mediated DSBs,

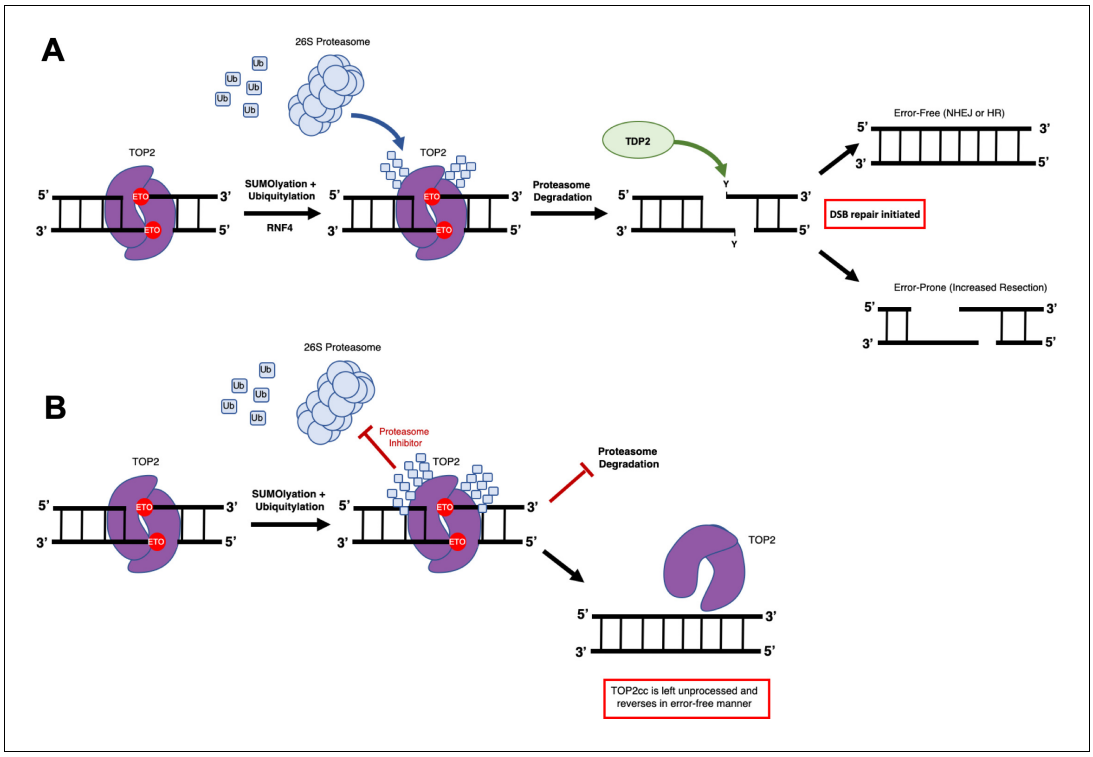


Figure 7. Model of proteasome mediated repair of TOP2ccs. (**A**) ETO-stabilized TOP2ccs are targeted for proteasomal degradation by the 26S proteasome via both SUMOylation and Ubiquitination, mediated by the SUMO-targeting ubiquitin ligase RNF4 and possibly other enzymes. Upon degradation of the TOP2 protein within the TOP2cc, the remaining 5'-phosphotyrosyl bonds must be processed by TDP2, generating clean protein-free DSBs with overhangs that can be repaired in a manner that is not always error-free. (**B**) If the proteasome is not recruited due to loss of RNF4-mediated TOP2cc polyubiquitylation or its activity is inhibited chemically, the TOP2cc is not recognized by the DNA repair machinery and the TOP2 protein retains full enzymatic ability to reverse itself once ETO is washed out.

error-prone repair is not engaged, thereby facilitating the reversal of TOP2ccs through completion of the enzymes' catalytic cycle upon drug removal, and preserving genome integrity (***Figure 7B***).

Both proteasome inhibitors (e.g. BTZ) and topoisomerase poisons (e.g. ETO, doxorubicin) are used clinically as therapeutic anti-cancer agents, sometimes in combination (***Cowell and Austin, 2012***; ***Dittus et al., 2018***; ***Manasanch and Orlowski, 2017***; ***Nitiss, 2009b***; ***Palumbo et al., 2008***; ***Thomas et al., 2017***). Notably, prior studies have shown that proteasome inhibitors potentiate the cytotoxicity of TOP2 poisons (***Ceruti et al., 2006***; ***Destanovic et al., 2018***; ***Dittus et al., 2018***; ***Lee et al., 2016***; ***von Metzler et al., 2009***), which at first glance appeared to be at odds with our observation that proteasome inhibition enhances genome stability and survival of ETO-treated cells. We addressed this apparent discrepancy by showing that the timing of proteasome inhibition has a remarkable impact on the outcome of ETO treatment. Inhibiting proteasomal function prior to or during ETO administration significantly improved genome stability by minimizing DSB mis-repair associated with TOP2 proteolysis. However, inhibiting the proteasome after ETO treatment did not effectively protect cells from ETO-induced genome instability, presumably because proteasome-mediated degradation of TOP2ccs is an extremely fast process. Taken together, our study suggests that caution needs to be taken when scheduling therapeutic regimens combining topoisomerase poisons with proteasome inhibitors.

Our analysis also revealed that TOP2-mediated DSBs, especially ones that are not promptly repaired, undergo extensive DNA end-resection. We interpreted this as an indicator of active DNA repair occurring at TOP2ccs that have been partially or completely degraded by the proteasome. By contrast, when proteasomal degradation was inhibited by BTZ, we could only detect trace levels of DNA end-resection. This suggests the possibility that MRE11, which is required for DNA end-resection, may not recognize unprocessed TOP2. Similar to TOP2, the meiotic recombination initiator SPO11 forms protein-linked DSBs, which are thought to trigger MRE11 nuclease removal of SPO11 (***Neale et al., 2005***). However, recent studies reveal that mammalian meiotic cells accumulate significant levels of DNA-bound SPO11, which could, in principal be partially proteolyzed (***Paiano et al., 2020***). While the mechanism by which TOP2 (and perhaps SPO11) shields its associated DSB from early detection by DDR surveillance factors is unclear, our results highlight that γ-H2AX is only a reliable marker for proteolytically degraded TOP2ccs. In contrast, END-seq can be used to distinguish intact TOP2ccs, partially proteolyzed TOP2ccs, as well as fully processed protein-free DSBs.

Finally, we found that the SUMO-targeted ubiquitin ligase RNF4 played a critical role in recruiting the proteasome to degrade TOP2ccs, confirming a report by another group (***Sun et al., 2019***). The loss of RNF4 afforded cells significant protection from the genotoxic effects of ETO, further highlighting how impairing proteasome-mediated degradation of TOP2ccs can preserve genome integrity. However, RNF4 deletion did not prevent ETO-induced chromosomal aberrations or blunt the γ-H2AX response as efficiently as chemical proteasome inhibition. This indicates the likely existence of additional ubiquitin ligases that are functionally redundant to RNF4. Nevertheless, our data suggests that RNF4 could be potentially used as a biomarker to predict the effectiveness of chemotherapeutic regimens that incorporate topoisomerase poisons. In addition, de-regulation of RNF4 may represent a potential mechanism for tumors to acquire resistance to topoisomerase poisons by favoring TOP2cc reversal over degradation.

Materials and methods

Key resources table

Reagent type (Species or Source)	Designation	Source	Identifier	Add. info
Antibody	Anti-phospho-Histone H2A.X (Ser139) (mouse monoclonal)	Millipore	Cat# JBW301; RRID:AB_309864	(1:5000 - 1:10000)
Antibody	Anti-Tubulin (mouse monoclonal)	Sigma	Cat# T-5168 RRID:AB_477585	(1:5000)
Antibody	Anti-p53 (1CA2) (mouse monoclonal)	Cell Signaling	Cat# 2524 RRID:AB_331743	(1:2000)

Continued on next page

Continued

Reagent type (Species or Source)	Designation	Source	Identifier	Add. info
Antibody	Anti-Vinculin (rabbit polyclonal)	Cell Signaling	Cat# 4650 RRID:AB_10559207	(1:2000)
Antibody	Anti-Top2b (rabbit polyclonal)	Novus	Cat# NB100-10842 RRID:AB_1522570	(1:5000)
Antibody	Anti-Rabbit Li-Cor (goat polyclonal)	Li-Cor	Cat# 926–32211; RRID:AB_621843	(1:10000)
Antibody	Anti-mouse IgG1 Alexa 488 (goat polyclonal)	Invitrogen	Cat# A-21121; RRID:AB_141514	(1:10000)
Antibody	Anti-H2AX (rabbit polycolonal)	Abcam	Cat# ab11175 RRID:AB_297814	(1:5000)
Chemical compound, Drug	Etoposide	Sigma	Cat# E1383	10–50 µM
Chemical compound, Drug	Bortezomib	Millipore	Cat# 179324-69-7	1–5 µM
Chemical compound, Drug	MG132	Sigma	Cat# 1211877-36-9	10 µM
Chemical compound, Drug	Camptothecin	Sigma	Cat# 7689-03-4	12.5 µM
Chemical compound, Drug	Ixaomib	Selleckchem	Cat# S2180	10 µM
Chemical compound, Drug	Eponomicin	Sigma	Cat# 134381-21-8	2 µM
Chemical compound, Drug	Doxycycline hyclate	Sigma	Cat# D9891	1 µg/ml
Chemical compound, Drug	Imatinib mesylate (STI571)	Selleckchem	Cat# S1026	
Cell Line (*M. musculus*)	pre-B cell lines	***Bredemeyer et al., 2006***		
Cell Line (*M. musculus*)	RNF4$^{-/-}$ MEFs	Generated by Gary Lyons, ***Hu et al., 2010***		
Biological Sample (*M. musculus*)	Primary splenic B-cells		Freshly isolated from *M. musculus* spleen	
Strain (*M. musculus*)	C57BL/6NCr mice (WT)	Charles River	Strain code# 027	
Strain (*M. musculus*)	Conditional BRCA1-Δ11f/f CD19-cre expressing mice	NCI mouse repository		
Strain (*M. musculus*)	H2AX$^{-/-}$ mice	***Celeste et al., 2002***		
Strain (*M. musculus*)	Conditional Lig4f/f expressing ERT2-Cre mice	Provided by P. Mckinnon		
Strain (*M. musculus*)	ATM$^{-/-}$ mice	Provided by A. Wynshaw-Boris		
Strain (*M. musculus*)	Conditional RNF4 f/f mice	Provided by S. Bunting		
Strain (*M. musculus*)	Conditional BRCA2 f/f CD19-cre expressing mice	Provided by S. Sharan		
Strain (*M. musculus*)	PolQ$^{-/-}$ mice	Provided by A. D'Andrea		
Recombinant DNA reagent	YFP^{u} Degron reporter (plasmid)	Provided by A. Weissman ***Bence et al., 2005***; ***Bence et al., 2001***		(0.5 ng/µL)

Continued on next page

Continued

Reagent type (Species or Source)	Designation	Source	Identifier	Add. info
Peptide, recombinant protein	Lipopolysaccharide (LPS)	Sigma	Cat# L-2630	25 µg/ml
Peptide, recombinant protein	IL-4 from mouse, Interleukin-4, recombinant	Sigma	Cat# I-1020	5 ng/ml
Peptide, recombinant protein	RP105	BD Biosciences	Cat # 552128	0.5 µg/ml
Peptide, recombinant protein	Cy3-labeled (CCCTAA) peptide nucleic acid probe	PNA Bio	Cat# F1002	
Peptide, recombinant protein	Puregene Proteinase K solution	Qiagen	Cat# 19133	170 µL
Peptide, recombinant protein	Puregene RNaseA	Qiagen	Cat# 19101	50 µL
Peptide, recombinant protein	T4 DNA polymerase	NEB	Cat# M0203L	15 U
Peptide, recombinant protein	Klenow fragment	NEB	Cat# M0210M	5 U
Peptide, recombinant protein	T4 polynucleotide kinase	NEB	Cat# M0201L	15 U
Peptide, recombinant protein	Klenow exo- fragment	NEB	Cat# M0212M	15 U
Peptide, recombinant protein	Quick Ligase	NEB	Cat# M2200L	
Peptide, recombinant protein	USER enzyme	NEB	Cat# M5508L	
Peptide, recombinant protein	beta-agarase I	NEB	Cat# M0392S	1.5 µL
Commercial Assay Kit	CHEF Mammalian Genomic DNA plug kit	Bio-Rad	Cat# 1703591	
Commercial Assay Kit	Click-IT EdU Alexa Fluor 488 Flow Cytometry Assay Kit	ThermoFisher	Cat# C10425	
Commercial Assay Kit	KAPA Library Quantification Kit	Kapa Biosciences	Cat# KK4824	
Commercial Assay Kit	Celltiter-Glo Cell Viability Assay Kit	Promega	Cat# G7571	
Other	Streptavidin beads MyOne C1	ThermoFisher	Cat# 650–01	35 µL
Other	Anti-CD43 (Ly-48) MicroBeads (mouse)	Miltenyi Biotech	Cat# 130-049-80	
Other	Glycogen (20 mg/ml)	Roche	Cat# 10901393001	1 µL
Other	Kapa HiFi Hot Start Ready mix	Kapa Biosciences	Cat# KK2502	

Continued on next page

Continued

Reagent type (Species or Source)	Designation	Source	Identifier	Add. info
Other	Agencourt AM Pure XP beads	Beckman Coulter	Cat# A63881	
Other	IrysPrep Lysis Buffer	BioNano Genomics	Cat# 20270	
Other	X-tremeGene 9 Transfection Reagent	Sigma	Cat # XTG9-RO	
Sequence-based reagent	END-seq hairpin adaptor 1, 5′-Phos-GATCGGAAGAGCGTCGTGT AGGGAAAGAGTGUU[Biotin-dT] U[Biotin-dT]UUACACTCTTTCCCTACACGA CGCTCTTCCGATC*T-3′	**Canela et al., 2016**	END-seq sequence adapter	
Sequence-based reagent	END-seq hairpin adaptor 2, 5′-Phos-GATCGGAAGAGCACACGTC UUUUUUUUAGACGTGTGCT CTTCCGATC*T-3′	**Canela et al., 2016**	END-seq sequence adapter	
Sequence-based reagent	TruSeq barcoded primer p5, AATGATACGGCGACCACCGAGATCTA CACNNNNNNNNACACTCTTTCCCTA CACGACGCTCTTCCGATC*T	**Canela et al., 2019**	END-seq sequence primer	
Sequence-based reagent	TruSeq barcoded primer p7, CAAGCAGAAGACGGCATACGAGA NNNNNNNGTGACTGGAGTTCAG ACGTGTGCTCTTCCGATC*T	**Canela et al., 2019**	END-seq sequence primer	
Software	Prism (7)	GraphPad	https://www.graphpad.com/scientific-software/prism/ RRID:SCR_002798	
Software	Bowtie 1.1.2	**Langmead et al., 2009**	https://sourceforge.net/projects/bowtie-bio/files/bowtie/1.1.2/ RRID:SCR_005476	
Software	MACS 1.4.3	**Zhang et al., 2008**	https://pypi.python.org/pypi/MACS/1.4.3	
Software	UCSC database	**Karolchik et al., 2004**	https://genome.ucsc.edu RRID:SCR_005780	
Software	UCSC Genome Browser	**Kent et al., 2002**	https://genome.ucsc.edu RRID:SCR_005780	
Software	Bedtools	**Quinlan and Hall, 2010**	https://github.com/arq5x/bedtools2 RRID:SCR_006646	
Software	Samtools	**Li et al., 2009**	https://github.com/samtools/samtools RRID:SCR_002105	
Software	Knight-Ruiz algorithm	**Rao et al., 2014**	https://academic.oup.com/imajna/article-lookup/DOI: 10.1093/imanum/drs019	
Software	R	**R Development Core Team (2008)**	https://www.r-project.org/ RRID:SCR_001905	
Software	FlowJo (10.1)	FlowJo LLC	https://www.flowjo.com/ RRID:SCR_008520	
Software	Metapher ISIS	Metapher Systems		
Software	hiSKY 7.2.7	ADS Biotech		

Mice

C57BL/6 WT (NCI mouse repository), POLQ$^{-/-}$ (provided by A. D'Andrea), H2AX$^{-/-}$ (**Celeste et al., 2002**), ATM$^{-/-}$ (provided by A. Wynshaw-Boris) and conditional Lig4$^{-/-}$ (Lig4$^{f/f;ERT2\text{-}Cre}$, provided by P. McKinnon), BRCA1$^{\Delta 11}$BRCA1$^{f/f;CD19Cre}$, **Zong et al. (2019)**, BRCA2$^{-/-}$ (BRCA2$^{f/f;CD19Cre}$, provided by

S. Sharan) and RNF4$^{-/-}$ (RNF4$^{f/f;CD19Cre}$) mice between 8 and 18 weeks of age were used to prepare single cell suspensions of primary splenic B-cells. All animal experiments were approved by the NCI Animal Care and Use Committee (Protocol Numbers: EIB-064–3 and 17–042).

Cell culture methods

Mature resting B-cells were isolated from mouse spleen with anti-CD43 MicroBeads (Miltenyi Biotech). B-cells were activated with LPS (25 µg/ml; Sigma), IL-4 (5 ng/ml; Sigma-Aldrich) and RP105 (0.5 µg/ml; BD Biosciences) for 12 hr as described (***Barlow et al., 2013***; ***Callén et al., 2007***). Lymphocyte separation media (Corning) was used after harvesting the B-cells to separate dead cells from live cells prior to use for END-seq protocol. Cellular Viability was measured using CellTiter-Glo Cell Viability Assay Kit (Promega) according to manufactures directions. Luminescence was quantified using a Tecan Infinite M200 Pro plate reader. B-cells were pre-treated with proteasome inhibitor, either 10 µM MG132 (Sigma-Aldrich) or 5 µM Bortezomib (Millipore) for 1 hr then co-treated for an additional 2 hr with 50 µM Etoposide (E1383, Sigma-Aldrich). Following the ETO pulse, the cells were either harvested in the presence of the drugs, or washed three times in ice-cold drug free media (3 × 5 min spin at 1500 rpm and 4˚C to pellet B-cells between washes; 15–20 min total time to complete washout), and returned to 37˚C to recover in fresh drug free media, until they were harvested at various timepoints post-washout for END-seq, western blot, Immunofluorescence, ICE assay, or for mitotic spread analysis as described in ***Figure 1A***. Asynchronous WT B-cells were treated 24 hr post-activation for 2 hr with 50 µM ETO or 12.5 µM Camptothecin (Sigma-Aldrich) ±1 hr 10 µM MG132 pre-treatment, then fixed following a 24 hr recovery in drug free media for mitotic spread analysis. Abelson-transformed pre-B cells (***Bredemeyer et al., 2006***) were retrovirally transduced with the tetracycline-inducible ER-AsiSI, pTRE3G-HA-ER-AsiSI as previously described (***Callen et al., 2013***). WT and RNF4$^{-/-}$ MEFs (generated by Gary Lyons) (***Hu et al., 2010***) were cultured in DMEM (Invitrogen) + 10% FBS and 1% PenStrep (Invitrogen). MEFs were treated with 5–20 µM ETO for 1 hr ±1 µM BTZ 1 hr pretreatment. For END-seq experiments that used Zinc Finger Nuclease (ZFN) spike-in normalization, induction of ZFN was done in G1-arrested Lig4$^{-/-}$ pre B cells (***Bredemeyer et al., 2006***) by treating with imatinib for 48 hr and 1 µg/ml Doxycycline during the last 24 hr as described (***Canela et al., 2016***; ***Bredemeyer et al., 2006***).

WT and Lig4-/- mouse pre B cell lines were provided by B. Sleckman. WT and RNF4$^{-/-}$ mouse embryonic fibroblasts (MEFs) were provided by S. Bunting. These cell lines are negative for mycoplasma.

γ-H2AX and pan-p53 western blot

Western blotting for γ-H2AX and pan-p53 was performed using γ-H2AX antibody (Millipore) at 1:5000 dilution and p53 antibody (Cell Signaling) at 1:2000 dilution. Antibodies recognizing tubulin (Sigma-Aldrich, 1:5000), H2AX (Millipore, 1:5000) and vinculin (Cell Signaling, 1:2000) were used to control for protein loading. Image analysis was done using Li-Cor Odyssey CLx to quantify band intensity.

In vitro of complex (ICE) Assay

Topoisomerase II DNA-protein complexes (TOP2-DPCs) were isolated and detected using in vivo complex of enzyme (ICE) assay as previously described (***Anand et al., 2018***). Briefly, 5 million cells were lysed in sarkosyl solution (1% w/v) after treatment. Cell lysates were sheared through a 25 g 5/8 needle (10 strokes) to reduce the viscosity of DNA and layered onto CsCl solution (150% w/v), followed by centrifugation in NVT 65.2 rotor (Beckman coulter) at 42,000 RPM for 20 hr at 25˚C. The resulting pellet containing nucleic acids and TOP2-DPCs was obtained and dissolved in TE buffer. The samples were subjected to immunoblotting (slot blot) with anti-mouse TOP2β antibody (Novus). 2 µg of DNA is applied per sample. TOP2-DPCs were quantified by densitometric analysis using ImageJ.

Metaphase spread, FISH and SKY analysis

5 million cells were harvested 24 hr after drugs were washed out for metaphase analysis as described (***Callén et al., 2007***). Quantitative FISH analysis using a Cy3-labeled (CCCTAA) peptide nucleic acid probe (Applied Biosystems) was performed as described previously (***Callén et al., 2007***). Telomere

length measurements were performed on least 15 metaphases for each cell type. DAPI chromosome and Cy3 telomere images were acquired with a constant exposure time that ensured all captured fluorescent signals were within the linear range using Metafer Software on a Zeiss Axios Imager Z2 Microscope. Image analysis was done in the Metafer ISIS software. SKY was performed and analyzed as previously described, using the hiSKY 7.2.7 Software (ASI) on a Leica DMRXA Microscope (***Liyanage et al., 1996***). A minimum of 35 metaphases were imaged and analyzed.

YFP^{u}degron quantification

WT eHAP cells were transfected for 48 hr using Xtreme Gene 9 transfection reagent with YFP^{u} Degron reporter (0.5 ng/µL) as described (***Bence et al., 2005***; ***Bence et al., 2001***). eHAP cells were then treated 48 hr post-transfection with 1 µM BTZ and 5 µM MG132 for 2 hr to mimic the total amount of time the proteasome inhibitors are incubated in primary B-cells, after which the cells were washed three times and left to recover at 37˚C for up to 18 hr. Cells were fixed at 2 hr, 6 hr, and 18 hr after washout with 4% PFA and YFP expression was measured on a FACS CantoII (BD biosciences).

EdU staining

To measure DNA synthesis, primary B-cell cultures were stimulated for 12 hr, 24 hr, 28 hr or 36 hr pulsed with 10 µM of EdU (5-ethynyl-2'-deoxyuridine) for 30 min at 37˚C and stained using the Click-IT EdU Alexa Fluor 488 Flow Cytometry Assay Kit according to the manufacturer's specifications (Thermo Fisher C10425). Samples were acquired on a FACS CantoII (BD biosciences).

Immunofluorescence

Immunofluorescence staining with γ-H2AX antibody (Millipore, 1:10,000) was performed in parallel in WT pre B cells (***Bredemeyer et al., 2006***) to verify DSB induction.

END-seq

Single cell suspensions of primary B-cells (15–20 million) were untreated or treated with drugs to inhibit the proteasome and/or etoposide as indicated in the cell culture methods section. Primary B-cells were washed twice in cold PBS and resuspended in cold cell suspension buffer (Bio-Rad CHEF Mammalian Genomic DNA plug kit), equilibrated for 5 min at room temperature, mixed with 2% melted CleanCut agarose (Bio-Rad CHEF Mammalian Genomic DNA plug kit) prewarmed at 37˚C for a final concentration of 0.75%, and transferred immediately into plug molds and let them solidify at 4˚C for 20 min. Embedded cells were lysed and digested using Proteinase K (50˚C, 1 hr then 37˚C for 7 hr). Etoposide and drugs were maintained at the same experimental concentrations in all the steps until digestion with Proteinase K. Plugs were rinsed in TE buffer and treated with RNaseA at 37˚C, 1 hr. The next enzymatic reactions were performed with the DNA in agarose plugs to prevent shearing. DNA ends were blunted for 1 hr at 37˚C with Exonuclease VII (ExoVII) followed by Exonuclease T (ExoT) (NEB) for 1 hr at 25˚C to detect all TOP2-mediated DSBs (protein-linked and protein-free), or only with Exonuclease T (NEB) for 1 hr at 25˚C to detect only protein-free DSBs. After blunting, A-tailing was performed to attach dA to the free 3'-OH, followed by ligation of ''END-seq hairpin adaptor 1,'' listed in reagents section, using NEB Quick Ligase. Agarose plugs were then melted and dialyzed, and DNA was sonicated to a median shear length of 170 bp using Covaris S220 sonicator for 4 min at 10% duty cycle, peak incident power 175, 200 cycles per burst, at 4˚C. DNA was ethanol-precipitated and dissolved in 70 ml TE buffer. Biotinylated DNA was isolated using MyOne Streptavidin C1 Beads (Thermo Fisher #650–01), followed by end repair (dNTPs, T4 polymerase (NEB), Klenow (NEB), T4 PNK) and dA-tailing (Klenow exo- (NEB), dATP). The second end was ligated to ''END-seq hairpin adaptor 2'' using NEB Quick Ligase. Hairpins were digested using USER (NEB), and the resulting DNA fragments were PCR amplified using TruSeq barcoded primer p5, AATGATACGGCGACCACCGAGATCTACACNNNNNNNNACACTCTTTCCCTACACGACGCTCTTCCGATC*T and TruSeq barcoded primer p7, CAAGCAGAAGACGGCATACGAGANNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T, (NNNNNNNN represents barcode a * a phosphothiorate bond) listed in reagents. PCR fragments were isolated by size selection from agarose gel, selecting 200–500 bp fragments followed by DNA purification using QIAquick Gel Extraction Kit. Libraries were quantified using KAPA Library Quantification Kit and sequenced using

Illumina NextSeq 500 or 550. A detailed END-seq protocol can be found in ***Canela et al. (2017)***, as well as more information about the ExoVII and ExoT processing of TOP2-mediated DSBs in ***Canela et al. (2019)***.

Quantification and statistical analysis

Genome alignment

END-seq single end reads were aligned to the mouse genome (GRCm38p2/mm10) using Bowtie v1.1.2 (***Langmead and Salzberg, 2012***) with parameters (-n 3 k 1 l 50) and alignment files were generated and sorted using SAMtools (***Li et al., 2009***) and BEDtools (***Quinlan and Hall, 2010***).

Peak calling

Peaks were called using MACS 1.4.3 (***Zhang et al., 2008***) with the parameters: –nolambda,–nomodel and–keep-dup = all (keep all redundant reads). For ETO-treated END-seq data peak calling, we used the corresponding non-treated samples as control, keeping >10 fold-enriched peaks. The peaks within blacklisted regions (https://sites.google.com/site/anshulkundaje/projects/blacklists) were filtered. For primary B-cell data, in order to perform a direct comparison of the breaks across experiments, peak calling was done on the END-seq (ExoT+ExoVII) data presented in ***Figure 2*** (carried out after 2 hr of ETO) and in ***Figure 2—figure supplement 1***. As the END-seq signal is very reproducible between replicates (***Figure 2—figure supplement 1***), all of the analyses thereafter including persistence, reversibility and resection presented in ***Figures 2*** and ***3***, utilized the peak dataset from replicate 1. Analysis using the peak dataset from replicate two is presented in ***Figure 2—figure supplement 1*** and ***Figure 3—figure supplement 1***.

END-seq data analysis

For comparisons between different genotypes and treatments, END-seq signal was calculated, as RPKM, within END-seq peaks, for the corresponding cell type. For END-seq data, all reads were counted. For END-seq data, where cells with ZFN break were spiked-in to the library at a 1:40 (2.5%) ratio, the RPKM value for each peak was divided by the signal around the spiked-in breaks and then multiplied by 2.5, to get values as cell-percentage.

Resection quantification

To quantify the width of maximum resection endpoint, a sliding window containing 10 50 bp bins was used to start from the summit of each break, out to 2.5 kb on either side of the summit to detect the END-seq signal around the break summit. Maximum END-seq signal for 50 bp bins within 2.5 kb ~5 kb around break summit was used as background. When more than half of the bins within this sliding window had an RPKM values equal to or lower than the background, then the last bin within the window which had a detectable signal over background is regarded as the maximum resection endpoint from the break summit. Only breaks having >100 bp resection signal were considered for the comparison of resection length in ***Figure 3***.

Classification of TOP2-mediated DSBs

The fraction of the different TOP2-mediated DSB species (Reversible TOP2cc, Irreversible TOP2cc and protein-free DSB) were calculated as follows: First, ExoT END-seq signal was divided by the END-seq signal (using ExoVII+ExoT) for all TOP2-mediated DSB sites (determined by peak calling as described above), to obtain the protein-free DSB fraction ([DSB]) (***Chase and Richardson, 1974***). Similarly, TOP2cc fraction, [TOP2CC], was defined as [TOP2CC]=1-[DSB] fraction for each lesion. To calculate the reversible TOP2cc fraction ($[TOP2CC]_R$), we subtracted the data obtained from the END-seq ETO washout experiment (see above), from the data obtained from END-seq ETO experiment without washout, to get $[TOP2CC]_R$. Finally, the irreversible TOP2cc fraction ($[TOP2CC]_I$) was defined as $[TOP2CC]_I$ = 1-[DSB]-$[TOP2CC]_R$. A more detailed description of this analysis can be found in ***Canela et al. (2019)***.

Data visualization

Aligned-reads bed files were first converted to bedgraph files using bedtools genomecov (***Quinlan and Hall, 2010***) following by bedGraphToBigWig to make a bigwig file (***Kent et al., 2002***).

Visualization of genomic profiles was done by the UCSC browser (***Kent et al., 2002***). Genome browser profiles were normalized to the library size (RPM) and spike-in.

Statistical analysis

Statistical analysis was performed using R version 3.5.0 (http://www.r-project.org). The statistical tests are reported in the figure legend and main text.

Acknowledgements

We thank Keith Caldecott for comments on the manuscript, Allan Weissman for providing the YFP^{u} Degron, Rachel Burga for support with statistical analysis and generating the graphic schematic and Jennifer Wise and Kelly Smith for assistance with animal work. The AN laboratory is supported by the Intramural Research Program of the NIH, a Department of Defense awards W81XWH-16-1-599) and W81XWH-19-1-0712 and an NIH Intramural FLEX Award.

Additional information

Funding

Funder	Grant reference number	Author
National Institutes of Health	Intramural Research Program	André Nussenzweig
Ellison Medical Foundation	Senior Scholar in Aging Award AG-SS- 2633-11	André Nussenzweig
U.S. Department of Defense	Idea Expansion Award W81XWH-15-2-006	André Nussenzweig
U.S. Department of Defense	Idea Breakthrough Award W81XWH-16-1-599	André Nussenzweig
Alex's Lemonade Stand Foundation		André Nussenzweig
National Institutes of Health	Intramural FLEX Award	André Nussenzweig

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Nicholas Sciascia, Conceptualization, Data curation, Software, Formal analysis, Validation, Investigation, Visualization, Methodology, Project administration; Wei Wu, Data curation, Software, Formal analysis, Visualization, Methodology; Dali Zong, Formal analysis, Investigation, Methodology, Writing - original draft, Writing - review and editing; Yilun Sun, Formal analysis, Investigation, Visualization; Nancy Wong, Investigation; Sam John, Writing - original draft, Writing - review and editing; Darawalee Wangsa, Data curation, Formal analysis, Investigation, Visualization; Thomas Ried, Formal analysis, Methodology; Samuel F Bunting, Resources, Provided the RNF4 conditional knockout mice; Yves Pommier, Conceptualization, Methodology; André Nussenzweig, Conceptualization, Resources, Supervision, Funding acquisition, Validation, Methodology, Writing - original draft, Project administration, Writing - review and editing

Author ORCIDs

Nicholas Sciascia https://orcid.org/0000-0003-4169-4929
André Nussenzweig https://orcid.org/0000-0002-8952-7268

Ethics

Animal experimentation: All mouse breeding and experimentation followed protocols approved by the National Institutes of Health Institutional Animal Care and Use Committee (Protocol Numbers: EIB-064-3 and 17-042).

Decision letter and Author response
Decision letter https://doi.org/10.7554/eLife.53447.sa1
Author response https://doi.org/10.7554/eLife.53447.sa2

Additional files

Supplementary files

- Supplementary file 1. Table of experimental replicates for each figure panel.
- Transparent reporting form

Data availability

Sequencing data has been deposited in GEO under the accession code GSE140372.

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Sciascia N, Wu W, Zong D, Sun Y, Wong N, Wangsa D, John S, Ried T, Bunting S, Pommier Y, Nussenzweig A	2019	Suppressing Proteasome Mediated Processing of Topoisomerase II DNA-Protein Adducts Preserves Genome Integrity	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE140372	NCBI Gene Expression Omnibus, GSE140372

References

An WG, Hwang SG, Trepel JB, Blagosklonny MV. 2000. Protease inhibitor-induced apoptosis: accumulation of wt p53, p21WAF1/CIP1, and induction of apoptosis are independent markers of proteasome inhibition. *Leukemia* **14**:1276–1283. DOI: https://doi.org/10.1038/sj.leu.2401812, PMID: 10914553

Anand J, Sun Y, Zhao Y, Nitiss KC, Nitiss JL. 2018. Detection of topoisomerase covalent complexes in eukaryotic cells. *Methods in Molecular Biology* **1703**:283–299. DOI: https://doi.org/10.1007/978-1-4939-7459-7_20, PMID: 29177749

Aras B, Yerlikaya A. 2016. Bortezomib and etoposide combinations exert synergistic effects on the human prostate Cancer cell line PC-3. *Oncology Letters* **11**:3179–3184. DOI: https://doi.org/10.3892/ol.2016.4340, PMID: 27123085

Baranello L, Kouzine F, Wojtowicz D, Cui K, Przytycka TM, Zhao K, Levens D. 2014. DNA break mapping reveals topoisomerase II activity genome-wide. *International Journal of Molecular Sciences* **15**:13111–13122. DOI: https://doi.org/10.3390/ijms150713111, PMID: 25056547

Barlow JH, Faryabi RB, Callén E, Wong N, Malhowski A, Chen HT, Gutierrez-Cruz G, Sun HW, McKinnon P, Wright G, Casellas R, Robbiani DF, Staudt L, Fernandez-Capetillo O, Nussenzweig A. 2013. Identification of early replicating fragile sites that contribute to genome instability. *Cell* **152**:620–632. DOI: https://doi.org/10.1016/j.cell.2013.01.006, PMID: 23352430

Bence NF, Sampat RM, Kopito RR. 2001. Impairment of the ubiquitin-proteasome system by protein aggregation. *Science* **292**:1552–1555. DOI: https://doi.org/10.1126/science.292.5521.1552, PMID: 11375494

Bence NF, Bennett EJ, Kopito RR. 2005. Application and Analysis of the GFPu Family of Ubiquitin-Proteasome System Reporters. In: *Methods in Enzymology*. Academic Press. p. 481–490. DOI: https://doi.org/10.1016/S0076-6879(05)99033-2

Bredemeyer AL, Sharma GG, Huang CY, Helmink BA, Walker LM, Khor KC, Nuskey B, Sullivan KE, Pandita TK, Bassing CH, Sleckman BP. 2006. ATM stabilizes DNA double-strand-break complexes during V(D)J recombination. *Nature* **442**:466–470. DOI: https://doi.org/10.1038/nature04866, PMID: 16799570

Callén E, Jankovic M, Difilippantonio S, Daniel JA, Chen HT, Celeste A, Pellegrini M, McBride K, Wangsa D, Bredemeyer AL, Sleckman BP, Ried T, Nussenzweig M, Nussenzweig A. 2007. ATM prevents the persistence and propagation of chromosome breaks in lymphocytes. *Cell* **130**:63–75. DOI: https://doi.org/10.1016/j.cell.2007.06.016, PMID: 17599403

Callen E, Di Virgilio M, Kruhlak MJ, Nieto-Soler M, Wong N, Chen HT, Faryabi RB, Polato F, Santos M, Starnes LM, Wesemann DR, Lee JE, Tubbs A, Sleckman BP, Daniel JA, Ge K, Alt FW, Fernandez-Capetillo O, Nussenzweig MC, Nussenzweig A. 2013. 53bp1 mediates productive and mutagenic DNA repair through distinct phosphoprotein interactions. *Cell* **153**:1266–1280. DOI: https://doi.org/10.1016/j.cell.2013.05.023, PMID: 23727112

Canela A, Sridharan S, Sciascia N, Tubbs A, Meltzer P, Sleckman BP, Nussenzweig A. 2016. DNA Breaks and End Resection Measured Genome-wide by End Sequencing. *Molecular Cell* **63**:898–911. DOI: https://doi.org/10.1016/j.molcel.2016.06.034

Canela A, Maman Y, Jung S, Wong N, Callen E, Day A, Kieffer-Kwon K-R, Pekowska A, Zhang H, Rao SSP, Huang S-C, Mckinnon PJ, Aplan PD, Pommier Y, Aiden EL, Casellas R, Nussenzweig A. 2017. Genome organization drives chromosome fragility. *Cell* **170**:507–521. DOI: https://doi.org/10.1016/j.cell.2017.06.034

Canela A, Maman Y, Huang SN, Wutz G, Tang W, Zagnoli-Vieira G, Callen E, Wong N, Day A, Peters JM, Caldecott KW, Pommier Y, Nussenzweig A. 2019. Topoisomerase II-Induced chromosome breakage and translocation is determined by chromosome architecture and transcriptional activity. *Molecular Cell* **75**:252–266. DOI: https://doi.org/10.1016/j.molcel.2019.04.030, PMID: 31202577

Celeste A, Petersen S, Romanienko PJ, Fernandez-Capetillo O, Chen HT, Sedelnikova OA, Reina-San-Martin B, Coppola V, Meffre E, Difilippantonio MJ, Redon C, Pilch DR, Olaru A, Eckhaus M, Camerini-Otero RD, Tessarollo L, Livak F, Manova K, Bonner WM, Nussenzweig MC, et al. 2002. Genomic instability in mice lacking histone H2AX. *Science* **296**:922–927. DOI: https://doi.org/10.1126/science.1069398, PMID: 11934988

Ceruti S, Mazzola A, Abbracchio MP. 2006. Proteasome inhibitors potentiate etoposide-induced cell death in human astrocytoma cells bearing a mutated p53 isoform. *Journal of Pharmacology and Experimental Therapeutics* **319**:1424–1434. DOI: https://doi.org/10.1124/jpet.106.109397, PMID: 16971507

Chase JW, Richardson CC. 1974. Exonuclease VII of *Escherichia coli*. Mechanism of action. *The Journal of Biological Chemistry* **249**:4553–4561. PMID: 4602030

Cowell IG, Austin CA. 2012. Mechanism of generation of therapy related leukemia in response to anti-topoisomerase II agents. *International Journal of Environmental Research and Public Health* **9**:2075–2091. DOI: https://doi.org/10.3390/ijerph9062075, PMID: 22829791

Destanovic E, Boos J, Lanvers-Kaminsky C. 2018. Preclinical evaluation of combined topoisomerase and proteasome inhibition against pediatric malignancies. *Anticancer Research* **38**:3977–3984. DOI: https://doi.org/10.21873/anticanres.12684, PMID: 29970520

Dittus C, Grover N, Ellsworth S, Tan X, Park SI. 2018. Bortezomib in combination with dose-adjusted EPOCH (etoposide, prednisone, vincristine, cyclophosphamide, and doxorubicin) induces long-term survival in patients with plasmablastic lymphoma: a retrospective analysis. *Leukemia & Lymphoma* **59**:2121–2127. DOI: https://doi.org/10.1080/10428194.2017.1416365, PMID: 29303024

Errington F, Willmore E, Leontiou C, Tilby MJ, Austin CA. 2004. Differences in the longevity of topo iiα and topo iiβ drug-stabilized cleavable complexes and the relationship to drug sensitivity. *Cancer Chemotherapy and Pharmacology* **53**:155–162. DOI: https://doi.org/10.1007/s00280-003-0701-1, PMID: 14504921

Gittens WH, Johnson DJ, Allison RM, Cooper TJ, Thomas H, Neale MJ. 2019. A nucleotide resolution map of Top2-linked DNA breaks in the yeast and human genome. *Nature Communications* **10**:4846. DOI: https://doi.org/10.1038/s41467-019-12802-5

Gómez-Herreros F, Romero-Granados R, Zeng Z, Alvarez-Quilón A, Quintero C, Ju L, Umans L, Vermeire L, Huylebroeck D, Caldecott KW, Cortés-Ledesma F. 2013. TDP2-dependent non-homologous end-joining protects against topoisomerase II-induced DNA breaks and genome instability in cells and in vivo. *PLOS Genetics* **9**:e1003226. DOI: https://doi.org/10.1371/journal.pgen.1003226, PMID: 23505375

Gómez-Herreros F, Zagnoli-Vieira G, Ntai I, Martínez-Macías MI, Anderson RM, Herrero-Ruíz A, Caldecott KW. 2017. TDP2 suppresses chromosomal translocations induced by DNA topoisomerase II during gene transcription. *Nature Communications* **8**:1–9. DOI: https://doi.org/10.1038/s41467-017-00307-y

Halasi M, Pandit B, Gartel AL. 2014. Proteasome inhibitors suppress the protein expression of mutant p53. *Cell Cycle* **13**:3202–3206. DOI: https://doi.org/10.4161/15384101.2014.950132, PMID: 25485499

Hoa NN, Shimizu T, Zhou ZW, Wang ZQ, Deshpande RA, Paull TT, Akter S, Tsuda M, Furuta R, Tsutsui K, Takeda S, Sasanuma H. 2016. Mre11 is essential for the removal of lethal topoisomerase 2 covalent cleavage complexes. *Molecular Cell* **64**:580–592. DOI: https://doi.org/10.1016/j.molcel.2016.10.011, PMID: 27814490

Hsiang YH, Liu LF. 1989. Evidence for the reversibility of cellular DNA lesion induced by mammalian topoisomerase II poisons. *The Journal of Biological Chemistry* **264**:9713–9715. PMID: 2542330

Hu XV, Rodrigues TM, Tao H, Baker RK, Miraglia L, Orth AP, Lyons GE, Schultz PG, Wu X. 2010. Identification of RING finger protein 4 (RNF4) as a modulator of DNA demethylation through a functional genomics screen. *PNAS* **107**:15087–15092. DOI: https://doi.org/10.1073/pnas.1009025107, PMID: 20696907

Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, Kent WJ. 2004. The UCSC table browser data retrieval tool. *Nucleic Acids Research* **32**:493D–496. DOI: https://doi.org/10.1093/nar/gkh103, PMID: 14681465

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The human genome browser at UCSC. *Genome Research* **12**:996–1006. DOI: https://doi.org/10.1101/gr.229102, PMID: 12045153

Kisselev AF, Goldberg AL. 2001. Proteasome inhibitors: from research tools to drug candidates. *Chemistry & Biology* **8**:739–758. DOI: https://doi.org/10.1016/S1074-5521(01)00056-4

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**:R25. DOI: https://doi.org/10.1186/gb-2009-10-3-r25, PMID: 19261174

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with bowtie 2. *Nature Methods* **9**:357–359. DOI: https://doi.org/10.1038/nmeth.1923, PMID: 22388286

Ledesma FC, El Khamisy SF, Zuma MC, Osborn K, Caldecott KW. 2009. A human 5′-tyrosyl DNA phosphodiesterase that repairs topoisomerase-mediated DNA damage. *Nature* **461**:674–678. DOI: https://doi.org/10.1038/nature08444

Lee KC, Padget K, Curtis H, Cowell IG, Moiani D, Sondka Z, Morris NJ, Jackson GH, Cockell SJ, Tainer JA, Austin CA. 2012. MRE11 facilitates the removal of human topoisomerase II complexes from genomic DNA. *Biology Open* **1**:863–873. DOI: https://doi.org/10.1242/bio.20121834, PMID: 23213480

Lee KC, Bramley RL, Cowell IG, Jackson GH, Austin CA. 2016. Proteasomal inhibition potentiates drugs targeting DNA topoisomerase II. *Biochemical Pharmacology* **103**:29–39. DOI: https://doi.org/10.1016/j.bcp.2015.12.015

Lee K, Swan R, Sondka Z, Padget K, Cowell I, Austin C. 2018. Effect of TDP2 on the level of TOP2-DNA complexes and SUMOylated TOP2-DNA complexes. *International Journal of Molecular Sciences* **19**:2056. DOI: https://doi.org/10.3390/ijms19072056

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/Map format and SAMtools. *Bioinformatics* **25**:2078–2079. DOI: https://doi.org/10.1093/bioinformatics/btp352, PMID: 19505943

Lin CP, Ban Y, Lyu YL, Desai SD, Liu LF. 2008. A ubiquitin-proteasome pathway for the repair of topoisomerase I-DNA covalent complexes. *Journal of Biological Chemistry* **283**:21074–21083. DOI: https://doi.org/10.1074/jbc.M803493200, PMID: 18515798

Liyanage M, Coleman A, du Manoir S, Veldman T, McCormack S, Dickson RB, Barlow C, Wynshaw-Boris A, Janz S, Wienberg J, Ferguson-Smith MA, Schröck E, Ried T. 1996. Multicolour spectral karyotyping of mouse chromosomes. *Nature Genetics* **14**:312–315. DOI: https://doi.org/10.1038/ng1196-312, PMID: 8896561

Long BH, Musial ST, Brattain MG. 1985. Single- and double-strand DNA breakage and repair in human lung adenocarcinoma cells exposed to etoposide and teniposide. *Cancer Research* **45**:3106–3112. PMID: 3839166

Manasanch EE, Orlowski RZ. 2017. Proteasome inhibitors in cancer therapy. *Nature Reviews Clinical Oncology* **14**:417–433. DOI: https://doi.org/10.1038/nrclinonc.2016.206

Mao Y, Desai SD, Ting CY, Hwang J, Liu LF. 2001. 26 S proteasome-mediated degradation of topoisomerase II cleavable complexes. *Journal of Biological Chemistry* **276**:40652–40658. DOI: https://doi.org/10.1074/jbc.M104009200, PMID: 11546768

Mårtensson S, Nygren J, Osheroff N, Hammarsten O. 2003. Activation of the DNA-dependent protein kinase by drug-induced and radiation-induced DNA strand breaks. *Radiation Research* **160**:291–301. DOI: https://doi.org/10.1667/0033-7587(2003)160[0291:AOTDPK]2.0.CO;2, PMID: 12926987

Muslimović A, Nyström S, Gao Y, Hammarsten O. 2009. Numerical analysis of etoposide induced DNA breaks. *PLOS ONE* **4**:e5859. DOI: https://doi.org/10.1371/journal.pone.0005859, PMID: 19516899

Neale MJ, Pan J, Keeney S. 2005. Endonucleolytic processing of covalent protein-linked DNA double-strand breaks. *Nature* **436**:1053–1057. DOI: https://doi.org/10.1038/nature03872, PMID: 16107854

Nitiss JL. 2009a. DNA topoisomerase II and its growing repertoire of biological functions. *Nature Reviews Cancer* **9**:327–337. DOI: https://doi.org/10.1038/nrc2608, PMID: 19377505

Nitiss JL. 2009b. Targeting DNA topoisomerase II in Cancer chemotherapy. *Nature Reviews Cancer* **9**:338–350. DOI: https://doi.org/10.1038/nrc2607, PMID: 19377506

Paiano J, Wu W, Yamada S, Sciascia N, Callen E, Paola Cotrim A, Deshpande RA, Maman Y, Day A, Paull TT, Nussenzweig A. 2020. ATM and PRDM9 regulate SPO11-bound recombination intermediates during meiosis. *Nature Communications* **11**:857. DOI: https://doi.org/10.1038/s41467-020-14654-w, PMID: 32051414

Palumbo A, Gay F, Bringhen S, Falcone A, Pescosta N, Callea V, Caravita T, Morabito F, Magarotto V, Ruggeri M, Avonto I, Musto P, Cascavilla N, Bruno B, Boccadoro M. 2008. Bortezomib, doxorubicin and dexamethasone in advanced multiple myeloma. *Annals of Oncology* **19**:1160–1165. DOI: https://doi.org/10.1093/annonc/mdn018, PMID: 18326520

Pommier Y, Sun Y, Huang S-yinN, Nitiss JL. 2016. Roles of eukaryotic topoisomerases in transcription, replication and genomic stability. *Nature Reviews Molecular Cell Biology* **17**:703–721. DOI: https://doi.org/10.1038/nrm.2016.111

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**:841–842. DOI: https://doi.org/10.1093/bioinformatics/btq033, PMID: 20110278

R Development Core Team. 2008. R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. Vienna, Austria,http://www.R-project.org/

Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL. 2014. A 3D map of the human genome at Kilobase resolution reveals principles of chromatin looping. *Cell* **159**:1665–1680. DOI: https://doi.org/10.1016/j.cell.2014.11.021, PMID: 25497547

Rastogi N, Mishra DP. 2012. Therapeutic targeting of Cancer cell cycle using proteasome inhibitors. *Cell Division* **7**:26. DOI: https://doi.org/10.1186/1747-1028-7-26, PMID: 23268747

Schellenberg MJ, Lieberman JA, Herrero-Ruiz A, Butler LR, Williams JG, Muñoz-Cabello AM, Mueller GA, London RE, Cortés-Ledesma F, Williams RS. 2017. ZATT (ZNF451)-mediated resolution of topoisomerase 2 DNA-protein cross-links. *Science* **357**:1412–1416. DOI: https://doi.org/10.1126/science.aam6468, PMID: 28912134

Sun Y, Jenkins LMM, Su YP, Nitiss KC, Nitiss JL, Pommier Y. 2019. A conserved SUMO-Ubiquitin pathway directed by RNF4/SLX5-SLX8 and PIAS4/SIZ1 drives proteasomal degradation of topoisomerase DNA-protein crosslinks. *bioRxiv*. DOI: https://doi.org/10.1101/707661

Sunter NJ, Cowell IG, Willmore E, Watters GP, Austin CA. 2010. Role of topoisomerase iiβ in DNA damage response following IR and etoposide. *Journal of Nucleic Acids* **2010**:710589. DOI: https://doi.org/10.4061/2010/710589, PMID: 20847952

Thomas A, Bates S, Figg WD, Pommier Y. 2017. DNA topoisomerase targeting drugs. In: Bast R. C, Croce C. M, Hait W. N, Hong W. K, Kufe D, Piccard-Gebhart M, Pollock R. E, Weichselbaum R. R, Wang H, Holland J. F (Eds). *Holland-Frei Cancer Medicine*. Hoboken, NJ: John Willey & Sons, Inc. p. 665–678.

von Metzler I, Heider U, Mieth M, Lamottke B, Kaiser M, Jakob C, Sezer O. 2009. Synergistic interaction of proteasome and topoisomerase II inhibition in multiple myeloma. *Experimental Cell Research* **315**:2471–2478. DOI: https://doi.org/10.1016/j.yexcr.2009.04.019, PMID: 19410573

Willmore E, Frank AJ, Padget K, Tilby MJ, Austin CA. 1998. Etoposide targets topoisomerase IIalpha and IIbeta in leukemic cells: isoform-specific cleavable complexes visualized and *quantified* in situ by a novel immunofluorescence technique. *Molecular Pharmacology* **54**:78–85. DOI: https://doi.org/10.1124/mol.54.1.78, PMID: 9658192

Zagnoli-Vieiral G, Caldecott KW. 2017. Tdp2, top2 And SUMO: What Is ZATT About? *Cell Research* **27**:1405–1406. DOI: https://doi.org/10.1038/cr.2017.147

Zhang A, Lyu YL, Lin CP, Zhou N, Azarova AM, Wood LM, Liu LF. 2006. A protease pathway for the repair of topoisomerase II-DNA covalent complexes. *Journal of Biological Chemistry* **281**:35997–36003. DOI: https://doi.org/10.1074/jbc.M604149200, PMID: 16973621

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biology* **9**:R137. DOI: https://doi.org/10.1186/gb-2008-9-9-r137, PMID: 18798982

Zhang CZ, Leibowitz ML, Pellman D. 2013. Chromothripsis and beyond: rapid genome evolution from complex chromosomal rearrangements. *Genes & Development* **27**:2513–2530. DOI: https://doi.org/10.1101/gad.229559.113, PMID: 24298051

Zong D, Adam S, Wang Y, Sasanuma H, Callén E, Murga M, Day A, Kruhlak MJ, Wong N, Munro M, Ray Chaudhuri A, Karim B, Xia B, Takeda S, Johnson N, Durocher D, Nussenzweig A. 2019. BRCA1 haploinsufficiency is masked by RNF168-Mediated chromatin ubiquitylation. *Molecular Cell* **73**:1267–1281. DOI: https://doi.org/10.1016/j.molcel.2018.12.010, PMID: 30704900

nature COMMUNICATIONS

ARTICLE

https://doi.org/10.1038/s41467-020-14654-w OPEN

ATM and PRDM9 regulate SPO11-bound recombination intermediates during meiosis

Jacob Paiano[1,2,8], Wei Wu[1,8], Shintaro Yamada[3,4], Nicholas Sciascia[1,5], Elsa Callen[1], Ana Paola Cotrim[1], Rajashree A. Deshpande[6,7], Yaakov Maman[1], Amanda Day[1], Tanya T. Paull[6,7] & André Nussenzweig[1]✉

Meiotic recombination is initiated by SPO11-induced double-strand breaks (DSBs). In most mammals, the methyltransferase PRDM9 guides SPO11 targeting, and the ATM kinase controls meiotic DSB numbers. Following MRE11 nuclease removal of SPO11, the DSB is resected and loaded with DMC1 filaments for homolog invasion. Here, we demonstrate the direct detection of meiotic DSBs and resection using END-seq on mouse spermatocytes with low sample input. We find that DMC1 limits both minimum and maximum resection lengths, whereas 53BP1, BRCA1 and EXO1 play surprisingly minimal roles. Through enzymatic modifications to END-seq, we identify a SPO11-bound meiotic recombination intermediate (SPO11-RI) present at all hotspots. We propose that SPO11-RI forms because chromatin-bound PRDM9 asymmetrically blocks MRE11 from releasing SPO11. In *Atm*$^{-/-}$ spermatocytes, trapped SPO11 cleavage complexes accumulate due to defective MRE11 initiation of resection. Thus, in addition to governing SPO11 breakage, ATM and PRDM9 are critical local regulators of mammalian SPO11 processing.

[1] Laboratory of Genome Integrity, National Cancer Institute, NIH, Bethesda, MD, USA. [2] Immunology Graduate Group, University of Pennsylvania, Philadelphia, PA, USA. [3] Molecular Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [4] Department of Radiation Genetics, Graduate School of Medicine, Kyoto University, Kyoto 606-8501, Japan. [5] Institute for Biomedical Sciences, George Washington University, Washington, DC, USA. [6] The Howard Hughes Medical Institute and The University of Texas at Austin, Austin, TX 78712, USA. [7] The Department of Molecular Biosciences, The University of Texas at Austin, Austin, TX 78712, USA. [8] These authors contributed equally: Jacob Paiano, Wei Wu. ✉email: andre_nussenzweig@nih.gov

Recombination between homologous chromosomes during meiosis requires DNA double-strand break (DSB) formation by the topoisomerase-like protein SPO11[1]. After cutting, SPO11 remains covalently bound to a two-nucleotide, 5′ overhang at both ends of the DNA via phosphotyrosyl linkage. Recombination then begins with the processing of SPO11-bound DSBs into resected 3′ single-stranded DNA (ssDNA) tails that preferentially invade the homologous chromosome by the recombinases DMC1 and RAD51. Studies in budding yeast *Saccharomyces cerevisiae* determined that the MRE11/RAD50/NBS1 (MRN) complex detects SPO11 and cooperates with Sae2 to produce a nick on the SPO11-bound strand via MRE11 endonuclease activity[2]. The nick serves as an entry point for both short-range MRE11 3′−5′ exonuclease activity to degrade back to the DSB, thereby removing covalently bound SPO11 attached to a ssDNA oligonucleotide, as well as for more-extensive long-range processing of 5′ strands (Fig. 1a)[2]. In budding yeast, Exo1 nuclease is uniquely responsible for this long-range 5′−3′ resection[3]. Moreover, short- and long-range resection are tightly coupled in a single processive reaction (Fig. 1a). As a result, meiotic DSBs are maximally resected as soon as they appear and unresected SPO11-bound DSBs are extremely rare[4–6]. Although ATM has been shown to regulate DSB numbers and locations[7,8], it remains unclear whether it also functions downstream in regulating SPO11 processing and resection.

Distinct from yeast, DSB hotspots in mice and humans are determined by the DNA-binding specificity of the PRDM9 methyltransferase[9]. Besides positioning DSBs, PRDM9-binding activity also reorganizes nucleosomes in a manner that creates a nucleosome-depleted region (NDR) within which DSBs and PRDM9 itself are centered[10]. Moreover, PRDM9 has been suggested to have a role in DSB repair post cleavage[11,12]. Crossover resolution is facilitated by PRDM9 binding symmetrically to the template (uncut) homolog, which generates a NDR within which the DSB-initiating chromosome can stably engage[13–15]. If PRDM9 remains bound to DNA post-cleavage, it is possible that it could influence the downstream processing events that facilitate synapsis and crossover recombination.

Here, we demonstrate that ATM regulates multiple steps of initial SPO11 processing, including the activation of MRE11, the coordination of short- and long-range resection, and the assurance of minimal SPO11 cutting per hotspot. However, we find that ATM-mediated SPO11 processing can be hindered by the binding of PRDM9 to hotspots post-SPO11 cutting. We suggest that PRDM9 acts as a physical barrier to MRE11 activity, preventing SPO11 release during the early stages of homolog engagement. This generates a PRDM9-dependent SPO11-bound recombination intermediate (SPO11-RI). We propose that SPO11-RI may favor high-fidelity homologous recombination by facilitating crossover (CO) and non-crossover (NCO) events.

Results

END-seq robustly detects mouse meiotic DSB hotspots. To probe the early steps of mouse meitotic recombination, we utilized END-seq[16–18]. In this method, a sequencing adapter is ligated to each end of a DNA break inside an agarose plug after a combination of nucleases ExoVII and ExoT removes ssDNA overhangs. As a result, sequencing reads begin at the terminal end of physiological resection, resulting in libraries of ssDNA-DSB

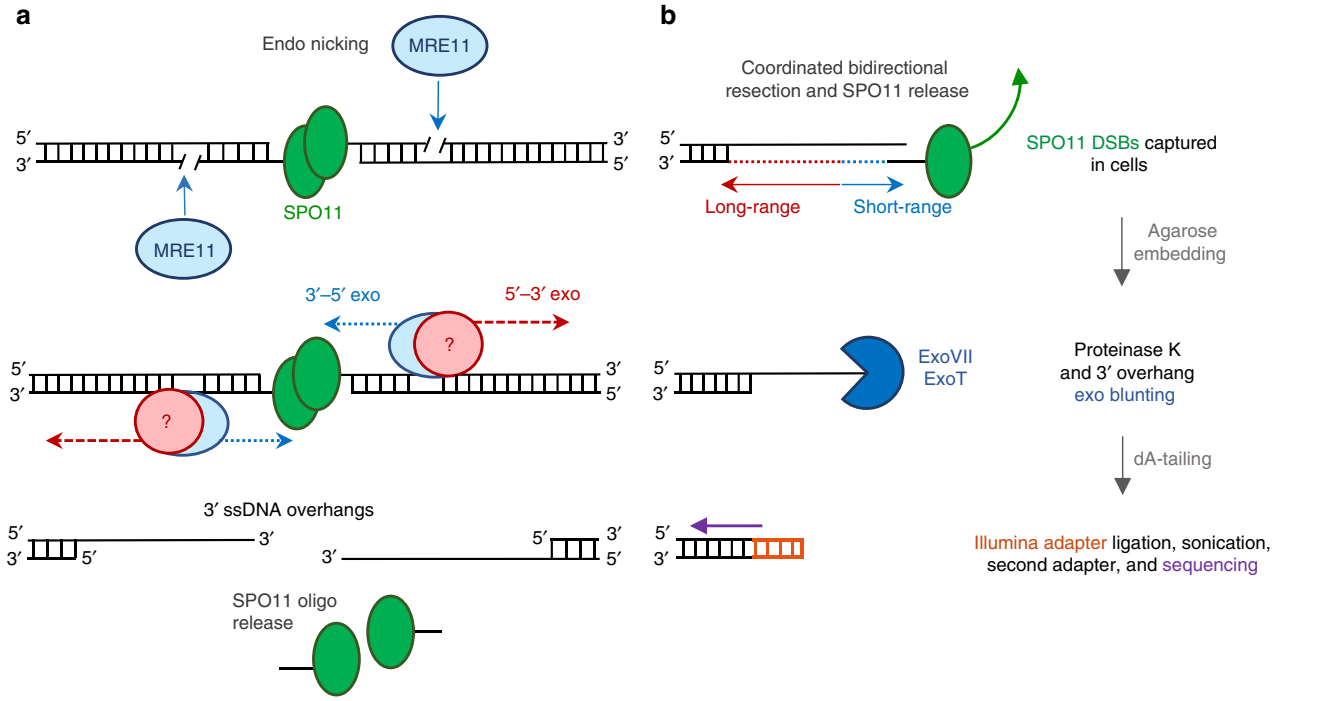


Fig. 1 SPO11 generates meiotic DSBs that are detectable by END-seq. a Illustration of meiotic break generation and processing. SPO11 induces a double-strand break (DSB) and remains covalently bound to both DNA ends. MRE11 recognizes the DSB and induces a nick on the SPO11-bound strand. Tightly coordinated short-range 3′−5′ resection by MRE11 and long-range 5′−3′ resection by an unknown nuclease generates 3′ overhangs for homology search. MRE11 activity releases SPO11 bound to short oligonucleotides (SPO11 oligos). **b** Brief schematic of END-seq detection of SPO11 DSBs (only one side of the DSB is shown for simplicity). In vivo processing of SPO11 by coordinated bidirectional resection removes covalently bound SPO11 and produces a 3′ overhang present at the time of END-seq preparation and agarose cell embedding. Initial END-seq processing degrades all proteins by proteinase K and blunts ssDNA overhangs by in vitro nuclease digestion (dark blue). Once fully blunted and dA-tailed, the DNA end is ligated to a biotinylated Illumina sequencing adapter (orange), sheared, and streptavidin captured. A second Illumina adapter is ligated at the other end of the sonicated fragment after end repair and sequenced.

junctions (Fig. 1b). These exonucleases can be used to detect DSB termini that are either protein-bound or protein-free. For example, etoposide-induced DSBs, which are covalently attached to topoisomerase 2 (TOP2) via an active site tyrosine at the 5′-termini, require ExoVII to remove covalently bound TOP2[16], whereas ExoT can only blunt protein-free overhangs, resulting in a ligatable DNA end[18]. Like TOP2, the topoisomerase-like protein SPO11 remains attached to DSB 5′-ends prior to release by MRE11-mediated nicking as short 20–40 bp oligonucleotides (SPO11 oligos)[7,19] (Fig. 1a).

We assayed spermatocytes from juvenile mouse testes during the first wave of semi-synchronous meiosis I. We embedded spermatocytes from 20 pooled, 12–14 dpp C57BL/6 J (B6) mice in agarose plugs and blunted meiotic ssDNA overhangs with ExoVII and ExoT before ligating sequencing adapters (Fig. 1b). Approximately 5000 reproducible broken hotspots were called with a threshold of at least 2.5-fold enrichment (Supplementary Fig. 1a, Supplementary Data 1). We overlapped these END-seq peaks with previously reported B6 meiotic hotspots determined by SPO11-oligo sequencing and DMC1 single-strand DNA sequencing (SSDS)[8,11,20]. By visual inspection and correlative analysis, both END-seq break location and peak intensity overlapped with SPO11 oligos and SSDS, which accounted for 97% and 98% of END-seq peaks, respectively (Fig. 2a, Supplementary Fig. 1b, c). END-seq therefore provides a map of directly detected meiotic DSB resection in mammals.

Using the 2.5-fold-enrichment criteria, END-seq peaks called a third of the total hotspots determined by SPO11-oligo seq or SSDS (Supplementary Fig. 1b, c). However, END-seq peaks that did not meet this cutoff were nevertheless associated with previously mapped hotspots. For example, 66% of SPO11-oligo loci that were “END-seq negative” showed the same DSB detection pattern when these END-seq reads were aggregated (Supplementary Fig. 1d, e). Moreover, these END-seq-negative SPO11-oligo sites show a significant reduction in SPO11-oligo reads (Supplementary Fig. 1f), indicating that these are the coldest meiotic hotspots. We conclude that END-seq detects breakage at all ~ 15,000 previously mapped hotspots, yet the more frequently broken top 33% of hotspots yield the most robust signal. Therefore, unless otherwise stated, the subsequent analyses were performed on the 5000 strongest END-seq breaks.

The landscape of mouse meiotic DSB resection. END-seq captured a strikingly uniform pattern of breakage at all sites that consisted of a strong central peak directly at the site of SPO11 cutting with an accumulation of reads flanking the cut site at a defined distance away (Fig. 2b, c). We interpret the central peak to be the direct detection of SPO11 breakage while adjacent, distal reads reflect minimum and maximum resection endpoints in the population of spermatocytes. Reads comprising the central peak are entirely coincident with SPO11-oligo mapping and reflect a subset of breaks within the spermatocyte population that has not yet released covalently bound SPO11 from the DSB (Fig. 2d). As discussed later, the central peak in WT cells is dependent on engagement of the cut chromosome with the uncut homolog, and therefore represents unreleased SPO11 associated with recombination.

END-seq detects the terminal end of physiologic resection after in vitro blunting of the 3′ overhang, with the first nucleotide sequenced corresponding to the position of the ssDNA-DSB junction (Figs. 2c, 1b)[18]. If bidirectional resection of DNA-bound SPO11 by short- and long-range resection machineries are entirely coordinated as in yeast[4–6], then the ssDNA-DSB junction will always be beyond the most distal MRE11 nick and correspond to the terminal end of long-range resection (Figs. 2c, 1b). In this case, the location of short-range 3′−5′ MRE11 exonuclease activity would not yield any sequencing reads as it does not operate independently of 5′−3′ resection (Figs. 2c, 1b). Indeed, at every hotspot, we observed a read-less “gap” in resection, consistent with the tight coupling of resection initiation (by MRE11) and 5′−3′ extension by the long-range resection machinery (Fig. 2b, c). The length of this gap reflects the minimum resection endpoints in the spermatocyte population and corresponds to the maximum distance from the DSB at which MRE11 nicks the strand plus any constant, minimum distance that 5′−3′ exonucleases traverse. This pattern is reminiscent of S1 nuclease detection of meiotic recombination in yeast[6] (Fig. 2e). Thus, both SPO11 cutting and its initial processing by coordinated resection mechanisms are highly conserved evolutionary features of meiotic recombination that span unicellular eukaryotes to mammals.

The gap size was extremely uniform at all hotspots, with mean maximum distance of 647 nts (Supplementary Fig. 2a), and was largely restricted to the two-nucleosome H3K4/K36me3 signal surrounding SPO11 cut sites as determined by ChIP-seq (Fig. 2f, Supplementary Fig. 2b). Thus, minimum resection distances correlate well with PRDM9-mediated methylated histone deposition[10]. Interestingly, the majority of CO and NCO boundaries in mice are also restricted to this region, rarely extending beyond the 650 bp gap (Supplementary Fig. 2c)[15,21]. Because minimum resection correlated well with methylated histones, we asked if long-range resection endpoints were similarly well positioned. Between two replicate END-seq samples, we found high correlation in the pattern of resection endpoints among the hotter hotspots detected with recurrent subpeaks apparent at a mean distance of 210 nucleotides (Supplementary Fig. 2d, e). This suggests that nucleosome occupancy far from the break site may influence long-range resection endpoints.

Meiotic DSB hotspots are detectable in a single mouse. To determine the sensitivity of the method, we compared 20 pooled juvenile mice to a library made from a single 12 dpp mouse. Both samples called ~ 5000 peaks, with 77–89% shared breaks with highly correlated ($r = 0.98$) END-seq intensities (Fig. 3a). Importantly, the break pattern at individual hotspots was fully retained in testes from a single mouse (Fig. 3b). To compare signal-to-noise ratio (S/N) in the two libraries, we followed ENCODE’s assessment of fraction of reads in peaks (FRiP) and cross-correlation profiles (CCPs) for ChIP-seq data sets[22,23]. FRiP and CCP values for both libraries exceeded ENCODE’s criteria for signal-to-noise ratios (Fig. 3c, d). We conclude that END-seq can accurately assess individual meiotic DSB locations and processing with remarkably little biological material. This high sensitivity bypasses the limitations of other hotspot mapping methods that require either impractical quantities of mice (SPO11-oligo seq) or the availability of species-specific, high-quality antibodies (DMC1 SSDS).

Increased breakage in *Atm*$^{-/-}$ spermatocytes. ATM is thought to negatively regulate SPO11 cutting, and in its absence, hotspot breakage and SPO11 oligos have been shown to markedly increase in mice[7]. This ultimately results in early meiotic arrest and apoptosis of *Atm*$^{-/-}$ spermatocytes that carry an excess of unrepaired DSBs[7]. To further validate this model using direct, genome-wide detection, we performed END-seq on *Atm*$^{-/-}$ spermatocytes that were backcrossed 11 times to the B6 background. Indeed, *Atm*$^{-/-}$ END-seq detected 99% of WT END-seq breaks while calling ~ 6300 more peaks (Supplementary Fig. 3a, Supplementary Data 2). All *Atm*$^{-/-}$ breaks overlapped better with SPO11 oligos and SSDS sites and showed amplified signal over WT at weak SSDS hotspots (Supplementary Fig. 3b–d). These

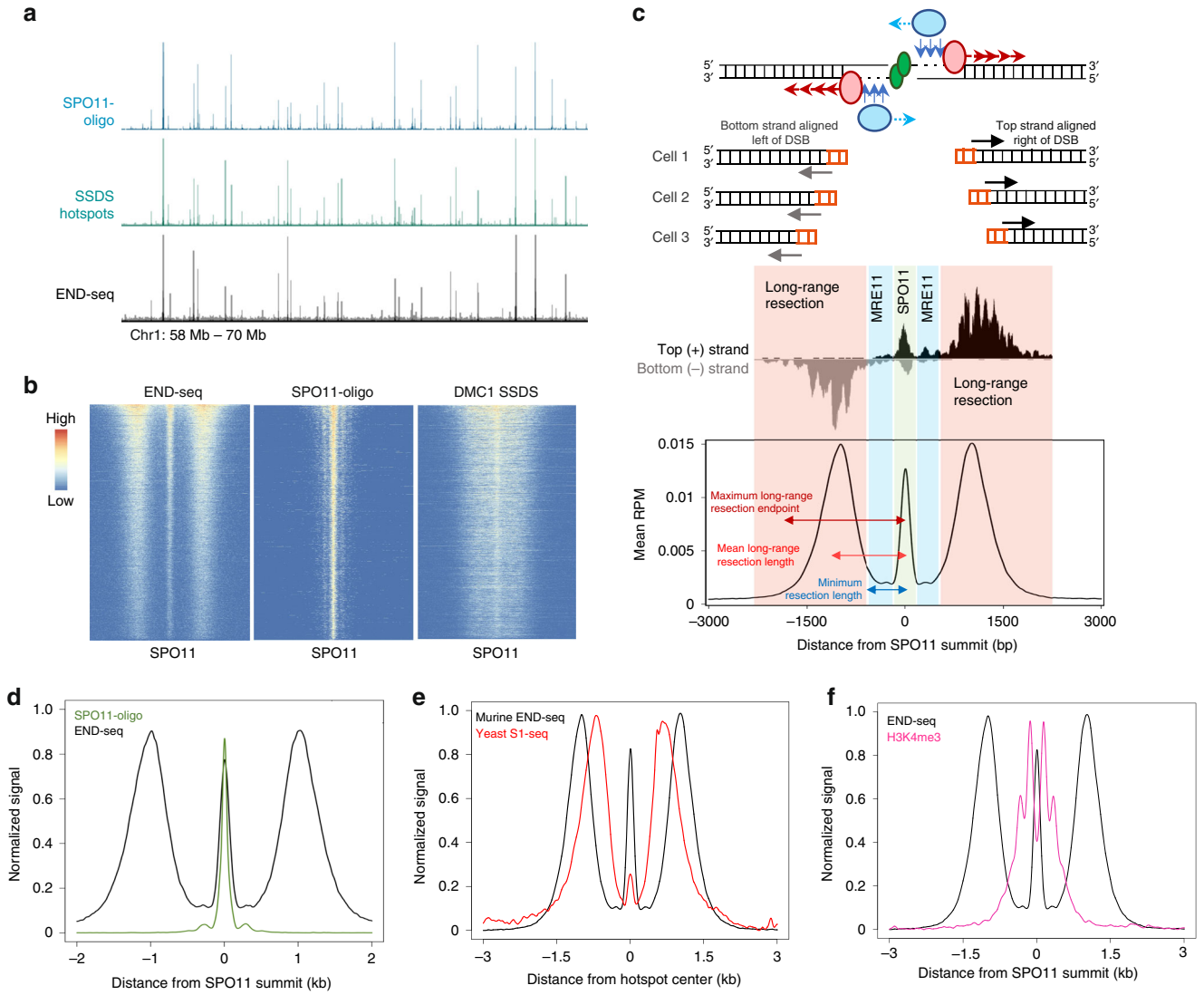


Fig. 2 END-seq correlates with previous hotspot mapping techniques and uncovers a uniform pattern of SPO11 processing at all hotspots. a Representative genome browser profiles of meiotic hotspots for SPO11-oligo sequencing, DMC1 SSDS, and END-seq. Browser axis scales are adjusted between techniques to show both hot and cold hotspots simultaneously. **b** Heatmaps of END-seq, SPO11-oligo, and SSDS ± 2.5 kb around hotspot centers (determined by SPO11-oligo summits), ordered by total read count of END-seq, for top 5000 END-seq breaks. WT END-seq peaks are provided in Supplementary Data 1. **c** Schematic of END-seq break pattern, consisting of (1) a central peak at the SPO11 break site (2) a read-less gap produced by MRE11-mediated short-range resection and minimum distance of long-range resection, and (3) distal reads at the terminal ends of long-range resection. This pattern is evident at individual hotspots (middle, chr1:68488000–68491500) and when signal from all hotspots is aggregated (bottom). Minimum resection lengths are calculated by the absence of sequencing reads (blue highlighted region); mean and maximum long-range resection are calculated by the average and most distal reads from the DSB, respectively (red highlighted region). **d** END-seq central peak and SPO11-oligo reads are coincident as evidenced by aggregated signal around SPO11-oligo summits (normalized to the same height). Both the primary SPO11 peak and adjacent secondary peaks are apparent. **e** Aggregate signal comparison of murine meiotic END-seq around SPO11-oligo summits and yeast meiotic S1-seq around yeast hotspot centers. **f** Aggregated END-seq and H3K4me3 ChIP-seq signal (normalized to the same height).

data confirm previous reports that colder hotspots become preferentially hotter in the absence of ATM owing to a loss of negative feedback of SPO11 cutting (Supplementary Fig. 3e)[7,8]. To quantify this increase, we added a spike-in normalization control to END-seq plugs, consisting of a G1-arrested Ableson-transformed murine pre-B cell line (*Lig4*$^{-/-}$) carrying a single zinc-finger-induced DSB at the TCRβ enhancer[16], which was mixed in at a 2% frequency with bulk testiscular cells. After normalizing END-seq reads at all ~15,000 SPO11-oligo hotspots using the signal at the TCRβ enhancer, we found that *Atm*$^{-/-}$ cells harbor a 4.5-fold increase in total breaks over WT. As discussed later, increased SPO11 double-cutting within the same hotspot may account for the 10-fold increase in SPO11 oligos released in *Atm*$^{-/-}$ spermatocytes;[7] yet END-seq would detect double-cutting as a single event, yielding a 4.5-fold increase in DSBs.

Unexpectedly we found that 7–16% (700–1800) of break sites were not shared in WT SSDS or SPO11-oligo maps, respectively (Supplementary Fig. 3b, c). Among these were several hundred promoter breaks, typically associated as "default" hotspots in organisms lacking PRDM9 (Supplementary Fig. 3f). Indeed, these break locations are among the hottest *Prdm9*$^{-/-}$ sites as determined by SSDS (Supplementary Fig. 3g)[11,24]. Thus, in the absence of ATM, colder hotspots and default hotspots become increasingly broken. This is consistent with previous observations that ATM-null spermatocytes have increased SPO11-oligo mapping at PRDM9-independent hotspots[14].

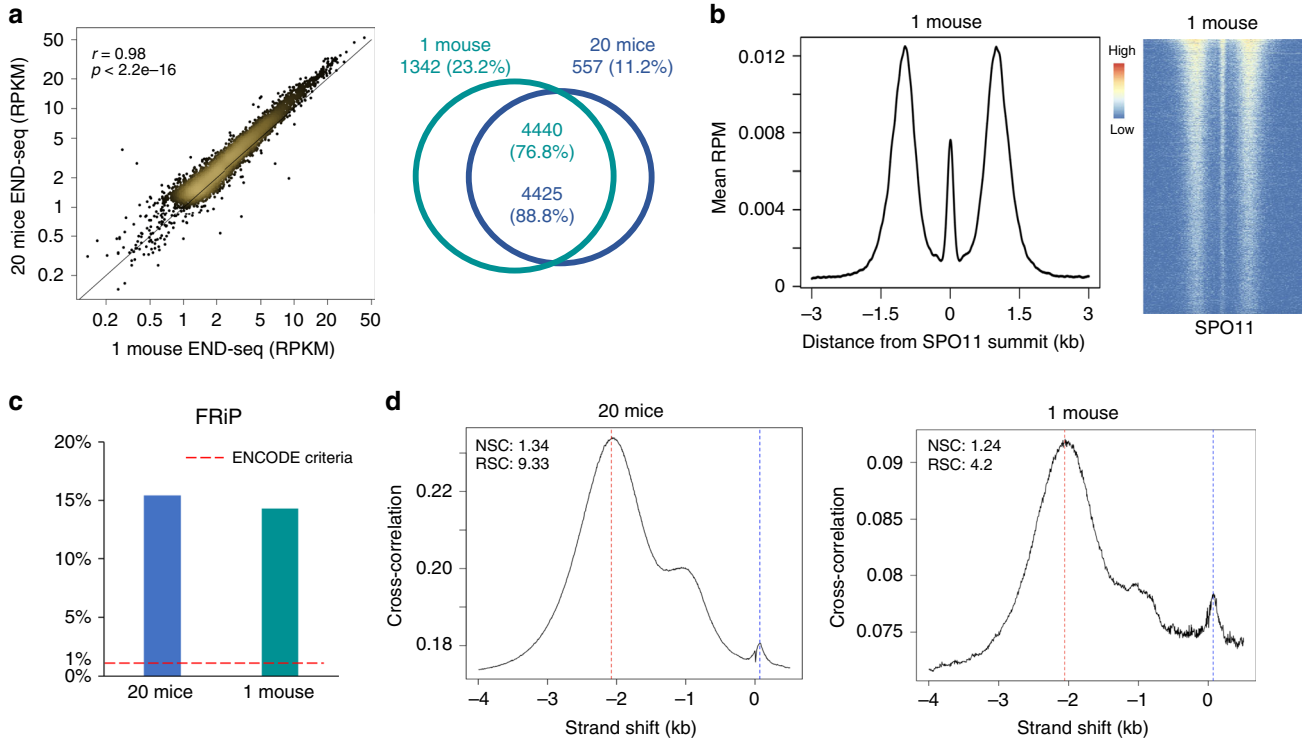


Fig. 3 END-seq detects meiotic DSBs in a single mouse. a Left panel: correlation (Pearson's *r*) between END-seq RPKM from 20 mice versus RPKM from one 12 dpp mouse ± 3 kb around SPO11 summits. Right panel: Venn diagram overlap of called peaks between 20 mice and one mouse. *P* value < 2.2e-16, Fisher's exact test. **b** Left panel: Full END-seq break pattern in aggregated signal from one mouse, centered around SPO11 oligos. Right panel: heatmap of END-seq signal from one mouse in a ± 3 kb window around SPO11 oligos, ordered by total read count. **c** FRiP values for 20 mice versus one mouse. Recommended ENCODE value denoted by dotted red line. **d** Cross-correlation plot profiles for 20 mice and one mouse. The plot shows Pearson cross-correlations (CCs, *y* axis) of read intensities between the plus strand and the minus strand, after shifting minus strand (*x* axis). One peak corresponds to read length (CCread, blue dash line) and the other one corresponds to the fragment length (CCfrag, red dash line). Normalized strand coefficient (NSC) is CCfrag divided by minimal CC value (CCmin) and relative strand coefficient (RSC) is the ratio of CCfrag-CCmin divided by CCread-CCmin. Higher NSC and RSC values mean more enrichment. ENCODE's recommendation for ChIP-seq: NSC ≥ 1.05 and RSC ≥ 0.8.

Minimal roles for EXO1, BRCA1, and 53BP1 during resection. Long-range resection endpoints showed greater variation (1–3 kb) relative to the minimum resection gap region (0.4–1 kb) (Supplementary Fig. 2a). Previous studies estimated total mammalian meiotic resection lengths based on the extent of DMC1 bound to ssDNA overhangs as measured by DMC1 SSDS[8]. Strikingly, we found that maximum resection endpoints extended significantly farther at all hotspots than DMC1-bound ssDNA (Fig. 4a, b). ssDNA occupied by DMC1 ranged from 800 to 2700 nts, whereas maximum long-range resection lengths (defined in Fig. 2c) determined by END-seq were 1.2–1.6-fold greater (Fig. 4b, c). These data indicate that DMC1 binds to only a portion of the available ssDNA and underestimates the total extent of meiotic resection. One potential explanation is that single tracks of ssDNA can be co-occupied by DMC1 and RAD51 filaments with DMC1 loaded more proximally to the break site than RAD51[25]. Although unlikely, this difference may also be due to technical limitations in SSDS library preparation that relies on microhomology-mediated hairpin formations naturally present in ssDNA tracks[26].

Because the resection pattern in mouse mirrored so closely that observed in yeast (Fig. 2e)[6], we wanted to know whether they utilized the same long-range end-processing machineries. Exo1-deficient yeast exhibit total loss of 5′−3′ long-range resection[3]. In contrast, EXO1 and DNA2 act redundantly in yeast vegetative and mammalian somatic cells to mediate end resection[27–29]. Surprisingly, END-seq analysis of juvenile *Exo1*$^{-/-}$ spermatocytes revealed that long-range resection was largely intact (Fig. 4d). When averaged genome-wide, we found a small, though significant, reduction in long-range resection distance in *Exo1*$^{-/-}$ cells (median resection tract: WT: 1,129 nts vs. *Exo1*$^{-/-}$: 996 nts (Fig. 4e). We therefore conclude that compared with yeast, mammalian meiosis has evolved additional mechanisms to achieve extensive 3′ overhangs, perhaps through utilization of redundant DNA2 exonuclease activity[27–29].

In somatic interphase cells, 53BP1 has been shown to inhibit long-range resection of DSBs[30]. One principal role of BRCA1 is to counteract 53BP1's block to resection in S phase, possibly by excluding it from chromatin proximal to DNA damage sites[31]. In addition, BRCA1 acts post-resection to load the RAD51 recombinase onto 3′ ssDNA[32]. Despite these extensive studies demonstrating their importance in somatic cells, the roles of 53BP1 and BRCA1 in regulating meiotic resection are unknown. We therefore performed END-seq on *53bp1*$^{-/-}$ and BRCA1-deficient spermatocytes and measured overall resection lengths. In striking contrast to 53BP1-deficient somatic cells in which DSB resection is consistently and acutely increased[18,30,33], we found that resection lengths were comparable at all meiotic hotspots in *53bp1*$^{-/-}$ and WT spermatocytes (Fig. 4e).

We then assayed *Brca1*$^{\Delta 11}$*p53*$^{+/-}$ mice, which exhibit known defects in BRCA1 function, yet are alive owing to partial p53 apoptotic suppression[34]. Contrary to our expectations[30,33], *Brca1*$^{\Delta 11}$*p53*$^{+/-}$ spermatocytes showed a mild increase in DSB resection relative to WT controls (Fig. 4e). Moreover, *Brca1*$^{\Delta 11}$*53bp1*$^{-/-}$ and *Brca1*$^{\Delta 11}$*53bp1*^{S25A} spermatocytes[33] exhibited a similar increase in resection as *Brca1*$^{\Delta 11}$*p53*$^{+/-}$

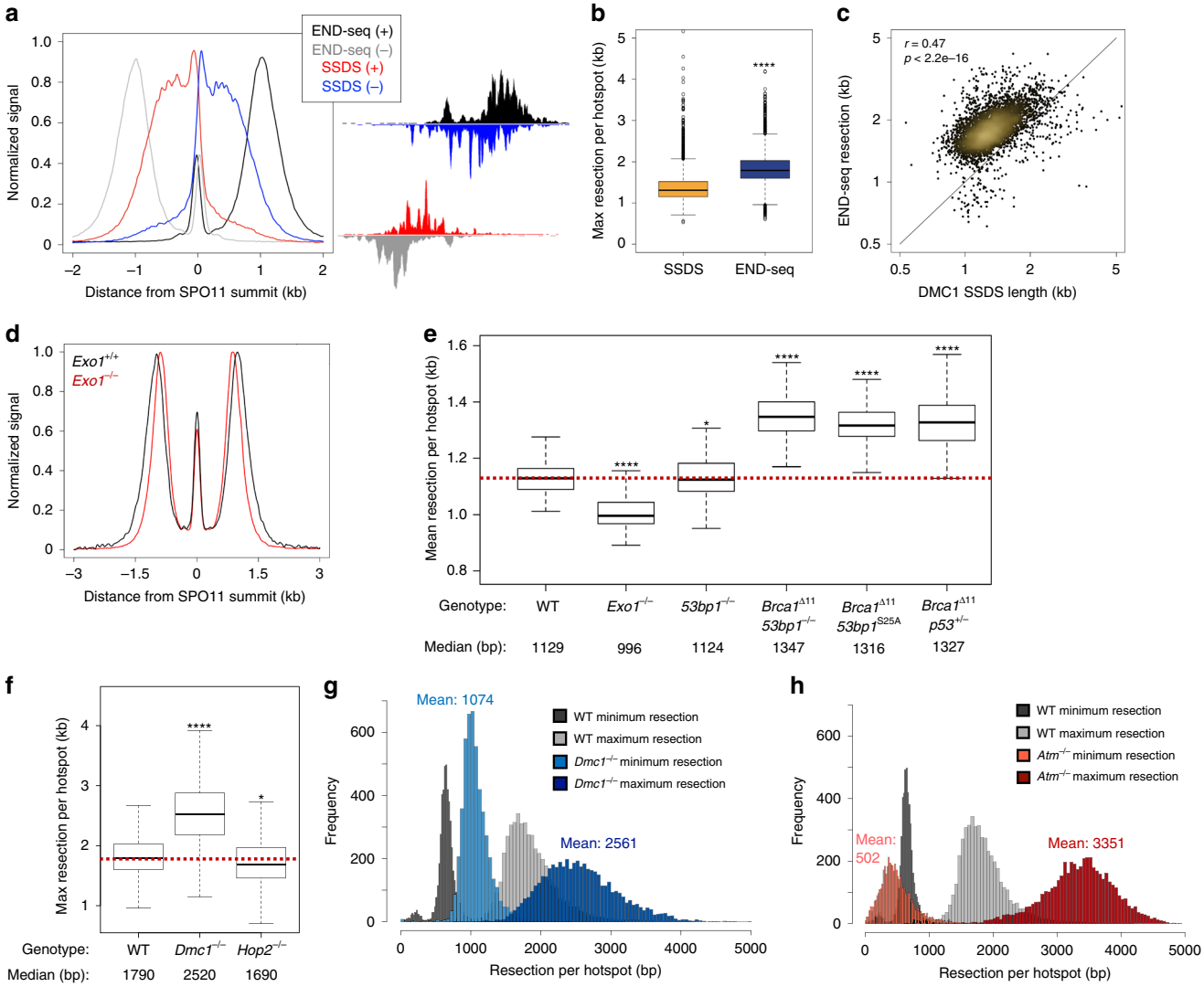


Fig. 4 END-seq accurately measures hotspot resection lengths and elucidates resection regulation. a Top (+) and bottom (-) strand distributions of END-seq and SSDS show increased resection detection by END-seq (left, signal normalized to the same height) that is evident at individual hotspots (right). **b** Boxplot of END-seq vs SSDS maximum resection per hotspot. ****$p < 1e\text{-}10$; *t* test. **c** Correlation (Pearson's *r*) of maximum resection endpoints detected by END-seq and SSDS. **d** Aggregate plot of END-seq signal in WT and *Exo1*$^{-/-}$ spermatocytes (signal normalized to the same height) around SPO11 oligos. **e** Boxplots of mean resection (as defined in Fig. 2c) between WT, *Exo1*$^{-/-}$, *53bp1*$^{-/-}$, *Brca1*$^{\Delta 11}$*53bp1*$^{-/-}$, *Brca1*$^{\Delta 11}$*53bp1*^{S25A} and *Brca1*$^{\Delta 11}$*p53*$^{+/-}$ at top 250 hotspots. *$p < 0.01$; ****$p < 1e\text{-}10$; *t* test with mu = 10 (mu is estimated as standard deviation of WT replicates). **f** Boxplots of maximum resection endpoints between WT, *Dmc1*$^{-/-}$, and *Hop2*$^{-/-}$ at top 5000 hotspots. *$p < 0.01$; ****$p < 1e\text{-}10$; *t* test with mu = 43 (mu is estimated as standard deviation of WT replicates). **g**, **h** Histogram distributions comparing either WT and *Dmc1*$^{-/-}$ **g** or WT and *Atm*$^{-/-}$ **h** END-seq minimum and maximum resection endpoints at top 5000 hotspots. Mean values (bp) are listed.

spermatocytes (Fig. 4e). Based on these findings, we conclude that in contrast to their well-defined antagonistic relationship during DSB processing in interphase cells, 53BP1 does not inhibit resection while BRCA1 does not promote resection during meiotic recombination. As discussed below, one reason why end resection increases (rather than decreases) in BRCA1-deficient cells may be owing to inefficient loading of the recombinase. Consistent with this, *Brca1*$^{\Delta 11}$*p53*$^{+/-}$ spermatocytes have reduced RAD51 and DMC1 focus counts[35].

RAD51, DMC1, and ATM negatively regulate long-range resection. In budding yeast, cells lacking DMC1 show hyper-resection, perhaps owing to a negative feedback on resection mediated by recombinase loading[4,6]. Given our results that BRCA1-deficient spermatocytes exhibited increased resection (Fig. 4e) associated with defective RAD51/DMC1 foci[35], we hypothesized that RAD51 or DMC1 loading onto ssDNA might limit long-range resection. Consistent with this, *Dmc1*$^{-/-}$ spermatocytes showed substantially increased resection relative to WT at all hotspots (Fig. 4f, g and Supplementary Fig. 4a). Minimum resection lengths also increased in *Dmc1*$^{-/-}$ cells by ~ 400 nts relative to WT (Fig. 4g and Supplementary Fig. 4a).

HOP2 protein is required for proper homolog pairing after loading of DMC1/RAD51 recombinases[36]. In *Hop2*$^{-/-}$ spermatocytes, resection lengths were similar to WT (Fig. 4f). Thus, recombinase loading post-resection, independent of strand invasion, limits both the minimum and maximum length that the long-range resection machinery processes DSBs. DMC1/RAD51 loading onto ssDNA might limit resection by preventing the re-initiation of exonucleases on already resected ssDNA.

ATM negatively regulates DSB induction by SPO11[7]. Despite the fact that *Atm*$^{-/-}$spermatocytes harbor a 10-fold increase in

SPO11 oligos[7], RAD51/DMC1 foci form at similar levels as in WT, whereas RPA foci counts are increased in mutant cells[37]. One possible reason why the number of foci does not correlate with the large increase in DSBs could be that recombinase level or activity is limiting filament formation. In this scenario, when excess DSBs are generated in an ATM-null background, the pool of RAD51 and/or DMC1 is insufficient to load onto all ssDNA regions. In accord with this hypothesis, the mean maximum resection lengths increased almost twofold in $Atm^{-/-}$ spermatocytes relative to WT (WT: 1845 nts. vs. ATM: 3351 nts; Fig. 4h). Moreover, in ATM knockouts, resection lengths were significantly greater and more widely distributed than even DMC1 knockouts (Fig. 4g, h), perhaps owing to combined deficiency in RAD51 and DMC1 loading onto ssDNA[37]. In accord with the idea that increased DSB numbers rather than deficiency in ATM signaling contributes to the hyper-resection phenotype in the ATM-null background, S1-seq analysis of $Atm^{-/-}Spo11^{+/-}$ spermatocytes, in which SPO11-oligo complexes are reduced by half[7], revealed normal resection lengths[38]. Together these data support the idea that the availability of recombinases to form filaments is limiting and insufficient when there is excessive SPO11 cutting.

ATM coordinates short- and long-range resection. In addition to the hyper-resection observed at a subset of breaks, distinct boundaries between the short- and long-range resection practically disappeared in $Atm^{-/-}$ spermatocytes. This resulted in reads mapping within the gap region (Fig. 5a, b), in contrast to WT that always generates ssDNA-DSB junctions beyond the most distal MRE11 nick. The decreased minimum resection could indicate that ATM regulates the tightly coupled activities of short-range and long-range resection (Fig. 5c).

If some DSBs remain unresected in ATM-deficient spermatocytes, we would predict that MRE11 would accumulate at these unprocessed DNA ends. Consistent with this hypothesis, ChIP-seq for MRE11 revealed significant chromatin binding at all hotspots in $Atm^{-/-}$ cells (Fig. 5d, right, Supplementary Fig. 4b), which correlated with END-seq intensity (Supplementary Fig. 4c). This likely reflects a role for ATM in promoting normal MRE11 resection initiation as suggested for Tel1 in yeast[6]. In contrast, we detected no signal in similarly broken regions in WT cells (Fig. 5d, left), indicative of completed resection. Strikingly, the MRE11 signal in ATM-null cells was contained within the WT read-less gap (Fig. 5d, e, Supplementary Fig. 4b). This lends further support to the idea that the gap reflects MRE11-dependent end-processing. Interestingly, MRE11 binding greatly diminished in the small central H3K4me3/K36me3 nucleosome-depleted area (NDR) (Fig. 5d, e and Supplementary Fig. 4b). The NDR in $Atm^{-/-}$ cells may be depleted of DNA owing to SPO11 double-cutting (see below).

END-seq central peak reflects DNA-bound SPO11. At all hotspots, END-seq detected a uniform accumulation of reads aligned to the center of the DSB. This central peak was coincident with mapped SPO11 oligos, having a width (400 bp) similar to SPO11-oligo seq hotspots (300–400 bp) and was restricted to the NDR of H3K4/K36me3 (Fig. 2d,f and Supplementary Fig. 2b)[8]. Moreover, even low-level secondary oligos that are adjacent to the central SPO11 hotspot peak[8] were detectable by END-seq (Fig. 2d). This suggested that the central peak represented a fraction of the total DSBs in the spermatocyte population in which SPO11 is not yet released, thereby highlighting the heterogeneity of cellular DSB processing.

If 5′ covalently bound SPO11 DSBs existed at the time of END-seq processing, then the proteinase K digestion would leave behind a two-bp 5′ overhang with a phosphotyrosyl bond that requires ExoVII digestion to fully blunt the end, analogous to our studies on TOP2 cleavage complexes (TOP2cc)[16]. To test this hypothesis, we performed END-seq with ExoT blunting only, allowing adapter ligation only to protein-free DNA ends while excluding ends with protein adducts, such as any remaining trapped SPO11 cleavage complexes (SPO11cc). Strikingly, ExoT detected only fully resected DNA ends with a total absence of central signal at all hotspots (Fig. 6a and Supplementary Fig. 4d). These data indicate that the central peak, which represents ~ 11% of the total DSB signal and is present all hotspots, reflects SPO11 covalently bound to its break site.

To validate that the ExoVII-dependent central peak was owing to SPO11 bound to the break and not some other form of occlusion, we speculated that it might be possible to remove SPO11 tyrosyl-linked DNA through incubation with purified human MRE11/RAD50/NBS1 (MRN) and CtIP prior to blunting with ExoVII + ExoT, mimicking the in vivo processing of meiotic DSBs. Indeed, we observed a dramatic loss of central signal by preincubation with MRN and CtIP (Fig. 6b)[39]. Incubation with MRN alone (prior to ExoVII + ExoT) or MRN and CtIP in the absence of manganese did not efficiently remove the central peak (Fig. 6b and Supplementary Fig. 4e) consistent with the finding that CtIP and manganese is required for the in vitro MRE11 endonuclease processing of protein-bound DSBs[40,41]. Thus, purified MRN + CtIP recognizes and removes the remaining SPO11 phosphotyrosyl bonds associated with the central signal. These data therefore support the idea that a fraction of SPO11 remains physiologically bound to a subset of breaks (~ 11%) at virtually all hotspots after cutting.

Increased fraction of unresected SPO11-bound DSBs in $Atm^{-/-}$. Because Tel1 regulates MRE11-initiated resection, there is a marked increase in unresected DSBs in Tel1-deficient yeast[6]. If ATM functions similarly to Tel1, then we would expect an accumulation of unresected SPO11cc, well above WT levels, at the center of the hotspots. The abundant MRE11 ChIP-seq signal specifically in $Atm^{-/-}$ cells (Fig. 5c, d), indicative of incomplete processing, would also predict a vast increase in SPO11cc in the mutant.

We therefore sought to modify END-seq to specifically probe for SPO11cc. We hypothesized that preincubation with purified tyrosyl-DNA phosphodiesterase 2 (TDP2)[42] would remove the remaining phosphotyrosyl adduct after proteinase K treatment, generating a two-nucleotide, protein-free 5′ overhang that ExoT could readily blunt for adapter ligation. Although ExoT alone detected no central peak in ATM-null cells, similar to WT (Fig. 6c), TDP2 + ExoT END-seq captured an astonishingly robust central signal (Fig. 6c). Moreover, this peak was strongly detected at all hotspots (Fig. 6c, right). SPO11cc detection by TDP2 + ExoT far exceeded the efficiency of detection with ExoVII + ExoT in $Atm^{-/-}$ spermatocytes both at autosomes and at the non-PAR X chromosome (Supplementary Fig. 4f, g), most likely reflecting the biochemical preference for TDP2 over ExoVII to remove the phosphotyrosyl bonds and allow for adapter ligation[42]. These data indicate that ATM-deficient spermatocytes accumulate unresected SPO11cc, similar to yeast lacking Tel1[6].

In WT cells, the ExoVII + ExoT central signal, which is also SPO11-bound, represents 11% of the total DSB fraction. Yet, in contrast to $Atm^{-/-}$, combination of TDP2 and ExoT only recovered a small fraction (16%) of the ExoVII + ExoT central signal found in WT cells (Fig. 6d), corresponding to 0.5–2% of the total DSB signal. Given the high efficiency of TDP2 + ExoT in ATM-null cells for detection of SPO11cc, it was initially unclear why the WT TDP2 + ExoT central signal was so low. We

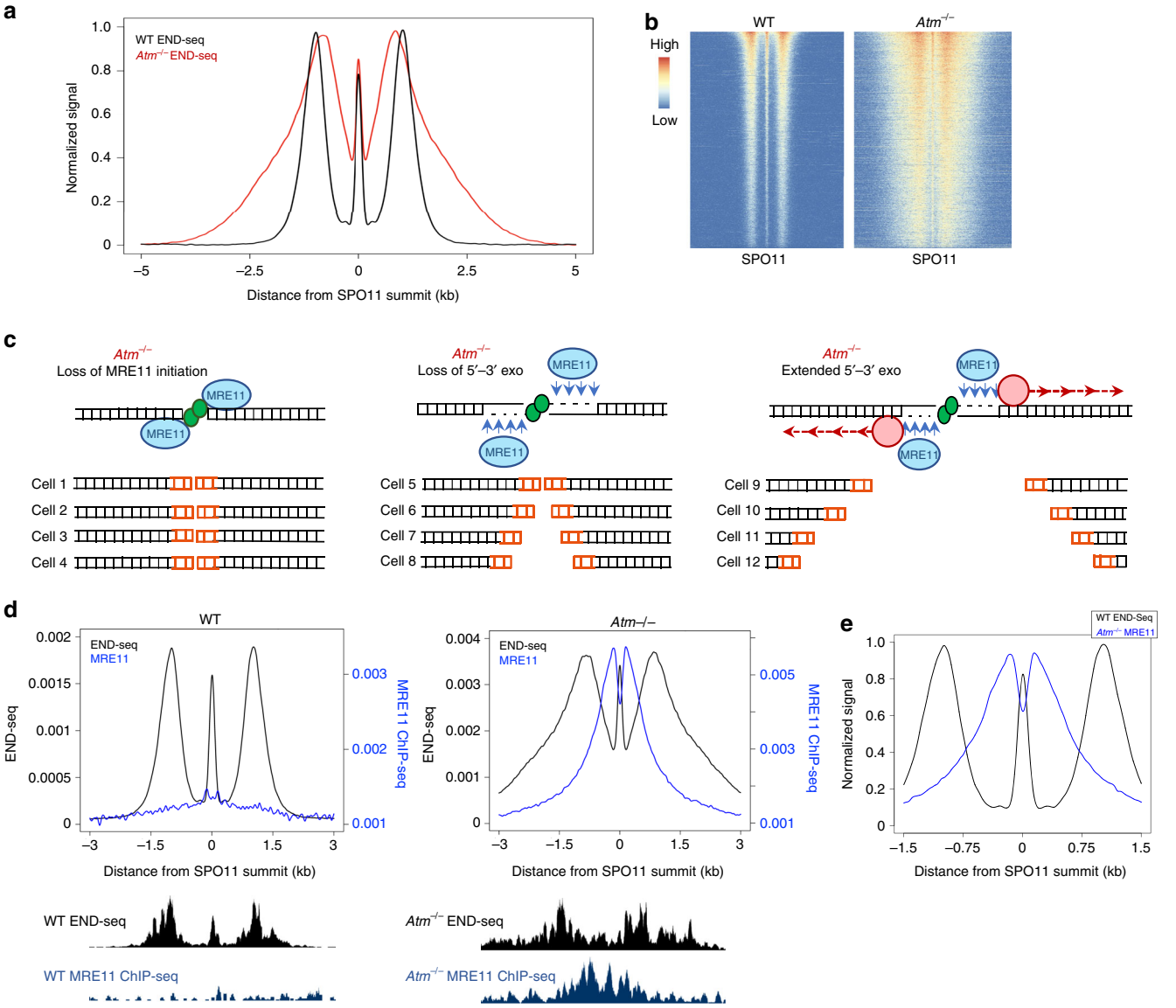


Fig. 5 ATM coordinates multiple levels of resection machinery. a Aggregate plots of END-seq signal in WT vs *Atm*$^{-/-}$ at top 5000 WT breaks around SPO11-oligo summits. Signals are normalized to the same height for DSB pattern clarity. **b** Heatmaps of WT and *Atm*$^{-/-}$ END-seq ± 5 kb around SPO11-oligo summits for top 5000 WT hotspots. Signals are normalized to spike-in control to show increased *Atm*$^{-/-}$ break intensity per hotspot. **c** Illustrations of ATM's multiple roles in coordinating resection that gives overall heterogeneous END-seq pattern in **a**. Left: MRE11 is recruited yet not activated in a subset of cells, resulting in reads directly at SPO11 DSB within the population. Middle: MRE11 is activated in another subset of cells, yet long-range resection is not sufficiently initiated, resulting in reads from only MRE11 nicking, and perhaps 3′−5′ exonuclease activity, flanking the DSB in regions that are read-less in WT. Right: resection is properly initiated in another subset of cells, yet long-range resection travels significantly farther than in WT. **d** Aggregated END-seq signal and MRE11 ChIP-seq RPM in WT (left) and *Atm*$^{-/-}$ (right). To fairly compare ChIP-seq signal between WT and *Atm*$^{-/-}$, MRE11 scale is proportional to spike-in normalized END-seq RPM for each genotype. Individual hotspot examples (chr12:34,592,264-34,598,265) are shown below. Note that decreased MRE11 coverage is observed within NDR of *Atm*$^{-/-}$. **e** Aggregate plot overlapping WT END-seq and *Atm*$^{-/-}$ MRE11 ChIP-seq, normalized to the same height. MRE11 shows prominent localization to the WT END-seq read-less gap.

speculated that in WT cells, the central signal might not simply reflect unresected SPO11cc, as observed in *Atm*$^{-/-}$ spermatocytes. Rather, there might be an additional structure at the break site associated with SPO11-bound DNA in WT cells that somehow prevented recognition by TDP2.

DNA-bound SPO11 is dependent on homolog engagement and PRDM9. Our first clue to understanding the distinct nature of SPO11-bound DNA in WT vs. *Atm*$^{-/-}$ cells was the observation that it was missing from all hotspots in *Dmc1*$^{-/-}$ spermatocytes (Fig. 7a). This prompted us to ask whether the central peak associated with DNA-bound SPO11 was also dependent on ssDNA strand invasion into the homologous partner. Remarkably, we found a loss of central signal at all non-PAR X chromosome hotspots (Fig. 7b), which repair from the sister chromatid as the X chromosome has no homolog in males. The residual central signal resembled the TDP2 + ExoT signal in WT cells (Fig. 6d) and therefore might simply reflect the amount of unresected SPO11 naturally present on all chromosomes. In striking contrast, the central signal associated with autosomes and the non-PAR X chromosome was virtually identical in *Atm*$^{-/-}$ spermatocytes, again confirming that the central signal is largely SPO11cc in the absence of ATM (Supplementary Fig. 4h–i).

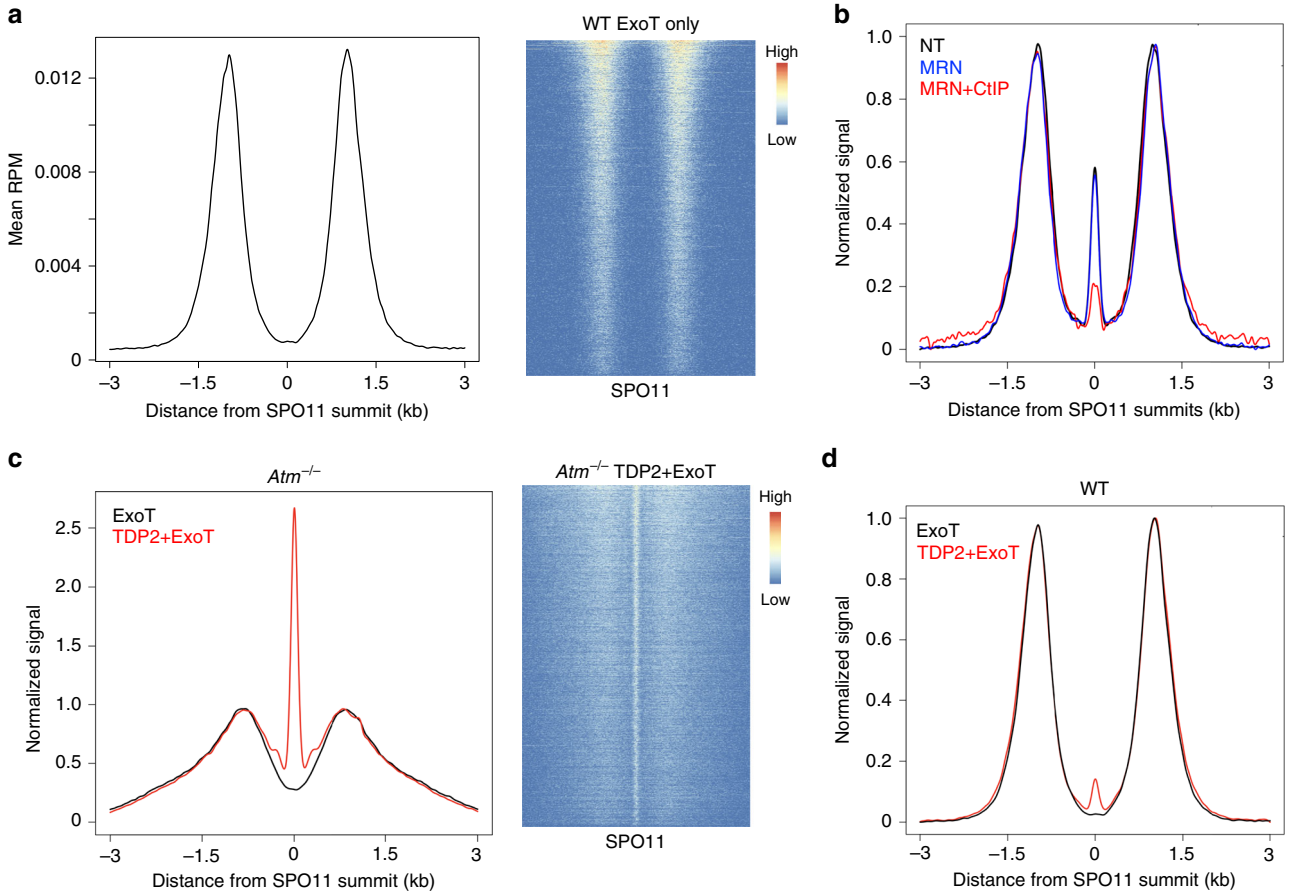


Fig. 6 WT and *Atm*$^{-/-}$ hotspots contain distinct species of DNA-bound SPO11. a END-seq processing with ExoT blunting alone shows total absence of SPO11 central peak when signal is aggregated around SPO11-oligo summits (left). Heatmap of ExoT only signal ± 3 kb around hotspot centers, ordered by total read count of END-seq (right). **b** Pretreatment with purified human MRN + CtIP reduces ExoVII + ExoT central peak detection (red) over no pretreatment (NT, black) and depends on the presence of CtIP (blue). Both MRN and MRN + CtIP reactions were carried out in the presence of manganese. One 12 dpp mouse used per condition. **c** END-seq with ExoT blunting alone also shows no central peak in *Atm*$^{-/-}$ cells (black line). Pretreatment with purified human TDP2 before ExoT incubation recovers SPO11 signal, indicating abundance of unresected, bona fide SPO11 cleavage complexes (red line). Aggregate plot (left, normalized to same resection height) and heatmap (right) ± 3 kb around SPO11-oligo summits. **d** END-seq aggregate plot of WT ExoT blunting with (red line) and without (black line) TDP2 pretreatment shows only minor recovery of central peak signal.

Finally, we found that the central signal was absent in *Hop2*$^{-/-}$ spermatocytes (Fig. 7c), which fully load recombinases but fail to engage the homologous template[36].

The analysis of hybrid mouse strains with different PRDM9 alleles has revealed that the degree of asymmetry in PRDM9 binding—that is, whether PRDM9 binds unequally to both homologs—predicts increased asynapsis and hybrid infertility[13,14]. When PRDM9 fails to bind the unbroken homologous chromatid, there is a severe reduction in both CO and NCO events[14,15]. Indeed, it has been suggested that similar to the X chromosome, asymmetric hotspots might be repaired from the sister chromatid[15]. We therefore performed END-seq on spermatocytes derived from juvenile B6xCAST hybrids with divergent genomes. Strikingly, the central peak was reduced to 2–3% of the total DSB signal in B6xCAST compared with the 11% observed in B6 (Fig. 7d) and much lower than in CAST alone (Supplementary Fig. 4j). Altogether, these findings support the idea that the central signal detected by ExoVII+ExoT in WT spermacotyes is associated with DNA-bound SPO11, and is dependent on the degree of homologous chromosome engagement.

In mice lacking PRDM9, DSBs occur at H3K4me3 sites mainly associated with promoters[9,11,12]. However, these DSBs are not repaired efficiently as COs resulting in meiotic arrest[11]. To test how PRDM9 deficiency impacts homolog engagement and resection, we performed END-seq on *Prdm9*$^{-/-}$ spermatocytes. As SPO11 can generate multiple breaks within the H3K4me3 sites at promoters, we focused our analyses on *Prdm9*$^{-/-}$ SSDS sites that exhibited the least overlap between top and bottom strand reads, i.e., hotspots that are most likely to have one main SPO11 cut site within the promoter (Supplementary Fig. 5a). Such SSDS hotspots showed limited signal at the center, in line with these sites having a strong preference for a single SPO11 cut site (Supplementary Fig. 5a, b). Strikingly, END-seq analysis of *Prdm9*$^{-/-}$ spermatocytes revealed a total absence of the central peak at hotspots with the least SSDS strand overlap, examined either on aggregate or individually in the genome browser, whereas short- and long-range resection appeared to be relatively intact (Fig. 7e, f and Supplementary Fig. 5c, d). The absence of central signal is consistent with the idea that PRDM9 promotes homolog engagement, which in turn facilitates COs. Alternatively, delay in homolog engagement might allow more time for processing DNA-bound SPO11, resulting in loss of SPO11-bound DNA.

SPO11 is associated with a recombination intermediate. At a meiotic DSB that has been fully resected on both sides and SPO11

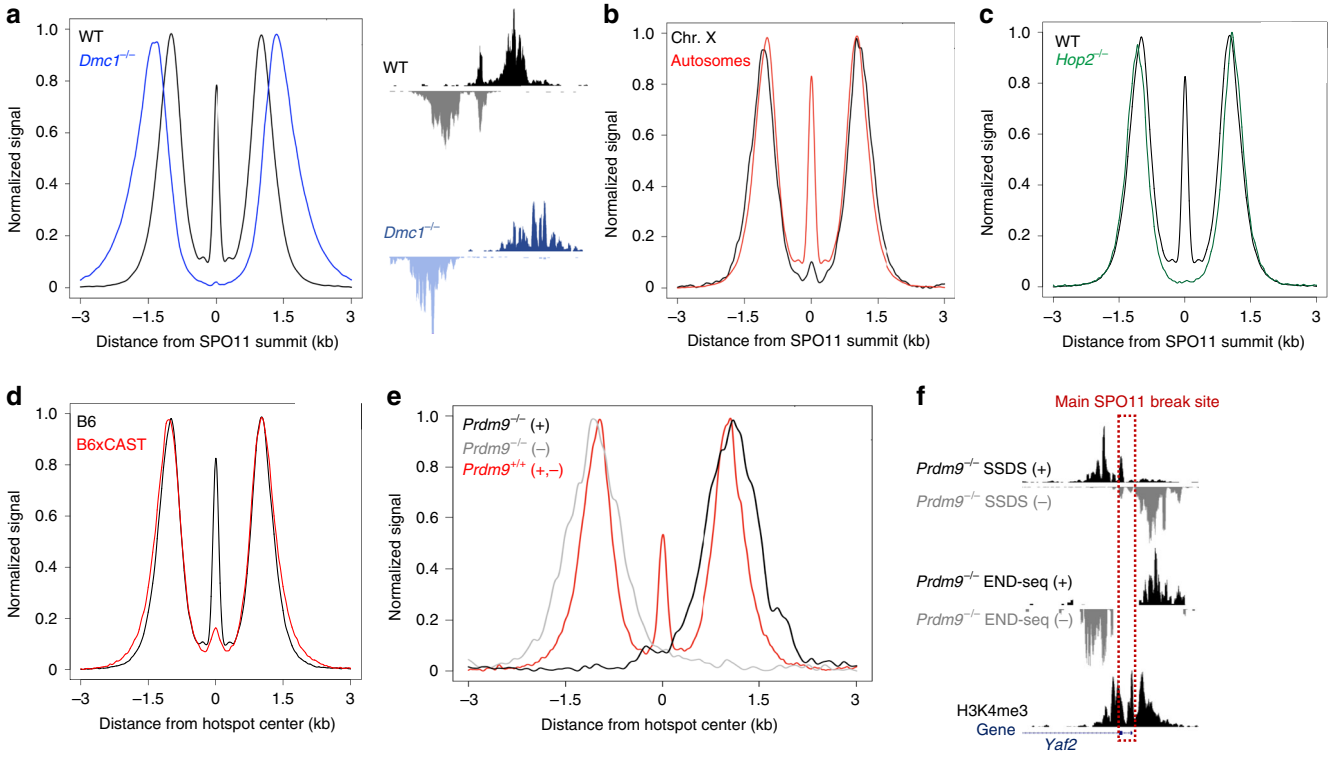


Fig. 7 WT central peak relies on homolog engagement and PRDM9. a Aggregated END-seq signal around SPO11 oligos, with hotspot example, showing absence of SPO11 central peak in *Dmc1*$^{-/-}$. **b** Reduction in central signal at non-PAR X chromosome hotspots; aggregated signal on ChrX compared with all autosomes. **c** Aggregated END-seq signal around SPO11 oligos showing absence of SPO11 central peak in *Hop2*$^{-/-}$. **d** Aggregated END-seq signal of B6 (centered on B6 SPO11 oligos) versus B6xCAST hybrid (centered on hybrid SSDS hotspot centers). **e** *Prdm9*$^{-/-}$ END-seq signal aggregated around default SSDS hotspot centers at ~200 sites with least overlap in SSDS top and bottom strands. WT END-seq is centered around WT SPO11 oligos at PRDM9-dependent hotspots. **f** *Prdm9*$^{-/-}$ SSDS and END-seq tracks at a single default hotspot (*Yaf2* gene) with minimal SSDS top and bottom strand overlap. Main SPO11 break site (red dotted line) is inferred from SSDS pattern. Aggregate plots in all panels are normalized to the same height for visual comparison.

completely released, adapters ligated to the right end of the break will align to the top (+) DNA strand, whereas adapters ligated to the left end will align to the bottom (–) strand (Fig. 2c, Fig. 8a, left). When reads from all hotspots are aggregated, the distal resection signal around SPO11 cuts exhibited this "correct" polarity for the top and bottom strands (Fig. 8a, left). However, a close examination of the WT END-seq reads associated with the central signal unexpectedly revealed a "wrong" polarity, in which top strand reads aligned slightly to the left within the NDR and bottom strand reads aligned slightly to the right (Fig. 8a, right). As SPO11 generates a DSB with only 2 nt overhangs, if the central signal were merely a collection of unresected SPO11cc, then aggregating the top and bottom strand DSB endpoints should show no separation between them. However, we observed a top and bottom strand shift of ~60 nts in the "wrong" orientation (Fig. 8a, right). Detecting a significant reversal in the expected polarity indicated that the central peak was not simply SPO11cc, which should generate a canonical DSB pattern (see *Atm*$^{-/-}$ below). This suggested that although SPO11 remains bound to a fraction DSBs, there is some kind of asymmetry associated with DNA-bound SPO11 that influenced END-seq detection compared with when SPO11 is released and both DNA ends are fully resected (Fig. 8a).

One potential mechanism that could contribute to the wrong polarity is asymmetric processing of SPO11. That is, if the two DNA ends bound by SPO11 are processed at different efficiencies by MRE11, one end might be incompletely processed, leaving SPO11 covalently bound to its cut site, whereas the other end is processed to completion and SPO11-oligo released. In this scenario, only the incompletely processed SPO11-bound end would contribute to central peak signal. The fully processed end (Fig. 8b, top, left end of the DSB) would result in the release of the SPO11-oligo that would generate a protein-free 3′ overhang. This (SPO11-free) DNA end would in turn be blunted by END-seq and sequencing reads would be detected within the distal, long-range resection peaks (Fig. 8b, bottom, left end of the DSB). In contrast, the other side of the DSB would be incompletely resected by MRE11 and retain SPO11 covalently bound to a two-nucleotide, 5′ overhang (Fig. 8b, top, right end of the DSB). END-seq detection (with ExoVII + ExoT) would then remove SPO11 and sequence the remaining dsDNA, with the first nucleotide sequenced being the SPO11 break site itself (Fig. 8b, bottom, right end of the DSB). This would result in top and bottom strand central peak reads with reversed polarity within the NDR, as SPO11 breaks to the left of the NDR center would contribute top strand reads aligning left of center, and SPO11 breaks to the right would contribute bottom strand reads aligning right of center (Fig. 8b). In a population of spermatocytes in which both events occur among the cells, END-seq would detect an overall aggregated signal of resection reads with correct polarity and central reads with reversed polarity (Fig. 8b, bottom).

How could such asymmetric MRE11-mediated processing arise? Most SPO11-oligo-sequencing reads cluster in the center of the nucleosome-free depleted region where PRDM9 is also bound, suggesting that PRDM9 does not block SPO11 access[8,43]. Rather, we imagine that DNA-bound PRDM9 may guide the position at which SPO11 cuts within the nucleosome-free region, which might be slightly displaced on average by 30 base-pairs (half the size of 60 bp shift) from PRDM9 itself (Fig. 8b). It has been suggested that PRDM9 often remains bound on the uncut

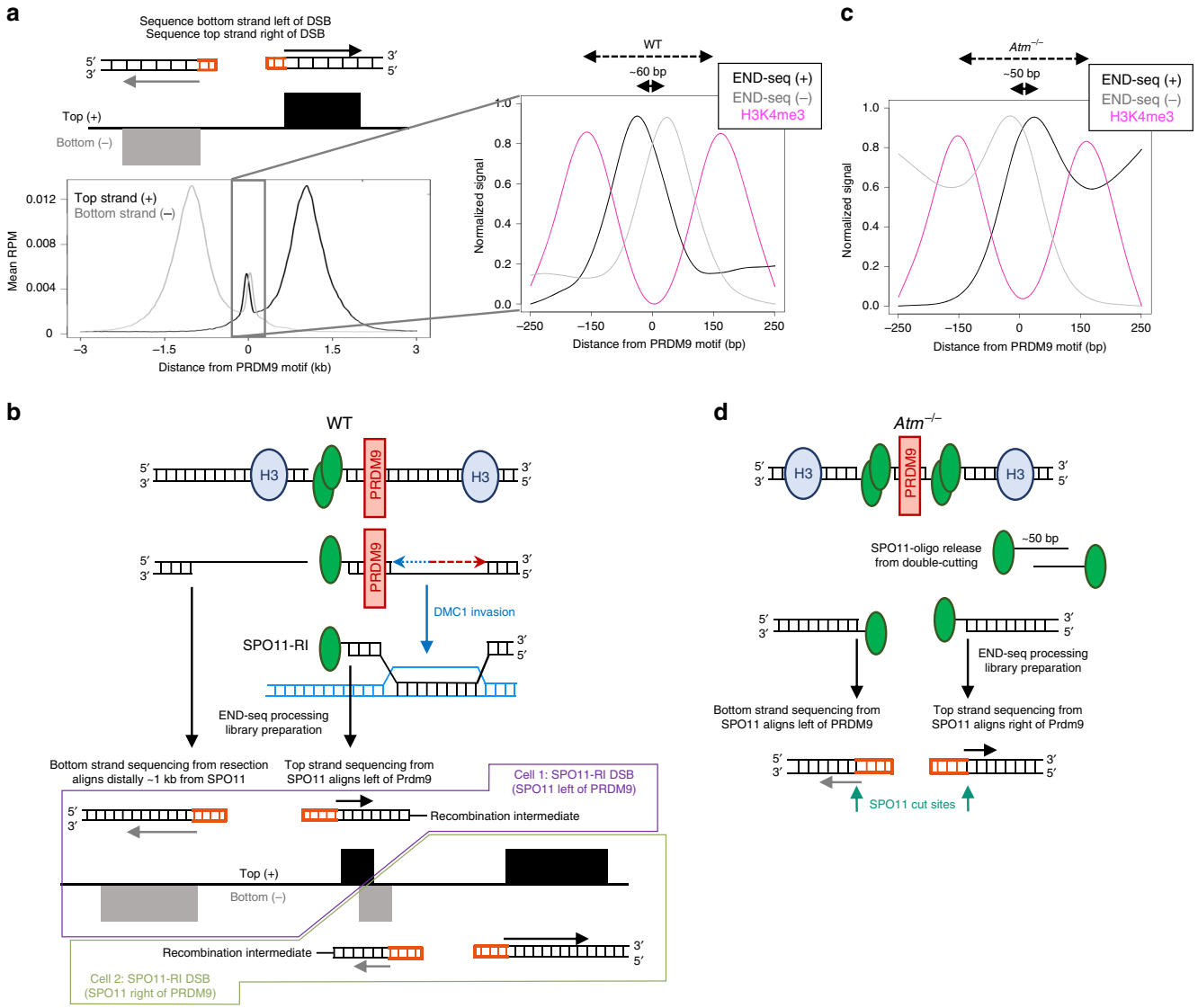


Fig. 8 Polarity of END-seq reads reflects distinct natures of SPO11 cutting and processing in WT and ATM-null cells. a At fully processed and resected SPO11 DSBs, END-seq top and bottom strand reads exhibit a "correct" polarity to the right and left of the DSB, respectively (left). WT reads at the center of hotspots show a strand polarity that is reversed from what is expected (right, zoomed in at NDR). **b** PRDM9 as a barrier to SPO11 processing that results in a SPO11-bound recombination intermediate (SPO11-RI) structure. In WT, SPO11 cutting to one side of chromatin-bound PRDM9 within the NDR may block MRE11 activity on one side of the break, leaving SPO11 covalently bound to a short stretch of dsDNA, capping the DMC1-loaded ssDNA that extends ~1 kb from the NDR. SPO11-RI is sequenced starting from the first SPO11-bound nucleotide where an adapter is ligated. SPO11 that cut left of PRDM9 would result in top strand reads aligning to the left of the PRDM9 motif and bottom strand reads from fully resected ssDNA aligning at a distance away from the break site. **c** *Atm*$^{-/-}$ END-seq central reads have correct separated strand polarity within the NDR. **d** Unresected SPO11 double-cutting within the same hotspot in *Atm*$^{-/-}$ cells would show the correct polarity of top and bottom strands after adapter ligation to SPO11-bound DSBs and ~50 bp separation, as observed in **c**. Decreased MRE11 activity at these breaks would result in the direct sequencing of SPO11 cleavage complexes within the NDR rather than SPO11-RI in WT.

chromosome[14], whereas SPO11 has been proposed to be associated with DNA ends[19] until or even subsequent to strand invasion[14,19]. If PRDM9 similarly remains bound to DNA post-cleavage between the SPO11 cut and MRE11-endonucleolytic nicking position (Fig. 8b), it could interfere with MRE11 release of SPO11. This would prevent MRE11 from generating a fully ssDNA overhang only on one side of the DSB (Fig. 8b, top). Because the natural on/off binding affinity of PRDM9 would determine the frequency at which MRE11 activity is blocked and SPO11-bound DNA is captured by END-seq, we would expect that the central peak to be detected at all hotspots genome-wide, as observed (Fig. 2b). Moreover, all hotspots had equal ratios of central peak to resection signal (~11%), indicating that no hotspot had preference over others, regardless of break frequency.

The central signal not only reflected asymmetric MRE11-mediated processing, but also required DMC1- and HOP2-mediated strand invasion and engagement with the homologous chromosome template (as shown above). Owing to this dependency, we infer that SPO11 remains bound post homolog engagement and during the formation of a recombination intermediate (RI). We therefore refer to this RI, with SPO11 capping the 3′ resected end, as SPO11-RI (Fig. 8b). Consistent with our results, an independent genomic sequencing method (S1-seq) corroborates the presence of SPO11-RI with a reversed central signal polarity in WT mouse spermatocytes[38].

Increased SPO11 double-cutting at the same hotspot in *Atm*$^{-/-}$. The central signal in ATM-null spermatocytes is largely comprised

of unresected SPO11cc, which accumulates MRE11 in vivo and is sensitive to TDP2-mediated processing (Figs. 5d, 6c, and Supplementary Fig. 4f–i). This is distinct from the enrichment of SPO11-RI observed in WT cells. Elevated levels of SPO11cc in *Atm*$^{-/-}$ cells could arise from decreased MRE11 endonucleolytic or 3′−5′ exonuclease activity (Fig. 5c, d), which is also characteristic of yeast *Tel1* deficiency[6]. If *Atm*$^{-/-}$ SPO11cc reflects fully unresected DSBs and not SPO11-RI, then the aggregated END-seq signal should have no obvious polarity. To examine this, we strand separated the central signal in *Atm*$^{-/-}$ cells. Unexpectedly, the separated strands exhibited a ~50 base-pair gap within the nucleosome-depleted region with the correct DSB polarity (Fig. 8c). If SPO11 cut once on each chromatid throughout the NDR within the population of cells and remained bound to DNA, there would be a 2-nt gap between top and bottom strand DSB endpoints. We therefore infer that the larger gap size reflects frequent SPO11 double-cutting within the same hotspot (Fig. 8c, d)[44]. These measurements are consistent with the increased 40–70 nucleotide SPO11-oligo species that were detected in ATM-null mice[7], as two distinct SPO11 cuts adjacent to one another could release these longer oligos even without MRE11 endonuclease activity (Fig. 8d). Moreover, MRE11 ChIP-seq revealed a notable dip in MRE11-binding exactly within the NDR (Fig. 5d, e and Supplementary Fig. 4b), consistent with the loss of DNA within hotspot centers. These results are also supported by genetic evidence of double-cutting in *Atm*$^{-/-}$ spermatocytes (A. Lukaszewicz, S. Keeney and M. Jasin, personal communication).

Increased double-cutting around PRDM9-binding sites would preclude it from blocking any MRE11 short-range resection that does occur, thereby reducing the frequency of SPO11-RIs in ATM-null cells (Fig. 8d). Our finding that central signal in *Atm*$^{-/-}$ spermatocytes exhibits the correct polarity (Fig. 8b), and that this signal is identical on the autosomes and non-PAR X chromosome (Supplementary Fig. 4h, i), is consistent with a significant reduction in SPO11-RI in *Atm*$^{-/-}$ spermatocytes. Therefore, through its regulation of SPO11 cutting and resection, ATM indirectly regulates the formation of SPO11-RI.

Discussion

We show here the capacity of END-seq to elucidate early meiotic pathways that are critical for proper chromosome segregation and fertility. Because the method requires little starting material (as low as a single mouse) and can be modified through differential enzymatic reactions to detect distinct DSB structures, we are able to bypass the limitations imposed by previous meiotic hotspot profiling techniques. In doing so, we uncovered a pattern of resection, strikingly uniform at all DSBs. This reflects the fact that short- and long-range resection are tightly coupled in a single processive reaction that is mediated by ATM. This pattern is highly reminiscent of meiotic resection in yeast, and reflects the evolutionary conservation of DSB processing pathways.

Although yeast do not possess obvious homologs of BRCA1, BRCA1 has a well-described function in supporting the resection of DSBs in somatic mammalian cells[32]. Surprisingly, we find that BRCA1 does not similarly promote resection during meiotic recombination. One potential reason could be that 53BP1 is not recruited to DSBs in mitotic cells or in early prophase meiocytes[45,46], which may reflect similarities in pathways that suppress DSB repair during mitosis and meiosis. If the primary function of BRCA1 is to counteract 53BP1's block to resection at DSBs, this function of BRCA1 would not be needed during mitosis or meiosis.

In addition to promoting resection in somatic cells, BRCA1 also facilitates the loading of RAD51/DMC1 onto ssDNA[47], a function which appears to be conserved in meiosis[35]. In several BRCA1-deficient mouse strains with known defects in RAD51/DMC1 filament formation, we observed a slight increase in resection tracts. We suggest that this reflects an indirect role for BRCA1 in limiting resection by promoting recombinase loading onto ssDNA. Consistent with this idea, *Dmc1*$^{-/-}$ but not recombinase-proficient *Hop2*$^{-/-}$ spermatocytes displayed a marked increase in minimum and maximum resection endpoints. We also observed hyper-resection in ATM-null spermatocytes, which we propose is owing in part to the limited availability of recombinases to form filaments in the prescence of excessive SPO11 cutting. Thus, we imagine that in BRCA1−, DMC1−, and ATM-deficient spermatocytes, defective recombinase loading permits reiterative engagement of the long-range resection machinery.

Although some aspects of resection are highly conserved in yeast and mammals, SPO11 processing appears not to be. Unlike yeast, mammalian cells accumulate significant levels of DNA-bound SPO11 that represent SPO11-RI. Our model suggests that incomplete processing of SPO11 in mammals, perhaps owing to PRDM9 blocking, still allows for, and may even promote, strand invasion via DMC1 into a homologous template (Fig. 8b). Although SPO11-RI is readily detectable by END-seq using ExoVII + ExoT and represents 11% of the total DSB signal, processing with TDP2 + ExoT only detects SPO11cc (Fig. 6d). We therefore estimate that out of the total DSB signal detected by ExoVII + ExoT in WT cells, 0.5–2% is truly unresected SPO11cc, as detected by TDP2 + ExoT, and the remaining ~ 10% is SPO11-RI.

What is the biological relevance of SPO11-RI? One possibility is that SPO11-RI facilitates NCO and/or CO events necessary for the exchange of genetic information between homologs. The majority of NCO and CO events occur only in the 4000 hottest hotspots[15], and we have found that SPO11-RI occurs across all 5000 DSBs detected by END-seq (Fig. 2b). NCO and CO events are reduced in strains with asymmetric PRDM9 hotspots, and we found strong depletion of the central signal in hybrid strains. CO were less likely to occur in gene-rich regions[15], and we have found that the central signal is decreased with increasing gene expression independent of DSB formation (Supplementary Fig. 5e, f). Finally, the positioning of NCO and CO tracts are highly enriched in a small region (generally <500 bp) surrounding the PRDM9 motifs[15,21]. It is possible that the "capping" of the 3′ ssDNA by SPO11 (Fig. 8b) delays polymerization or double Holliday junction migration, therefore restricting SPO11-RI to a 400 bp region surrounding the PRDM9 motif. Although SPO11-RI correlates with NCO and CO events, additional studies will be necessary to determine its precise physiological function.

The loss of the central signal in mouse hybrids, in *Dmc1*$^{-/-}$ mice, and on the WT non-PAR X chromosome could potentially be explained if DSB repair is delayed. With a fixed on/off rate for PRDM9, the longer DSBs persist, the more time MRE11 may have to process and remove SPO11-bound DNA, as PRDM9 will eventually exit the hotspot. Similarly, in PRDM9-deficient mice with defective repair, DSBs persist longer, and SPO11-RI is absent.

We find limited evidence of SPO11-RI in ATM-null cells. Instead there is a concomitant increase in SPO11cc and MRE11 binding owing to incompletely processed DSBs. Recent studies have demonstrated that some of these DSBs arising in *Atm*$^{-/-}$ mice can be repaired by detrimental non homologous end-joining (NHEJ) (A. Lukaszewicz, S. Keeney and M. Jasin, personal communication). Given the marked in vitro activity of TDP2 in *Atm*$^{-/-}$ spermatocytes, SPO11cc might be rejoined in part by TDP2-dependent NHEJ in vivo. Direct hydrolysis of PRDM9-blocked SPO11 by TDP2 in WT cells would also likely result in aberrant end-joining. However, we found that purified TDP2 does not act on SPO11-RI in vitro (Fig. 6d). Although it remains unclear why SPO11-RI is an inefficient substrate for TDP2, it is possible that TDP2 cannot recognize SPO11-RI owing to steric hindrance associated with the heteroduplex DNA. If TDP2 were similarly inactive on SPO11-RI

in vivo, the formation of SPO11-RI could serve to prevent NHEJ and instead favor repair towards high-fidelity HR.

In conclusion, both ATM and PRDM9 orchestrate mammalian SPO11 processing in a manner that influences meiotic DSB repair.

Methods

Mice. C57BL/6 mice (Jackson Laboratory) were bred in-house, and male pups were killed at 12–14 dpp for testes isolation. *Atm*$^{-/-}$ mice (Barlow et al. 1996) were backcrossed eleven generations to C57BL/6 mice. *Exo1*$^{-/-}$ mice were a gift from Winfried Edelmann. *Dmc1*$^{-/-}$ mice were a gift from Scott Keeney. *53bp1*$^{-/-}$ mice were a gift from Junjie Chen. *Prdm9*$^{-/-}$ mice were a gift from Petko Petkov. *Brca1*$^{+/\Delta 11}$ mice were obtained from the NCI mouse repository. *53bp1*^{S25A} mice were generated as described[33]. *p53*$^{-/-}$ mice were purchased from Taconic Biosciences. Strains were crossed to generate *Brca1*$^{\Delta 11}$*53bp1*$^{-/-}$, *Brca1*$^{\Delta 11}$*53bp1*^{S25A} and *Brca1*$^{\Delta 11}$*p53*$^{+/-}$ mice. CAST/EiJ (males) were purchased from Jackson Laboratory and were crossed with C57BL/6 (females) for one generation to produce B6xCAST F1 hybrids, respectively. All mouse breeding and experimentation followed protocols approved by the National Institutes of Health Institutional Animal Care and Use Committee.

Mouse testicular cell isolation. Adapted from Baker et al. 2014. Testes were dissected from 12–14 dpp male mice and placed into a 6 cm tissue culture dish containing Dulbecco's Modified Eagle Medium (DMEM). Tunica albuginea were removed under a microscope, and tubules were gently dissociated with forceps and placed into 50 mL tube containing 20 mL DMEM. After tubules settled to the bottom of the tube, DMEM was aspirated and replaced with 20 mL DMEM containing 0.5 mg/mL Liberase TM (Roche, 5401127001) and incubated at 32 °C for 15 min at 500 rpm. Tubules were washed once with fresh DMEM, replaced with 20 mL DMEM containing 0.5 mg/mL Liberase TM and 100 U DNase I (ThermoFisher, EN0521), and incubated at 32 °C for 15 min at 500 rpm.

Tubules were disrupted by gentle pipetting and passed through a 70-µm Nylon cell strainer (Falcon) repeatedly until tissue debris was fully removed. Cells were pelleted at 1500 rpm at 4 °C for 5 min and washed with 10 mL DMEM. Cells were filtered through a 40 µm Nylon cell strainer (Falcon) repeatedly until debris was fully removed and pelleted at 1500 rpm at 4 °C for 5 min. Cells were resuspended in 1 mL phosphate-buffered saline (PBS) and counted.

END-seq. Single-cell suspensions of bulk testicular cells were immediately embedded after isolation into 0.75% agarose plugs. After isolation, cells in 1 mL PBS were diluted to 5–7 million bulk cells/mL PBS and separated into 1 mL of cells per 1.5 mL tube for plug making. Spike-in cells were added at 2% of bulk cell number per tube/plug. Multiple plugs were made per sample if necessary, depending on number of mice and total cell number isolated, processed in the same tube, and DNA later combined after plug melting.

A detailed description for embedding cells into agarose plugs and general END-seq procedure can be found in Canela et al.[16], Canela et al.[17]. In brief, agarose embedded cells were solidified at 4 °C for 15 min then immediately lysed and digested with Proteinase K at 50 °C for 1 hr then 37 °C for 7 hr. Plugs were then washed with TE, treated with RNase at 37 °C for 1 h, and stored at 4 °C for no longer than 1 week before the next series of enzymatic reactions.

Unless otherwise noted, plugs were treated with sequential combination of Exonuclease VII (NEB) for 1 h at 37 °C followed by Exonuclease T (NEB) for 45 min at 24 °C to blunt DNA ends before Illumina adapter ligation[16]. For experiments in which only Exonuclease T was used to blunt, plugs were digested with Exonuclease T for 1 h at 24 °C. For experiments with purified human TDP2 (gift from Keith W. Caldecott), plugs were treated with 500 pM TDP2 (in 50 mM Tris-HCl pH 8.0, 10 mM $MgCl_2$, 80 mM KCl, 1 mM DTT, and 0.05% Tween−20) for 4 h at 24 °C, followed by Exonuclease T digestion for 1 h at 24 °C.

For experiments with purified human MRE11/RAD50/NBS1 (MRN) and CtIP (Tanya Paull lab), plugs were treated with 50 nM MRN (MR and NBS1 were pre-incubated for 10 min at 4 °C prior to reactions) with or without 80 nM CtIP (in 25 mM MOPS pH 7.0, 20 mM Tris-HCl pH 8.0, 80 mM NaCl, 1 mM DTT, 1 mM ATP, 5 mM $MgCl_2$, 1 mM $MnCl_2$, and 0.2 mg/mL bovine serum albumin) for 1.5 h at 37 °C. Plugs were treated again with proteinase K for 1 h at 50 °C to degrade any remaining bound protein, followed by Exonuclease VII and Exonuclease T digestion as previously described.

For all enzymatic reactions, subsequent steps of A-tailing, adapter ligation, plug melting, chromatin shearing, and second round of adapter ligation for sequencing were performed exactly as previously described[16,17].

ChIP-seq. Ten million bulk testicular cells were fixed in 1% formaldehyde (Sigma, F1635) at 37 °C for 10 min. Fixation was quenched with glycine (Sigma) at a final concentration of 125 mM. Cells were washed twice with cold PBS, and pellets were snap frozen on dry ice and stored at −80 °C until sonication. Sonication, immunoprecipation, and library preparation were performed as previously described in (Canela et al.[16]). In brief, frozen pellets were resuspended in 1 mL cold RIPA buffer (10 mM Tris-HCl pH 7.5, 1 mM ethylenediaminetetraacetic acid (EDTA), 0.1% sodium dodecyl sulphate, 0.1% sodium deoxycholate, 1% Triton X-100, and 1 Complete Mini EDTA-free proteinase inhibitor tablet (Roche)) and sonicated using a Covaris S220 at duty cycle 20%, peak incident power 175, and cycle/burst 200 for 30 min at 4 °C.

Chromatin was precleared with 40 µL prewashed Dynabeads Protein A (ThermoFisher) for 30 min at 4 °C, followed by incubation with 40 µL Dynabeads Protein A bound to either 6 µL anti-H3K4me3 (Millipore, 07–473) or 4 µL anti-MRE11 (Novus, NB100-142) overnight at 4 °C. Beads were washed and cross-linking reversed the next day as described in (Canela et al.[16]). Immunoprecipitated DNA was removed from beads and stored at −20 °C until library preparation, which was performed as described in (Canela et al.[16]).

END-seq data analysis. Reads were aligned to the mouse (GRCm38p2/mm10) genome using Bowtie version 1.1.2[48] and three mismatches were allowed and the best strata for reads were kept with multiple alignments (-n 3 -k 1 -l 50). Functions "view" and "sort" of samtools (version 1.6)[49,50] were used to convert and sort the mapping output to sorted bam file. Peaks were called using MACS 1.4.3[51]. As the strand separated aggregate plots (Fig. 2c) show ~ 2 kb distance, END-seq peaks were called using the parameters:–shiftsize = 1000, –nolambda, –nomodel, and–keep-dup = all. Peaks with >2.5-fold-enrichment are kept and those within blacklisted regions (https://sites.google.com/site/anshulkundaje/projects/blacklists) were filtered.

For estimation of total breaks and comparison between genotypes, we added a spike-in control into END-seq samples that consists of a G1-arrested Ableson-transformed pre-B cell line (*Lig4*$^{-/-}$) carrying a single zinc-finger-induced DSB at the TCRβ enhancer. This site is expected to break in all spike-in cells, which were mixed in at a 2% frequency with bulk testicular cells. END-seq signal was calculated, as RPKM, within ± 5 kb window around all hotspot centers. Total intensity was divided by the signal around the spiked-in breaks and then divided by 50 since the spiked-in was added at a 1:50 ratio (2%).

As the pattern of END-seq peaks are consisted of continuously sharp peaks and it is quite reproducible between experiments and to estimate the distance between the shar peaks and study its relation with nucleosome positioning, END-seq peaks were split as subpeaks by PeakSplitter tool of PeakAnalyzer with default parameters (https://www.bioinformatics.org/peakanalyzer/wiki/Main/Overview[52]) and the distances of sharp peak summits within each break were calculated.

Signal-to-noise analysis. ENCODE measures Signal-to-noise(S/N) ratio by fraction of reads in peaks (FRiP) and cross-correlation profiles (CCPs) for ChIP-seq. We used it for END-seq here.

The FRiP value is fractions of reads that mapped into called peaks without any filtering. The higher FRiP, the more enrichment. ENCODE recommends the threshold for FRiP is more than 1% for ChIP-seq.

The CCPs assess the quality of END-seq enrichment over background independent of peak calling. The plot shows the Pearson cross-correlations (CCs, *y* axis) of reads intensities between the plus strand and the minus strand, after shifting minus strand (*x* axis). There are two peaks, one corresponding to read length (CCread, blue dash line) and the other one corresponding to the fragment length (CCfrag, red dash line). Normalized strand coefficient (NSC) is CCfrag divided by minimal CC value (CCmin) and relative strand coefficient (RSC) is the ratio of CCfrag-CCmin divided by CCread-CCmin. The higher NSC and RSC, the more enrichment. ENCODE recommends an NSC ≥ 1.5 and RSC ≥ 0.8 for ChIP-seq.

The CCPs profile, NSC value and RSC value were generated by Phantompeakqualtools (https://github.com/kundajelab/phantompeakqualtools[22,53]). Different to ChIP-seq pattern that plus strand signal is at the left of minus strand signal, END-seq has opposite direction, so the shiftsize of the fragment is negative. Unsurprising, the absolute shift values are consistent with the distance of two long-range resection peaks showed by the aggregate plots.

Quantification of resections. Short-range resection was measured by a sliding window containing 10 bp bins, starting from the peak of long-range resection signal into the hotspot center. Mean value of ± 500 bp window around the peak site was used to estimate the normal distribution of the resection endpoints. Comparing the real signal of the sliding window to it, when the window got p value ≤ 0.05 by one-sample t test (alternative = "less"), the distance from hotspot center to the window midpoint was defined as the short-range resection length. Plus strand and minus strand were calculated separately.

For each hotspot, the long-range resection describes the distribution of resection endpoints among the cell population. Mean long-range resection was quantified based on the definition of expect value of a distribution as:

$$E(x) = \sum_{n=21}^{400} (x_i p_i)$$

RPM were calculated for 10-bp bins in a ± 4 kb window of each SPO11 summit. The first bins were not considered, the calculation started from the 21th bin. x_i is distance of the current bin to SPO11 summit and p_i is the probabilities calculated as RPM of the current bin/total RPM from 21th to 400th bins. Plus strand and minus strand were calculated separately. Top 250 breaks were used for the boxplot of all mean resection comparison.

To measure maximum long-range resection, a sliding window containing 20-bp bins was used, starting from the peak of long-range resection to further region until

more than half of bins within current window had lower intensity than background. The location of the last bin with a detectable signal over background is regarded as the maximum resection endpoint. Dynamic background was determined by the maximum intensity of 20-bp bins within 5 ~ 6 kb region away from SPO11 summit for individual hotspot. Plus strand and minus strand were calculated separately.

Estimation of SPO11-RI and unresected SPO11cc. The fraction of central signal detected by ExoVII + ExoT or TDP2 + ExoT in B6 was calculated using the total area substrate the area of ExoT blunting only after forcing them to the same height and then divided by ExoVII + ExoT total area. We got 11% for ExoVII + ExoT and 1.7% for TDP2 + ExoT. As the central signal of ExoVII + ExoT consists of both SPO11-RI and unresected SPO11cc, we used TDP2 to measure fraction of the unresected SPO11cc. The detection ability of ExoT for the small overhangs of SPO11cc and long overhangs of SPO11-RI may differ. Assuming ExoT detected unresected SPO11cc and SPO11-RI at the same level, the unresected SPO11cc fraction will be 1.7%. However, ExoT might work better for unresected SPO11cc detection than SPO11-RI and assume it captures all unresected SPO11cc. We first quantified the total break number for ExoT blunting sample using spike-in normalization which is 93. Then multiplied the ratio of the central signal with TDP2 + ExoT vs ExoT only (~2%). The number of unresected SPO11cc is ~ 2 and is 0.6% of total break number. Thus, the fraction of unresected SPO11cc could be 0.6% ~ 1.7% and SPO11-RI is 9.3% ~ 10.4%.

Mouse hybrid END-seq analysis. Single-nucleotide polymorphisms (SNP) were annotated for CAST genome by the Sanger Mouse Genome Project[54] based on B6 mouse reference genome (mm10). Only SNPs ($n = 20{,}668{,}274$) passed all filter criteria were employed to generate an "N-masked" genome. In all, 14,951 hotspots in B6xCAST hybrid genome were download from GSE75419. As Bowtie does not allow gapped alignment for Ns, the hybrid END-seq reads were aligned by Bowtie2 version 2.3.5.1[55] with options -N 1 -k 1.

SPO11 oligos, SSDS, and ChIP-seq and RNA-seq data analysis. For SPO11-oligo seq, reads were aligned to the mouse (GRCm38p2/mm10) genome using Bowtie version 1.1.2 with options -n 2 -m 1 -l 50. Peaks were called using MACS 1.4.3. Peaks were called using the parameters: –shiftsize = 73,–nomodel and then filtered by fold-enrichment (FC ≥ 10) as well as the blacklists. SPO11 summits from SPO11-oligo seq overlapped the B6 END-seq peaks (~ 5000) were used for most of the aggregate plots except the B6xCAST hybrid data. For DMC1 SSDS, type1 alignments and hotspots' coordinates of SSDS data were directly download from GEO database. B6xCAST hybrid SSDS was analyzed as described in END-seq part. For our H3K4me3 ChIP-seq and the public H3K36me3 ChIP-seq, reads were aligned to the mouse (GRCm38p2/mm10) genome using Bowtie version 1.1.2 with options -n 2 -m 1 -l 50. For RNA-seq, reads were aligned to the mouse (GRCm38p2/mm10) genome using TopHat version 2.1.1 (PMID: 19289445) with default parameters. The gene expression level of RPKM was determined by Cuffnorm version 2.2.1[56] based on the annotation from RefSeq. Mean values across samples were used.

Data visualization. Aligned-reads bed files were first converted to bedgraph files using bedtools genomecov (PMID: 20110278) following by bedGraphToBigWig to make a bigwig file[57]. Visualization of genomic profiles was done by the UCSC browser[58]. Genome browser profiles were normalized to the library size (RPM). Heatmaps were produced using the R package 'pheatmap'.

Aggregate plots for sequencing data around hotspot centers or SPO11 summits were performed as follows: a fixed window was defined around all sites genome wide. The number of reads overlapping each nucleotide was calculated. The aggregate signal was smoothed using smooth.spline function in R. For END-seq, only the adapter ligated endpoints were used.

Statistical analysis. Statistical analysis was performed using R version 3.5.0 (http://www.r-project.org). The statistical tests are reported in the figure legend and main text.

Ethical compliance. All mouse breeding and experimentation followed protocols approved by the National Institutes of Health Institutional Animal Care and Use Committee.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data that support the findings of this study (END-seq and ChIP-seq) are available in the GEO under accession GSE138915. RNA-seq data were downloaded from GEO accession GSE61613[59]. H3K36me3 ChIP-seq was downloaded from GEO accession GSE76416[60]. Prdm9$^{-/-}$ DMC1 SSDS were downloaded from GEO accession GSE99921[24]. DMC1 SSDS in the B6xCAST hybrid, B6 and CAST were downloaded from GEO accession GSE75419[20]. SPO11-oligo sequencing data were downloaded from GEO accession GSE84689[8]. The source data underlying Supplementary Data 1 (Peaks called from END-seq in WT mice) and Supplementary Data 2 (Peaks called from END-seq in *Atm*$^{-/-}$ mice) are provided as a Supplementary Data files. Any additional data are available on request from the authors.

Code availability

R code used for analysis is available at: https://github.com/weiwsmiling/meiosis_END-seq

Received: 24 December 2019; Accepted: 23 January 2020;
Published online: 12 February 2020

References

1. Keeney, S., Giroux, C. N., Kleckner, N. & Meiosis-specific, D. N. A. double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. *Cell* **88**, 375–384 (1997).
2. Garcia, V., Phelps, S. E., Gray, S. & Neale, M. J. Bidirectional resection of DNA double-strand breaks by Mre11 and Exo1. *Nature* **479**, 241–244 (2011).
3. Zakharyevich, K. et al. Temporally and biochemically distinct activities of Exo1 during meiosis: double-strand break resection and resolution of double Holliday junctions. *Mol. Cell* **40**, 1001–1015 (2010).
4. Bishop, D. K., Park, D., Xu, L. & Kleckner, N. DMC1: a meiosis-specific yeast homolog of E. coli recA required for recombination, synaptonemal complex formation, and cell cycle progression. *Cell* **69**, 439–456 (1992).
5. Joshi, N., Brown, M. S., Bishop, D. K. & Borner, G. V. Gradual implementation of the meiotic recombination program via checkpoint pathways controlled by global DSB levels. *Mol. Cell* **57**, 797–811 (2015).
6. Mimitou, E. P., Yamada, S. & Keeney, S. A global view of meiotic double-strand break end resection. *Science* **355**, 40–45 (2017).
7. Lange, J. et al. ATM controls meiotic double-strand-break formation. *Nature* **479**, 237–240 (2011).
8. Lange, J. et al. The landscape of mouse meiotic double-strand break formation, processing, and repair. *Cell* **167**, 695–708 e616 (2016).
9. Grey, C., Baudat, F. & de Massy, B. PRDM9, a driver of the genetic map. *PLoS Genet.* **14**, e1007479 (2018).
10. Baker, C. L., Walker, M., Kajita, S., Petkov, P. M. & Paigen, K. PRDM9 binding organizes hotspot nucleosomes and limits Holliday junction migration. *Genome Res.* **24**, 724–732 (2014).
11. Brick, K., Smagulova, F., Khil, P., Camerini-Otero, R. D. & Petukhova, G. V. Genetic recombination is directed away from functional genomic elements in mice. *Nature* **485**, 642–645 (2012).
12. Paigen, K. & Petkov, P. M. PRDM9 and its role in genetic recombination. *Trends Genet.* **34**, 291–300 (2018).
13. Davies, B. et al. Re-engineering the zinc fingers of PRDM9 reverses hybrid sterility in mice. *Nature* **530**, 171–176 (2016).
14. Hinch, A. G. et al. Factors influencing meiotic recombination revealed by whole-genome sequencing of single sperm. *Science* **363**, eaau8861 (2019).
15. Li, R. et al. A high-resolution map of non-crossover events reveals impacts of genetic diversity on mammalian meiotic recombination. *Nat. Commun.* **10**, 3900 (2019).
16. Canela, A. et al. Topoisomerase II-induced chromosome breakage and translocation is determined by chromosome architecture and transcriptional activity. *Mol. Cell* **75**, 252–266.e8 (2019).
17. Canela, A. et al. Genome organization drives chromosome fragility. *Cell* **170**, 507–521 e518 (2017).
18. Canela, A. et al. DNA breaks and end resection measured genome-wide by end sequencing. *Mol. Cell* **63**, 898–911 (2016).
19. Neale, M. J., Pan, J. & Keeney, S. Endonucleolytic processing of covalent protein-linked DNA double-strand breaks. *Nature* **436**, 1053–1057 (2005).
20. Smagulova, F., Brick, K., Pu, Y., Camerini-Otero, R. D. & Petukhova, G. V. The evolutionary turnover of recombination hot spots contributes to speciation in mice. *Genes Dev.* **30**, 266–280 (2016).
21. Cole, F. et al. Mouse tetrad analysis provides insights into recombination mechanisms and hotspot evolutionary dynamics. *Nat. Genet* **46**, 1072–1080 (2014).
22. Landt, S. G. et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* **22**, 1813–1831 (2012).
23. Nakato, R. & Shirahige, K. Recent advances in ChIP-seq analysis: from quality management to whole-genome annotation. *Brief. Bioinform.* **18**, 279–290 (2017).
24. Brick, K. et al. Extensive sex differences at the initiation of genetic recombination. *Nature* **561**, 338–342 (2018).
25. Brown, M. S., Grubb, J., Zhang, A., Rust, M. J. & Bishop, D. K. Small Rad51 and Dmc1 complexes often co-occupy both ends of a meiotic DNA double strand break. *PLoS Genet.* **11**, e1005653 (2015).

26. Khil, P. P., Smagulova, F., Brick, K. M., Camerini-Otero, R. D. & Petukhova, G. V. Sensitive mapping of recombination hotspots using sequencing-based detection of ssDNA. *Genome Res.* **22**, 957–965 (2012).
27. Gravel, S., Chapman, J. R., Magill, C. & Jackson, S. P. DNA helicases Sgs1 and BLM promote DNA double-strand break resection. *Genes Dev.* **22**, 2767–2772 (2008).
28. Mimitou, E. P. & Symington, L. S. Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing. *Nature* **455**, 770–774 (2008).
29. Zhu, Z., Chung, W. H., Shim, E. Y., Lee, S. E. & Ira, G. Sgs1 helicase and two nucleases Dna2 and Exo1 resect DNA double-strand break ends. *Cell* **134**, 981–994 (2008).
30. Bunting, S. F. et al. 53BP1 inhibits homologous recombination in Brca1-deficient cells by blocking resection of DNA breaks. *Cell* **141**, 243–254 (2010).
31. Chapman, J. R., Sossick, A. J., Boulton, S. J. & Jackson, S. P. BRCA1-associated exclusion of 53BP1 from DNA damage sites underlies temporal control of DNA repair. *J. Cell Sci.* **125**, 3529–3534 (2012).
32. Chen, C. C., Feng, W., Lim, P. X., Kass, E. M. & Jasin, M. Homology-directed repair and the role of BRCA1, BRCA2, and related proteins in genome integrity and cancer. *Annu Rev. Cancer Biol.* **2**, 313–336 (2018).
33. Callen, E. et al. 53BP1 enforces distinct pre- and post-resection blocks on homologous recombination. *Mol. Cell* **77**, 26–38 e27 (2020).
34. Xu, X. et al. Genetic interactions between tumor suppressors Brca1 and p53 in apoptosis, cell cycle and tumorigenesis. *Nat. Genet.* **28**, 266–271 (2001).
35. Xu, X., Aprelikova, O., Moens, P., Deng, C. X. & Furth, P. A. Impaired meiotic DNA-damage repair and lack of crossing-over during spermatogenesis in BRCA1 full-length isoform deficient mice. *Development* **130**, 2001–2012 (2003).
36. Petukhova, G. V., Romanienko, P. J. & Camerini-Otero, R. D. The Hop2 protein has a direct role in promoting interhomolog interactions during mouse meiosis. *Dev. Cell* **5**, 927–936 (2003).
37. Widger, A. et al. ATR is a multifunctional regulator of male mouse meiosis. *Nat. Commun.* **9**, 2621 (2018).
38. Yamada, S. et al. Molecular structures and mechanisms of DNA break processing in mouse meiosis. Preprint at: https://www.biorxiv.org/content/10.1101/2019.12.17.876854v1 (2020).
39. Myler, L. R. et al. Purification and biophysical characterization of the Mre11-Rad50-Nbs1 complex. *Methods Mol. Biol.* **2004**, 269–287 (2019).
40. Anand, R., Ranjha, L., Cannavo, E. & Cejka, P. Phosphorylated CtIP functions as a co-factor of the MRE11-RAD50-NBS1 endonuclease in DNA end resection. *Mol. Cell* **64**, 940–950 (2016).
41. Deshpande, R. A., Lee, J. H., Arora, S. & Paull, T. T. Nbs1 converts the human Mre11/Rad50 nuclease complex into an endo/exonuclease machine specific for protein-DNA adducts. *Mol. Cell* **64**, 593–606 (2016).
42. Cortes Ledesma, F., El Khamisy, S. F., Zuma, M. C., Osborn, K. & Caldecott, K. W. A human 5'-tyrosyl DNA phosphodiesterase that repairs topoisomerase-mediated DNA damage. *Nature* **461**, 674–678 (2009).
43. Grey, C. et al. In vivo binding of PRDM9 reveals interactions with noncanonical genomic sites. *Genome Res.* **27**, 580–590 (2017).
44. Garcia, V., Gray, S., Allison, R. M., Cooper, T. J. & Neale, M. J. Tel1(ATM)-mediated interference suppresses clustered meiotic double-strand-break formation. *Nature* **520**, 114–118 (2015).
45. Orthwein, A. et al. Mitosis inhibits DNA double-strand break repair to guard against telomere fusions. *Science* **344**, 189–193 (2014).
46. Ahmed, E. A. et al. Differences in DNA double strand breaks repair in male germ cell types: lessons learned from a differential expression of Mdc1 and 53BP1. *DNA Repair* **6**, 1243–1254 (2007).
47. Kowalczykowski, S. C. An overview of the molecular mechanisms of recombinational DNA repair. *Cold Spring Harb. Perspect. Biol.* **7**, a016410 (2015).
48. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
49. Li, H. et al. The sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
50. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).
51. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
52. Salmon-Divon, M., Dvinge, H., Tammoja, K. & Bertone, P. PeakAnalyzer: genome-wide annotation of chromatin binding and modification loci. *BMC Bioinformatics* **11**, 415 (2010).
53. Kharchenko, P. V., Tolstorukov, M. Y. & Park, P. J. Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat. Biotechnol.* **26**, 1351–1359 (2008).
54. Keane, T. M. et al. Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* **477**, 289–294 (2011).
55. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
56. Trapnell, C. et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
57. Kent, W. J., Zweig, A. S., Barber, G., Hinrichs, A. S. & Karolchik, D. BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* **26**, 2204–2207 (2010).
58. Kent, W. J. et al. The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
59. Walker, M. et al. Affinity-seq detects genome-wide PRDM9 binding sites and reveals the impact of prior chromatin modifications on mammalian recombination hotspot usage. *Epigenetics Chromatin* **8**, 31 (2015).
60. Powers, N. R. et al. The meiotic recombination activator PRDM9 trimethylates both H3K36 and H3K4 at recombination hotspots in vivo. *PLoS Genet.* **12**, e1006146 (2016).

Acknowledgements

We thank Todd Macfarlan, Mohamed Mahgoub, Michael Lichten, and Kevin Brick for comments on the manuscript; Francesca Cole, Todd Macfarlan, Mohamed Mahgoub, Scott Keeney, Agnieszka Lukaszewicz, Maria Jasin, Anjali Hinch, and Petko Petkov for sharing unpublished data and insightful suggestions; Petko Petkov for PRDM9 KO mice and Scott Keeney for DMC1 knockout mice; R. Daniel Camerini-Otero and Florencia Pratto for HOP2 KO mice; Keith W. Caldecott for purified human TDP2; and Jennifer Wise and Kelly Smith for assistance with animal work. The A.N. laboratory is supported by the Intramural Research Program of the NIH, an Ellison Medical Foundation Senior Scholar in Aging Award (AG-SS- 2633-11), the Department of Defense Idea Expansion (W81XWH-15-2-006) and Breakthrough (W81XWH-16-1-599) Awards, the Alex Lemonade Stand Foundation Award, and an NIH Intramural FLEX Award.

Author contributions

J.P., W.W. and A.N. designed experiments; J.P., N.S., A.C., E.C. and A.D. performed experiments; R.A.D. provided purified proteins; S.Y. provided critical insight and interpretations; W.W., J.P., Y.M. and A.N. analyzed and interpreted the data; T.P. and A.N. supervised the research and provided advice. J.P., W.W. and A.N. wrote the manuscript with comments from the authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41467-020-14654-w.

Correspondence and requests for materials should be addressed to A.N.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

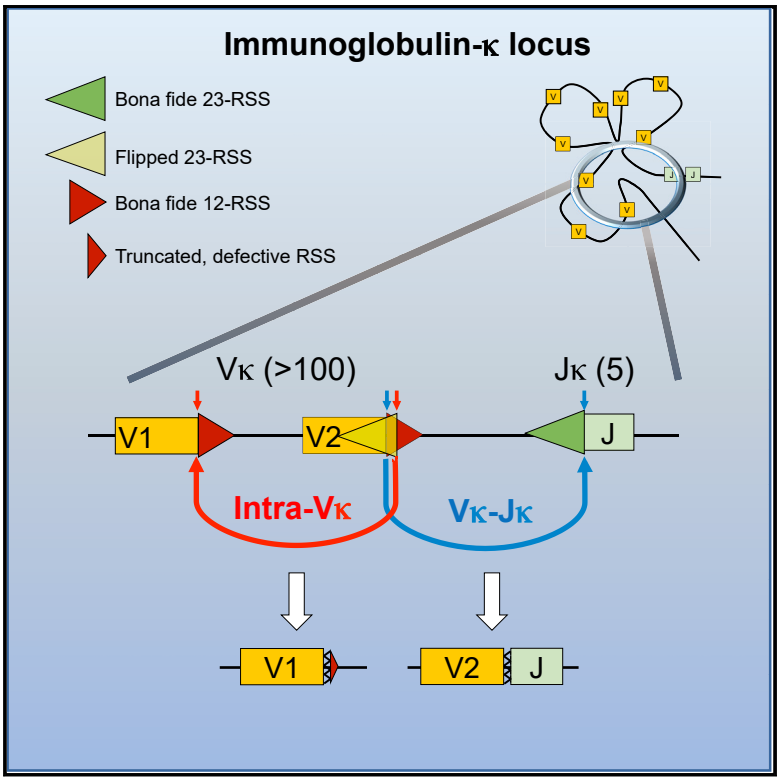
This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2020

Article

Cell Reports

Intra-Vκ Cluster Recombination Shapes the Ig Kappa Locus Repertoire

Graphical Abstract



Authors

Kenta Shinoda, Yaakov Maman, Andres Canela, David G. Schatz, Ferenc Livak, André Nussenzweig

Correspondence

ferenc.livak@nih.gov (F.L.), andre_nussenzweig@nih.gov (A.N.)

In Brief

Shinoda et al. demonstrate previously unknown, frequent RAG-dependent recombination within the Ig Vκ gene cluster, which is mediated by evolutionarily conserved flipped RSS embedded in bona fide RSS. These intra-Vκ cluster recombination events alter the available Vκ gene pool by deleting Vκ segments and shape the Vκ-Jκ immune repertoire.

Highlights

- Vκ gene segments rearrange within the Vκ cluster independent of Jκ gene segments
- Intra-Vκ recombination is mediated by flipped RSS embedded in bona fide RSS
- An independent RAG recombination center assembles in the Vκ cluster
- Intra-Vκ cluster recombination influences ultimate Vκ gene segment usage

Shinoda et al., 2019, Cell Reports *29*, 4471–4481
December 24, 2019 © 2019 The Author(s).
https://doi.org/10.1016/j.celrep.2019.11.088

CellPress

Cell Reports
Article

OPEN ACCESS
CellPress

Intra-Vκ Cluster Recombination Shapes the Ig Kappa Locus Repertoire

Kenta Shinoda,[1,5] Yaakov Maman,[1,2,5] Andres Canela,[1,3] David G. Schatz,[4] Ferenc Livak,[1,*] and André Nussenzweig[1,6,*]
[1]Laboratory of Genome Integrity, National Cancer Institute, NIH, Bethesda, MD, USA
[2]The Azrieli Faculty of Medicine, Bar-Ilan University, Safed, Israel
[3]The Hakubi Center for Advanced Research and Radiation Biology Center, Graduate School of Biostudies, Kyoto University, Kyoto, Japan
[4]Department of Immunobiology, Yale School of Medicine, New Haven, CT, USA
[5]These authors contributed equally
[6]Lead Contact
*Correspondence: ferenc.livak@nih.gov (F.L.), andre_nussenzweig@nih.gov (A.N.)
https://doi.org/10.1016/j.celrep.2019.11.088

SUMMARY

During V(D)J recombination, RAG proteins introduce DNA double-strand breaks (DSBs) at recombination signal sequences (RSSs) that contain either 12- or 23-nt spacer regions. Coordinated 12/23 cleavage predicts that DSBs at variable (V) gene segments should equal the level of breakage at joining (J) segments. Contrary to this, here we report abundant RAG-dependent DSBs at multiple Vκ gene segments independent of V-J rearrangement. We find that a large fraction of Vκ gene segments are flanked not only by a bone-fide 12 spacer but also an overlapping, 23-spacer flipped RSS. These compatible pairs of RSSs mediate recombination and deletion inside the Vκ cluster even in the complete absence of Jκ gene segments and support a V(D)J recombination center (RC) independent of the conventional Jκ-centered RC. We propose an improved model of Vκ-Jκ repertoire formation by incorporating these surprisingly frequent, evolutionarily conserved intra-Vκ cluster recombination events.

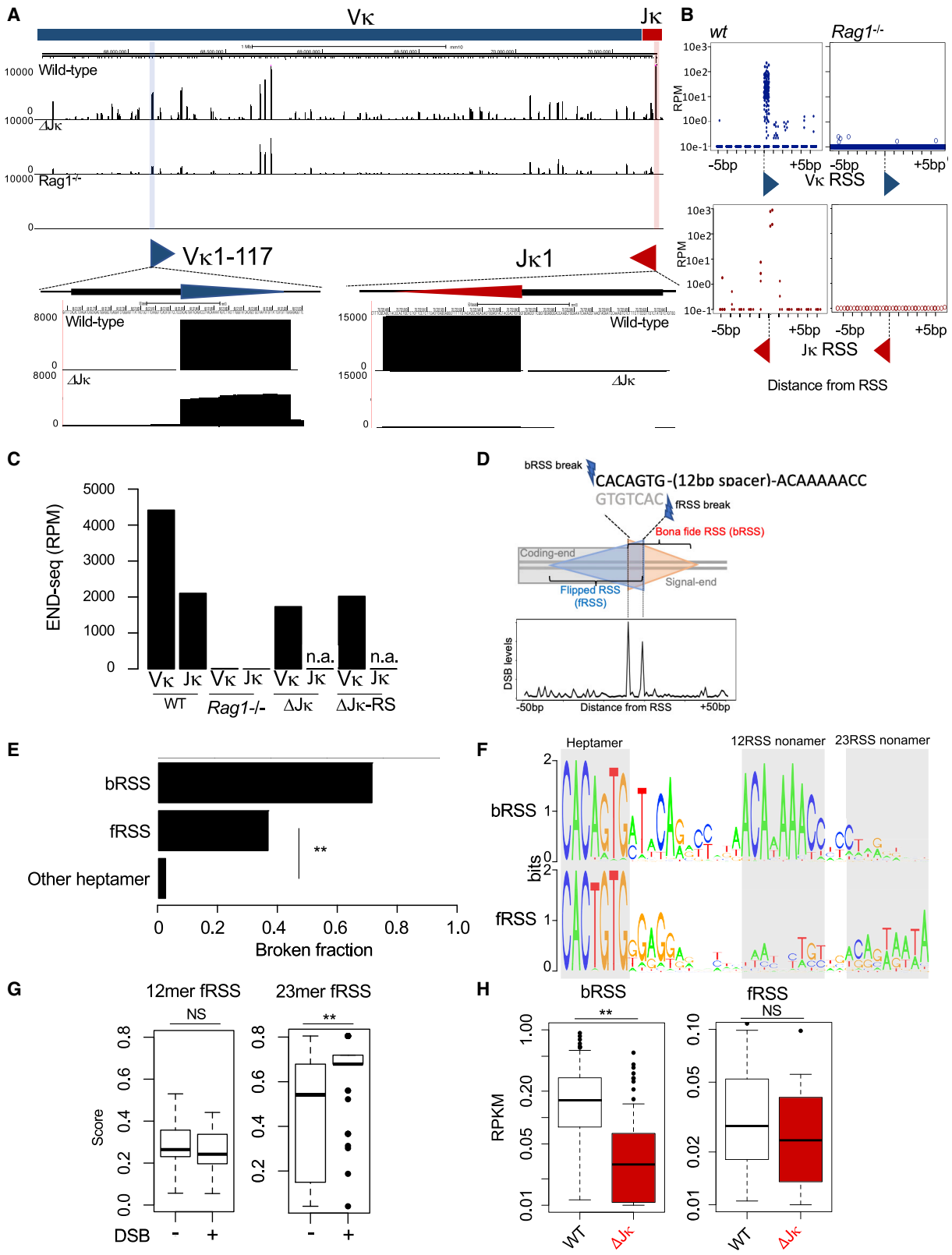
INTRODUCTION

The immensely diverse repertoire of B and T cell antigen receptors (AgRs) are encoded in discrete variable (V), diversity (D), and joining (J) gene segments that undergo somatic recombination in lymphocyte precursors. V(D)J recombination is initiated by the lymphocyte-specific recombination activating gene (RAG-1 and −2) proteins that generate DNA double-strand breaks (DSBs) at recombination signal sequences (RSSs) flanking the V, D, and J gene segments (Teng and Schatz, 2015). The RSS is composed of a conserved heptamer and nonamer motif, separated by either a 12-bp (12RSS) or a 23-bp (23RSS) spacer. In every AgR locus (Sakano et al., 1979), efficient recombination occurs only between different types of gene segments flanked by two different types of RSS known as the 12/23 rule (Tonegawa, 1983). This rule is enforced by the RAG proteins that, in physiological conditions, introduce DSBs only in the context of a synaptic complex composed of a pair of 12/23 RSS (van Gent et al., 1996). After cleavage, the broken signal ends (SEs) and coding ends (CEs) are joined by the non-homologous end joining (NHEJ) pathway (Helmink and Sleckman, 2012). Recombination can occur by deletion of the intervening sequences or by intrachromosomal inversion. During deletional rearrangement, joining of the two RSSs produce extrachromosomal excision circles (Shimizu and Yamagishi, 1992) that persist only in non-dividing precursors and are subsequently lost in mature lymphocytes (Livak and Schatz, 1996).

The mouse immunoglobulin κ (Igκ) locus contains more than 140 V gene segments and 4 functional J gene segments. In contrast to other AgR loci, there are many Vκ gene segments in either forward or reverse transcriptional orientation relative to the Jκ segments (Proudhon et al., 2015); thus, Vκ-to-Jκ joining can occur either deletionally or by inversion. In addition, the Igκ locus undergoes multiple rounds of rearrangements to generate a minimally self-reactive, tolerant Ig repertoire (Nemazee, 2017). Self-tolerance is also achieved through complete deletion of the Cκ gene segment mediated by the human Igκ deleting element (IGKDE) (Siminovitch et al., 1987) or its murine equivalent sequences (recombining sequence or RS) (Durdik et al., 1984; Feddersen et al., 1990). Precursors with such unconventional V(D)J recombination events eventually develop into Igλ-expressing mature B cells. Ultimately, the final VκJκ repertoire is remarkably similar in bone marrow pre-B cells and mature splenic B cells (Aoki-Ota et al., 2012), indicating that, like in the IgH and T cell receptor (TCR)-β loci (Livak et al., 2000; Yu et al., 2002), the frequency of Vκ gene segment usage is predetermined largely by the gene rearrangement process rather than by cellular selection (Rubelt et al., 2016).

A major conundrum of V(D)J recombination is to understand how the simple, transposon-derived RAG recombinase controls the astonishingly finely tuned shape of the Ig and TCR repertoires. V(D)J recombination is initiated by the recruitment of RAG to trimethylated lysine 4 residue of histone 3 (H3K4me3), in a region termed recombination center (RC), located at the J loci of the AgRs (Desiderio, 2010; Ji et al., 2010; Matthews and Oettinger, 2009). The recently introduced RAG-scanning model proposes that within convergent CTCF-binding-element (CBE)-based chromatin loop domains, RAG scans directionally from an initiating RC for megabase (Mb) distances for compatible RSS pairs (Hu et al., 2015; Jain et al., 2018). How the RAG-scanning model could explain the widespread, balanced use of more

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).



(legend on next page)

than a 100 Vκ gene segments, some of them Mbs away from the Jκ-associated RC (Teng et al., 2015), is not fully understood.

Here, we report that a large fraction of RAG-mediated cleavage at Vκ gene segments occurs independent of the Jκ gene segments both in mouse and human and identify a second, autonomous RC within the Vκ cluster. This RC uses the bona fide Vκ 12RSS (hereafter referred to as bRSS) with overlapping, cryptic 23RSS and directs recombination between Vκ gene segments in competition with the Jκ RC. We provide a predictive model of Vκ gene segment use that incorporates intra-cluster Vκ rearrangements. This study highlights an evolutionarily conserved strategy that, uniquely among the AgR loci, shapes the Igκ repertoire in diverse species.

RESULTS

Extensive RAG-Mediated DSBs at Vκ Gene Segments Independent of Jκ

To evaluate the repertoire of RAG-mediated breakage in the Igκ locus, we applied END-seq, a method that maps DSBs at single-base-pair resolution (Canela et al., 2016) in Abelson virus (v-Abl)-transformed mouse pre-B cells. Here, RAG expression and Igκ locus recombination are induced following treatment with the v-Abl kinase inhibitor STI-571 (Muljo and Schlissel, 2003). DSBs were found reproducibly at 80% of annotated Vκ gene segments and at all four functional Jκ gene segments (Jκ1, Jκ2, Jκ4, and Jκ5) (Figures 1A and 1B). *Rag1*$^{-/-}$ pre-B cells did not show any breakage around the RSS (Figures 1A and 1B), confirming that these breaks are RAG mediated. END-seq analysis performed on *ex-vivo*-cultured primary mouse bone marrow pre-B cells showed a similar pattern of DSBs to v-Abl-transformed pre-B cells (Figure S1A).

Coordinated RAG cleavage following the 12/23 rule of V(D)J recombination dictates that the total number of DSBs at Vκ 12RSS and Jκ 23RSS would be nearly equal. Surprisingly, the total number of DSBs detected at Vκ gene segments was more than 2-fold higher than that of Jκ gene segments (Figure 1C; for average of 7 replicates see Figure S1B).

To test whether all Vκ breaks are the result of attempted V-to-J rearrangement, we applied END-seq on a pre-B cell clone where all five Jκ gene segments were deleted (termed ΔJκ; see Figures S1C and S1D). Strikingly, Vκ breaks continued to be detected at substantial levels in the ΔJκ clone (Figures 1A and 1C). We considered the possibility that the previously described Cκ-deleting 23RSS elements (Durdik et al., 1984; Feddersen et al., 1990; Siminovitch et al., 1987) may contribute to the excess of Vκ breaks. However, the removal of the Jκ-Cκ intronic deleting elements (RS), in addition to Jκ (termed ΔJκ-RS; see Figures S1C and S1D) did not eliminate RAG-dependent Vκ breaks (Figure 1C). The above findings raised two alternatives: either DSBs at Vκ 12RSS are occurring in violation of the 12/23 rule or there are other, unidentified 23RSS in the Igκ locus.

Identification of a 23-mer RSS within the Vκ Locus

Taking advantage of the single-nucleotide resolution capacity of END-seq, we noticed a significant enrichment of breaks not only at the first nucleotide of the heptamer, where RAG normally introduces a precise DSB, but also at the last nucleotide of the heptamer of the bRSS (Figure 1D). Due to the inherent near-palindromic nature of the consensus heptamer sequence (5′-CACAGTG), this motif in the opposite direction on the other strand can be read as a flipped heptamer (5′-CACTGTG) and could be part of a flipped RSS (hereafter referred to as fRSS; see Figure 1D).

Although most bRSS and about one-third of fRSS showed DSBs precisely at the first position of their heptamers, virtually no other heptamer-like sequences were broken in the Vκ cluster (Figure 1E). We searched the DNA sequences following all heptamers for nonamer-like motifs by using a position weight matrix (PWM) approach (see STAR Methods). The heptamers of bRSS are followed by a strong nonamer at exactly a 12-bp distance, as expected for a well-defined 12RSS (Figure 1F). The flipped heptamers are also frequently followed by a nonamer-like motif but at a 23-bp distance (Figure 1F). We compared the sequences of fRSS with respect to their DSB status and found that broken fRSS had a higher 23RSS PWM score (see STAR Methods) than non-broken fRSS, whereas both cleaved and non-cleaved fRSS had equally low 12RSS PWM scores (Figure 1G; Figure S1E). Finally, although the number of bRSS breaks showed a significant reduction in the ΔJκ clone, the number of fRSS breaks remained similar between wild-type (WT) and ΔJκ cells

Figure 1. RAG-Dependent DSBs in the Vκ Cluster at Flipped RSS (fRSS) that Overlap with Bona Fide RSS (bRSS)

(A) Top panel: END-seq reads at the Igκ locus from WT (top track), ΔJκ (middle track), and *Rag1*$^{-/-}$ (bottom track) v-Abl-transformed pre-B cells. Positions of the Vκ1-117 (blue) and Jκ1 (red) genes are marked on the top. Bottom panel: magnification of the Vκ1-117 and Jκ1 gene segments. Vκ1-117 and Jκ1 breakages are highlighted in blue and red, respectively; blue and red triangles depict the corresponding 12RSS and 23RSS, respectively.

(B) Nucleotide resolution cumulative mapping of DSB sites by END-seq within ±5 bp surrounding Vκ RSS (top) or Jκ RSS (bottom) in WT (left) or *Rag1*$^{-/-}$ (right) v-Abl-transformed pre-B cells.

(C) Quantification of END-seq reads within the Vκ and Jκ gene clusters (represented as reads per milion [RPM]) in WT, ΔJκ, and ΔJκ-RS v-Abl-transformed pre-B cells. n.a., not applicable.

(D) Top panel: schematic diagram of bRSS and the partially overlapping fRSS. Double-strand breaks for bRSS occur at the 5′ end of the heptamer and for fRSS at the 3′ end of the heptamer. Orange triangle represents bRSS, and blue triangle represents fRSS. Bottom panel: aggregate plot for END-seq signal at Vκ RSS surrounding region (±50 bp).

(E) Fraction of broken bRSS, fRSS, and other heptamers within the Vκ cluster in v-Abl-transformed pre-B cells, as detected by END-seq. *$p < 1e^{-10}$, Fisher's test.

(F) Consensus sequence logo of Vκ bRSS (top) and fRSS (bottom). Heptamer and nonamer motifs at 12-bp and 23-bp spacer distances are highlighted in gray.

(G) Boxplots of average PWM score of 12-bp spacer RSS (left) or 23-bp spacer RSS (right) sequences of fRSS with (+) or without (–) DSBs, as detected by END-seq in WT v-Abl-transfomed pre-B cells. NS, p = 0.82; **$p < 1e^{-7}$; t test.

(H) Boxplot quantification of END-seq reads present at Vκ bRSS (left) and fRSS (right) DSBs in WT (black) and ΔJκ (red) v-Abl-transfomed pre-B cells. See also Figure S1. NS, p = 0.11; **$p < 1e^{-14}$; paired t test.

(Figure 1H; Figure S1F), indicating that fRSS are used independently of V-to-Jκ rearrangement.

Collectively, these results identify a number of 23-mer RSS in the mouse Vκ locus, which overlap with the heptamer of the Vκ bRSS in a flipped orientation. A substantial fraction of these fRSSs, specficially those with higher quality 23RSS, is cleaved by RAG. We hypothesize that these fRSS embedded within the Vκ cluster could explain many of the widespread, Jκ-independent DSBs described above.

Extensive Intra-Vκ Cluster Rearrangements Mediated by 12/23 Pairs of bRSS and fRSS

To determine whether these 23-bp spacer fRSSs actually mediate rearrangements within the Igκ locus, we used high-throughput genome-wide translocation sequencing (HTGTS) (Lin et al., 2016). This method captures joining between a pre-defined site, termed the "bait," with its fusion partner that originates elsewhere in the genome. We compared the number of HTGTS joints of V-Jκ genes with the number of intra-Vκ cluster bRSS/fRSS-mediated joining events. Consistent with the END-seq data, HTGTS performed with the Jκ1 CE as a bait revealed a large number of of Jκ1 joining events to Vκ gene segments using the Vκ bRSS (Figures 2A and 2B). In contrast, only one-half to two-thirds of the junctions from the highly used (Aoki-Ota et al., 2012) Vκ1-117 (Figures 2A and 2B) or Vκ8-34 RSSs (Figures S2A and S2B) were V-Jκ rearrangements, whereas a sizeable fraction involved recombination between bRSS and fRSS within the Vκ cluster in pre-B cells (Figures 2A and 2B; Figures S2A and S2B). These intra-Vκ cluster rearrangements were also entirely retained in ΔJκ pre-B cells (Figures 2A and 2B; Figures S2A and S2B). HTGTS performed with the Vκ1-117 CE bait on sorted, primary mouse bone marrow pre-B cells showed a similar pattern to the v-Abl pre-B cell clones (Figure S2C).

The Jκ1 CE bait products included diverse V gene segments across the entire Vκ cluster, whereas the Vκ CE baits captured predominantly nearby Vκ cluster targets (Figure 2A; Figure S2A). When we analyzed Vκ1-117 CE bait products according to the published Vκ locus chromatin loop distribution (Karki et al., 2018) (Figure S2D), we found that more than 80% of Vκ cluster bRSS-fRSS joints resided within the same loop as the Vκ1-117 gene segment (Figure S2E). In contrast, a pre-B cell clone with a CRISPR-Cas9 DSB bait site inserted near the Vκ1-117 gene segment captured prey sequences across the entire Vκ locus. Only about 30% of these junctions resided within the same loop as the Cas9 bait site, which was similar to the calculated random frequency (Figure S2E). These data suggest that intra-Vκ cluster rearrangements occur in accordance with the RAG scanning model of canonical 12/23 V(D)J recombination.

We also examined the precise location of prey sequences to which the Vκ1-117CE bait joined. We found that the majority of bRSS-fRSS-mediated joining occurred between the Vκ1-117 CE sequences and sequences flanking the heptamer of the fRSS of other partner Vκ gene segments (Figure 2C with ΔJk clone and Figure S2F with WT clone). HTGTS performed in DNA Ligase IV- or Artemis-deficient pre-B cells by using the Vκ1-117 CE bait confirmed that both V-Jκ and intra-Vκ cluster joints were NHEJ dependent (Figure S2G).

Strikingly, intra-Vκ cluster rearrangements mediated by bRSS-fRSS interactions were also detected in splenic, IgM$^+$ mature B cells in similar pattern, albeit at somewhat reduced proportions, as in pre-B cells (Figures 2A and 2B). Furthermore, HTGTS analysis of the Igκ locus in purified, human peripheral blood B cells showed a similarly large number of intra-Vκ cluster junctions from the Vκ2-28 gene segment—orthologous to the mouse Vκ1-117—or the Vκ3-20 gene segment (Figures 2D and 2E). Overall, mature B cells exhibited a surprisingly high frequency of intra-Vκ cluster rearrangements using 12/23 bRSS-fRSS pairs both in mice (24% of Vκ1-117 joints) and humans (37% of Vκ2-28 joints), suggesting that the fRSSs are extensively used during Igκ locus rearrangements. These results from peripheral B cells also support the notion that bRSS-fRSS-mediated intra-Vκ cluster joints are primary rearrangements on the chromosome rather than byproducts of secondary rearrangements on extrachromosomal circles, which are mostly lost in mature B cells.

Vκ fRSSs Exploit Conserved Framework Region 3 (FR3) Codons as a 23RSS Nonamer

Evolutionary conservation of non-coding 12 and 23RSS motifs is a hallmark of V(D)J recombination (Tonegawa, 1983). In contrast to bRSS, the fRSS is largely localized within the 3′ end of the Vκ coding sequences that could exert conflicting evolutionary selection on an otherwise regulatory, non-coding element. We found, however, that the sequence composition of Vκ gene segments allows the co-existence of both selection forces. Specifically, the 23-bp spacer encodes for the complementarity-determining region 3 (CDR3) of the Vκ domain, whereas the nonamer motif corresponds to the highly conserved YY[F/H]C anchor residues of the FR3 (Figure 3A). Frequently, the nine nucleotides of the predominant three codons of YYC resemble the canonical nonamer sequence 5′-ACAATAGTA (Figure 3A). Thus, in the Vκ locus, a highly conserved nonamer-like motif of FR3 and the flipped heptamer overlapping with the bRSS, separated exactly by 23 nt of less conserved CDR3, can create a 23-mer fRSS to pair with the Vκ 12-mer bRSS (Figure 3A).

Given the conservation of the YY[F/H]C motif in all Ig and TCR variable domains and the fact that this triplet can be frequently encoded by a nonamer-like motif (Figure 3B), we sought to determine whether the embedded fRSS could be a more general phenomenon extending beyond the Vκ gene segments. Because all other AgR loci have 23-mer bRSS flanking the V genes, we focused our search on 12-mer fRSS. In all AgR loci, except the Igκ and λ light chains, the V domains contain a shorter CDR3, typically 2–5 amino acids in length, following the conserved FR3 and could potentially encode a flipped 12RSS. However, we found only a few V gene segments that appear to contain a high scoring flipped 12RSS (see STAR Methods), primarily due to the frequent deviation of CDR3 sequences from the 12-nt length, which is absolutely necessary for a 12RSS spacer (Figure 3C) .

To test directly whether RAG-mediated DSBs occur at putative fRSSs in another AgR locus, we performed END-seq on CD4/CD8 double-positive (DP) thymocytes that actively undergo TCRα locus recombination and compared Vκ and Vα loci with respect to their bRSS and fRSS sequences and breakage

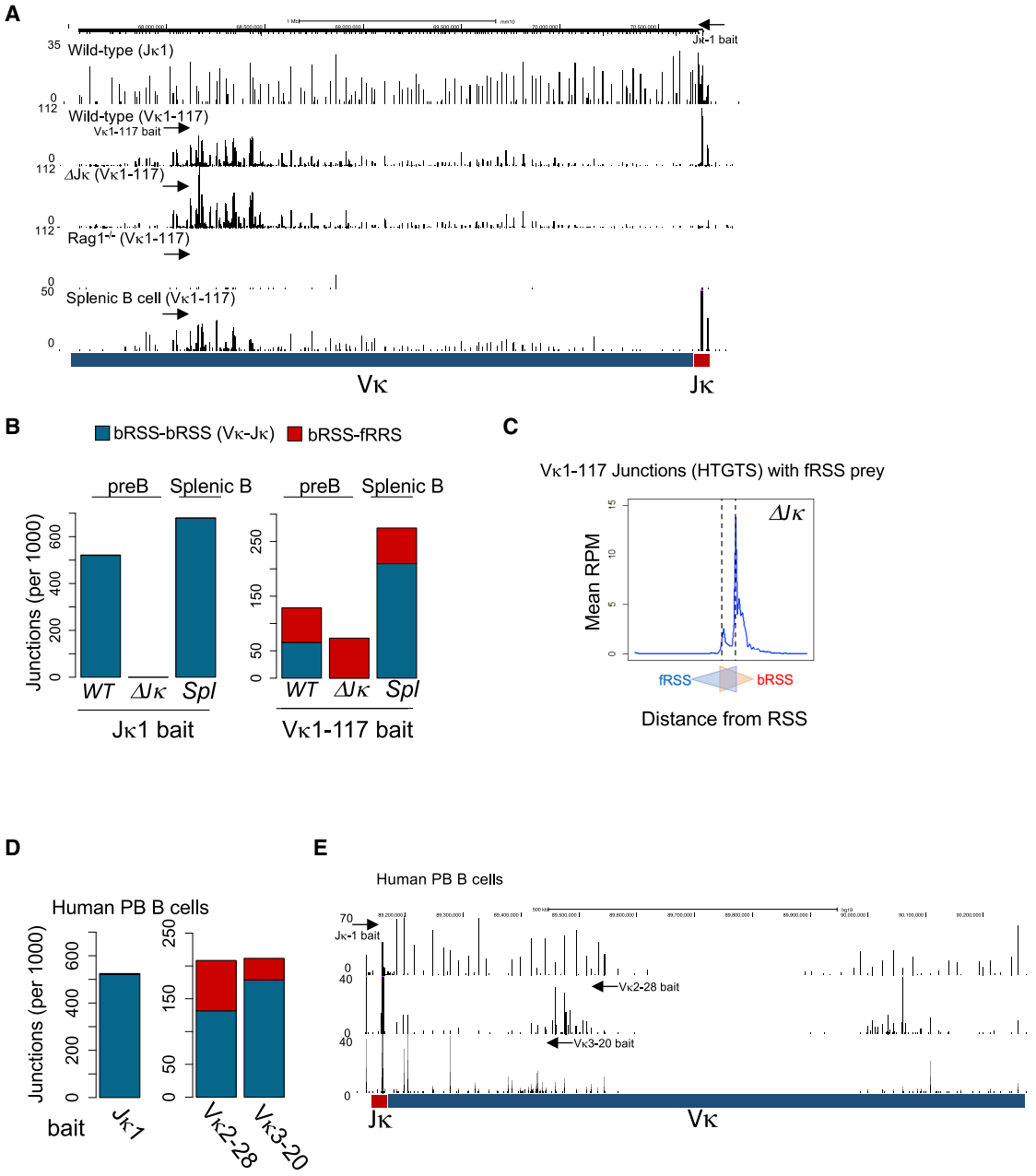


Figure 2. Frequent fRSS-Mediated Intra-Vκ Cluster Rearrangements in Mouse and Human B Cells

(A) HTGTS junction profiles in WT (Jκ1 bait and Vκ1-117 bait), ΔJκ (Vκ1-117 bait), and *Rag1*$^{-/-}$ (Vκ1-117 bait) v-Abl-transformed pre-B cells and in WT splenic B cells within the Igκ locus. Arrows indicate the approximate positions of the Jκ1 and Vκ1-117 bait primers.

(B) The number of junctions (per 1,000 total junctions) of bRSS-bRSS (Vκ-Jκ in blue) and bRSS-fRSS (intra-Vκ in red) recombinations in v-Abl-transformed pre-B cells (WT and ΔJκ) or splenic B cells detected by the Jκ1 (left) or Vκ1-117 (right) bait.

(C) Aggregate Vκ1-117 junctions surrounding the Vκ RSSs in ΔJκ v-Abl-transformed pre-B cells. The first nucleotides of bRSS and fRSS are depicted as two overlapping triangles.

(D) The number of junctions of bRSS-bRSS (Vκ-Jκ in blue) and bRSS-fRSS (intra-Vκ in red) recombinations in human peripheral blood B cells detected by HTGTS with Jκ1 (left), Vκ2-28, and Vκ3-20 (right) baits.

(E) HTGTS junction profiles of human peripheral B cells within the Igκ locus. The arrows indicate the approximate positions of the Jκ1, Vκ2-28, and Vκ3-20 bait primers.

See also Figure S2.

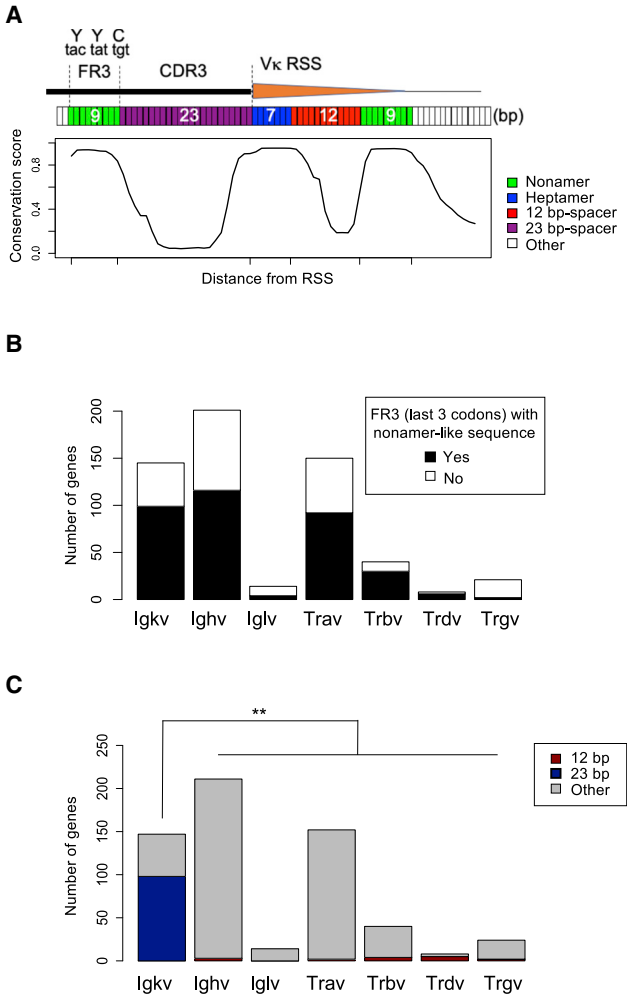


Figure 3. Vκ Gene Segments Harbor Unique, Evolutionarily Conserved fRSS

(A) Conservation score of the region surrounding the Vκ RSS region. Top diagram shows the position of FR3 and CDR3 at the 3′ end of Vκ coding region. Heptamer (blue), spacer (purple and red), and nonamer (green) regions are highlighted.

(B) Number of V gene segments with FR3 with the last 3 codons encoded by nonamer-like sequences in the Igk, Igh, Igl, Tra, Trb, Trd, and Trg loci.

(C) Number of V gene segments with CDR3 with length of 12 bp (red), 23 bp (blue), or other number of nucleotides (gray) in the Igk, Igh, Igl, Tra, Trb, Trd, and Trg loci.

See also Figure S3. **$p < 1e^{-}10$, Fisher's test.

(Figures S3A and S3B). High level bRSS breakage is seen both in pre-B cells at the Vκ gene segments and in thymocytes at the Vα gene segments, as expected (Figures 1A and S3A). Both Vκ and Vα clusters contain a large number of palindromic heptamer motifs (Figure S3C, marked as fRSS); however, only the Vκ fRSSs contain frequent DSBs (Figure S3B), whereas none of the Vα palindromic heptamers showed breakage at the opposite end of the heptamer in thymocytes (Figure 3B). Thus, we conclude that other AgR loci besides Vκ do not have conserved, functional fRSSs with either 12-bp or 23-bp spacers and, at least in the TCRα locus, do not undergo RAG-mediated cleavage in thymocytes at predicted, opposite end positions of palindromic heptamers anywhere within their V clusters.

An Autonomous RC in the Vicinity of fRSSs Initiate Intra-Vκ Cluster Recombination

V(D)J recombination is initiated by the binding of RAG1/2 to active, H3K4me3-marked chromatin at the J region of the AgR loci, termed the nascent RC (Jain et al., 2018; Ji et al., 2010). The chromatin-bound RAG complex associated with the RSS of a J gene segment linearly tracks the chromatin for compatible RSS to form a paired complex with the V or D gene segment RSS (Hu et al., 2015; Lin et al., 2018; Zhao et al., 2016). Given that intra-Vκ cluster rearrangements are independent of the Jκ gene segments and of the natural Igκ locus RC (Figure 1C), we hypothesized that these rearrangements are initiated by autonomous RAG binding within the Vκ locus. To test this prediction, we examined RAG1 chromatin immunoprecipitation sequencing (ChIP-seq) data obtained from RAG1-D708A mice, where a catalytically inactive, mutant RAG1 retains its binding to RSS but is incapable of cleaving DNA (Teng et al., 2015). In these mice, robust RAG binding to both Jκ and Jα clusters from bone marrow pre-B cells and DP thymocytes, respectively, were documented previously (Ji et al., 2010). However, when we compared RAG ChIP-seq signals in the V clusters, Vκ gene segments in bone marrow pre-B cells, but not Vα in DP thymocytes, also showed substantial RAG1 binding (Figures 4A and 4B). Strikingly, the highest level of RAG1 binding in the Vκ cluster is detected at bRSS with overlapping, cleaved fRSSs (Figure 4C). Further ChIP-seq analysis of v-Abl-transformed pre-B cells revealed that RAG1 and RAG2 binding, as well as the levels of H3K4me3 throughout the Vκ locus, are nearly completely retained in the ΔJk clone at the level observed in WT pre-B cells (Figures 4D and 4E).

These results support the hypothesis that RAG1/RAG2 are recruited locally to the Vκ locus to form an autonomous Vκ RC. This Vκ RC can initiate intra-Vκ cluster rearrangement by using the Vκ-associated bRSS and fRSS pairs, independent of the Jκ RC.

Predictive Model of the Igκ Light Chain Repertoire

Numerous studies have shown that the biochemical properties of RAG-RSS interactions significantly shape the developing combinatorial AgR repertoire and that this repertoire is minimally altered by subsequent cellular selection for the TCRβ (Livak et al., 2000; Wu et al., 2003), IgH (Yu et al., 2002), or Igκ loci (Aoki-Ota et al., 2012). Several algorithms have been developed to measure the functionality of the RSS, mostly based on sequence conservation, but these measures, such as the Recombination Information Content (RIC) score (Cowell et al., 2002), predict gene segment use or DSBs with modest success in AgR loci, including the mouse Igκ locus (Figures 5A and 5B). The majority of Vκ gene segments are flanked by high PWM nonamer score bRSS (Figure S4A), whereas RSS cleavage and gene segment use vary significantly across the locus (Figure S4A). Recombination mediated by bRSS-fRSS pairs would destroy the bRSS and render both participating Vκ gene segments non-functional. In addition, deletional bRSS-fRSS recombination would also remove all intervening

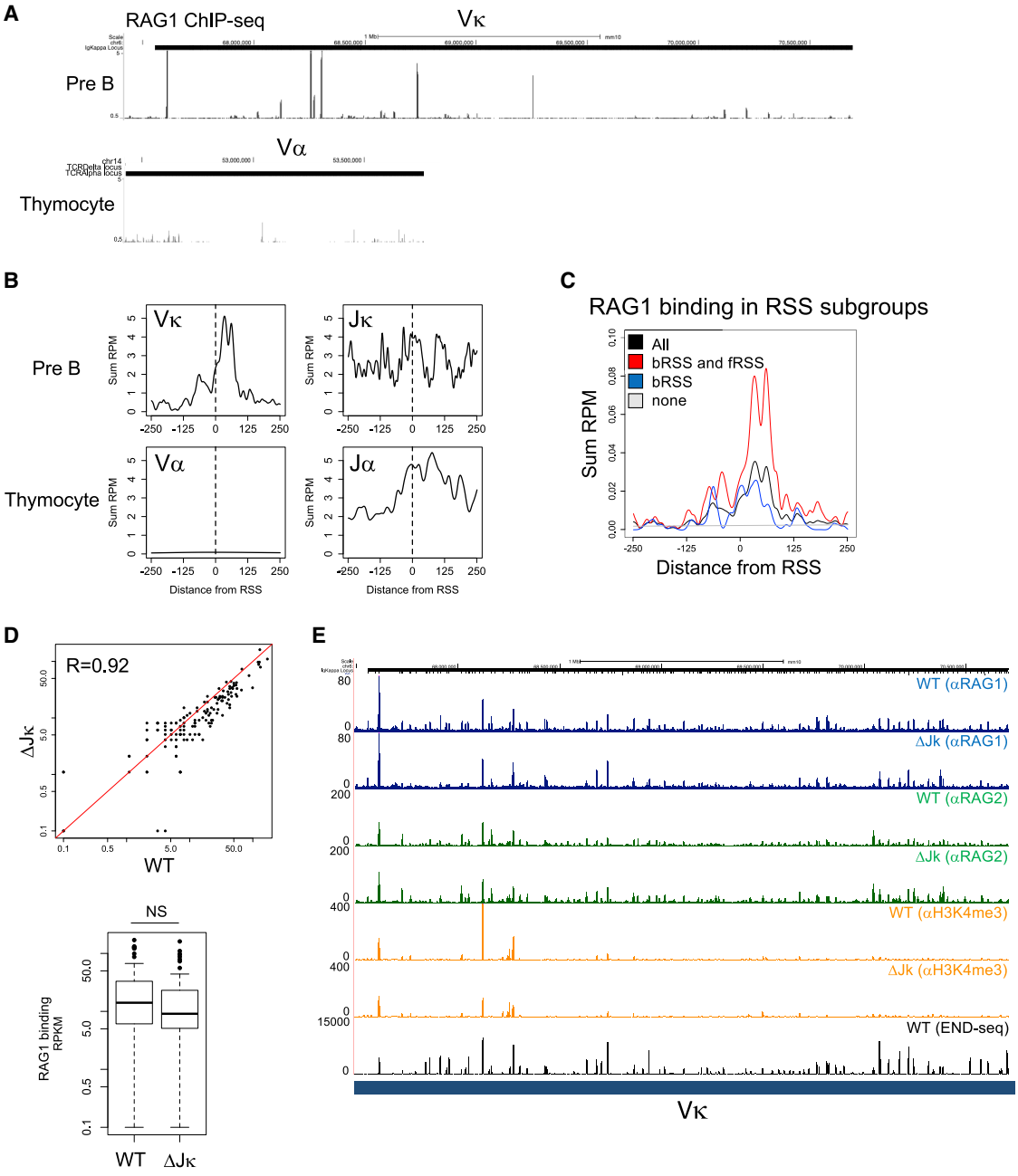


Figure 4. RAG1/2 Complex Forms an Autonomous RC at the Vκ Cluster

(A) Publically accessible ChIP-seq data (RPM mapped reads) for RAG1 binding in the Vκ cluster in primary bone marrow pre-B cells (top) or in the Vα cluster in DP thymocytes (bottom) from RAG1(D708A) mutant mice (Teng et al., 2015).

(B) Aggregate plot of RAG1 ChIP-seq signal around bRSS for Vκ, Jκ, Vα, and Jα loci (±250 bp).

(C) Aggregate plot of RAG1 ChIP-seq signal around Vκ RSS, classified as all RSS (black), cleaved bRSS with overlapping cleaved fRSS (red), cleaved bRSS with overlapping non-cleaved fRSS (blue), or at RSS with no cleavage at all (gray) (±250 bp).

(D) Scatterplot correlation of RAG1 ChIP-seq signal at each Vκ RSS in WT and ΔJκ v-Abl-transformed pre-B cells (top). Boxplot of total RAG1 ChIP-seq signal at Vκ RSS in WT and ΔJκ v-Abl-transformed pre-B cells (bottom).

(E) ChIP-seq (RPM mapped reads) for RAG1 (blue), RAG2 (green), and H3K4me3 (orange) in WT and ΔJκ v-Abl-transformed pre-B cells across the Igκ locus. END-seq tracks (a replicate sample different from the one shown in Figure 1A) are displayed in the bottom.

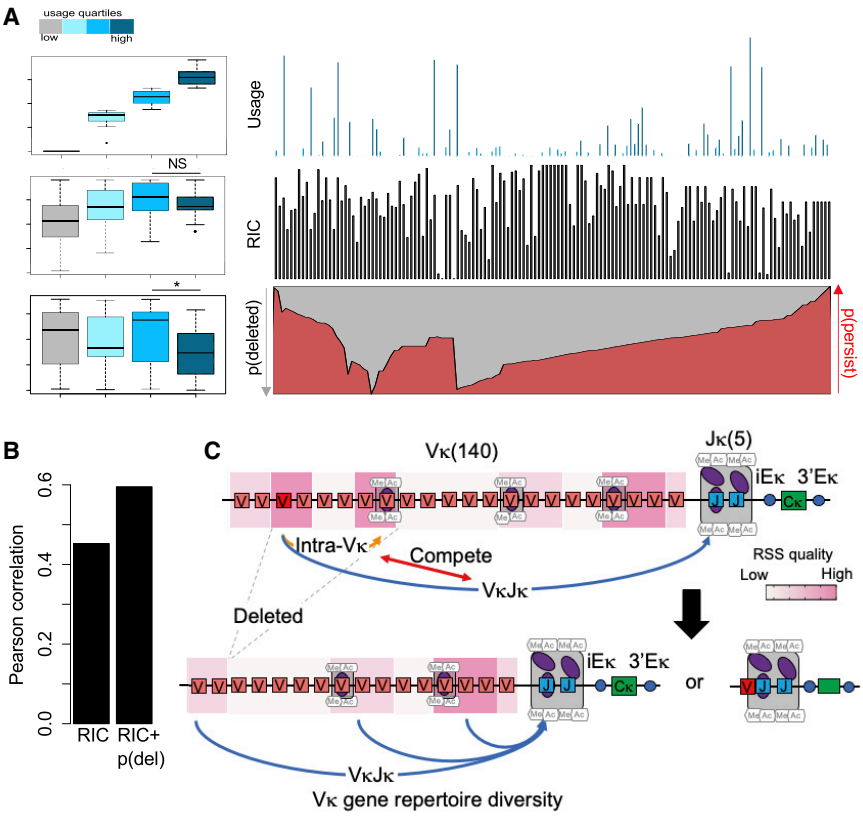


Figure 5. Predictive Model for the Generation of the Igκ Repertoire

(A) The Vκ RSSs are divided into four quartiles based on published use data in splenic B cells (Aoki-Ota et al., 2012). RIC score and probablility of deletion are calculated for each classified RSS. Boxplot summary (left) and distribution (right) of the Vκ gene segment use (top), RIC score (middle), and probability of deletion/persistence (gray/red, respectively) (bottom) are shown (for calculations see STAR Methods). NS, p = 0.16; *p = 0.006; t test.

(B) Pearson correlation coefficient between Vκ use against RIC score and combined RIC score and probability of deletion.

(C) Schematic model displaying competition between intra-Vκ cluster rearrangement and Vκ-Jκ rearrangement. Intra-Vκ cluster rearrangement results in the deletion of various Vκ gene segments, which allows for diverse V-Jκ rearrangements to fine tune the Vκ gene repertoire. See also Figure S4.

Vκ gene segments from the chromosome. We hypothesized that bRSS-fRSS-mediated recombination and deletion within the Vκ cluster could significantly influence V gene segment availability for V-Jκ rearrangements.

We, therefore, sought to incorporate the probability of loss of Vκ gene segments by bRSS-fRSS-mediated recombination into the prediction of Vκ gene segment use. By considering all possible bRSS/fRSS pairs, we assigned a probability of deletion [p(del)] score to each Vκ gene segment that reflects the probability that a given Vκ gene gets deleted through intra-Vκ cluster rearrangement. We assigned each RSS pair a weight proportional to the RIC score of its bRSS and fRSS component, their position relative to each other on chromosomal loop domains, and the binding of RAG1 measured by ChIP-seq (Figure S4B; see STAR Methods). We then binned the Vκ gene segments into four quartiles with respect to their usage in V-Jκ rearrangements (group 1–4 from low to high use, respectively) and compared these groups based on their bRSS RIC score and deletion probability. We noticed that the bRSS RIC score correctly predicted the ranking of use in three groups (3 > 2 > 1). However, RIC score values did not increase for the highest used group 4 RSS, i.e., the RIC score failed to place this group at the top of the use ranking (Figure 5A). Importantly, group 4 also has the lowest probability of being deleted. By incorporating the p(del) score together with the RIC score, the predictive power of Vκ gene segment use was significantly improved (Figure 5B). We, therefore, propose that the frequency at which a given Vκ gene segment is used is determined both by the strength of its bRSS and its probability to be deleted by intra-Vκ cluster rearrangement (see model in Figure 5C). Notably, stronger Vκ bRSSs tend to rearrange more not only with Jκ RSS but also with fRSS flanking other Vκ genes; therefore, the stronger a bRSS is, the more likely it is to undergo intra-Vκ cluster deletion. These two competing rerrangement pathways reduce the dominance of V gene segments flanked by the best bRSS and allow for the generation of a more balance Igκ repertoire.

DISCUSSION

The 12/23 rule is the cornerstone of ordered rearrangement of V, D, and J gene segments in vertebrate lymphocyte precursors. Indeed, it may have arisen very early at the time of the domestication of the RAG transposon (Zhang et al., 2019). It was, therefore, very surprising to observe extensive RAG-mediated DNA breaks and recombination events within the Vκ cluster in the complete absence of Jκ gene segments and Jκ RC. Although cryptic RSSs (Lieberman et al., 2009) and atypical rearrangement events in the Igκ locus had been reported (Li et al., 2016; Vinocur et al., 2009), their use appears to be relatively infrequent, as determined by HTGTS (Lin et al., 2016) and END-seq (Canela et al., 2016). In contrast, we observed RAG-dependent intra-Vκ cluster recombination frequencies for commonly used Vκ gene segments, in both mouse and human, that occur at comparable levels to the bona fide V-Jκ rearrangement frequencies of the same V gene segment. Many of these products are retained in mature, selected splenic B cells and should, therefore, reside on the chromosome rather than on extrachrosmosomal circles.

Sequence conservation of the fRSS is unique because it is embedded within the 3′ end of the V gene segment where it encodes the last 11 residues of the V domain. The nonamer encodes the highly conserved YYC motif that is present, with minor

variations, in the V domains of all AgRs (Figure 3B). This raised the possibility that V gene segment clusters of other AgR loci also might contain fRSS, but further experiments showed that the Vκ cluster is unique in that the majority of its CDR3-encoding regions are 23 bp long and, therefore, can serve as a compatible 23RSS for its bona fide 12RSS. The germline-encoded CDR3 in Vκ is uniquely long (7–8 amino acids) and functions to span the gap that in other loci is filled by non-templated nucleotide addition catalyzed by terminal deoxynucleotidyl transferase (TdT), which is not expressed in light-chain-rearranging pre-B cells (Victor et al., 1994). In summary, the unique organization of the Igκ locus with 12-bp spacer bRSSs flanking the V gene segments, the near perfect palindrome structure of the consensus heptamer, the lack of TdT activity in pre-B cells compensated by unusually long germline-encoded CDR3, and the nonamer-coding potential of the highly conserved YYC residues in FR3 all contributed to the evolutionarily conserved emergence of a flipped 23RSS to mediate intra-Vκ cluster recombination.

Even if structural requirements for a functional 12/23 RRS pair are met, recruitment of the RAG proteins mediated by RAG2-H3K4me3 interactions is critical to allow initiation of rearrangement. The discovery of exclusive J gene segment-associated RCs in all AgR loci (Ji et al., 2010) offered an elegant explanation for the spatially and temporarily ordered rearrangement process seen in lymphocyte precursors (Schatz and Ji, 2011). It was unexpected, therefore, that we found an exception to this rule and detected significant RAG protein binding to several Vκ gene segments even in the complete absence of Jκ RSS and Jκ RC. Indeed, the widespread cleavage and recombination activity at the Vκ gene segments independent of Jκ gene segments are fully consistent with the idea that an autonomous, Vκ-associated RC would also be present in pre-B cells. The presence and activity of the Vκ RC in ΔJκ pre-B cells clearly indicate that the two RCs can function independently of each other. We propose that the Vκ RC controls intra-Vκ cluster, whereas the Jκ RC controls V-to-J recombination. We note that an analogous situation exists in the TCRα/δ locus where the embedded TCRδ J gene segments form a RC in the middle of the larger TCRα locus. Early deletional TCRδ rearrangements have been shown to be important for subsequent balanced Vα-to-Jα recombinations to take place in developing thymocytes (Carico et al., 2017). Perhaps the reason why evolution did not maintain 12-mer fRSS and autonomous RC in the Vα cluster is because the embedded TCRδ locus provides an internal RC proximal to the Vα genes.

The sheer magnitude of intra-Vκ cluster recombination and its evolutionary conservation imply an important physiological role in Igκ repertoire formation. Rearrangements mediated by bRRS-fRSS pairs could occur either by deletion or inversion, depending on the orientation of the participating RSS. Both reactions would result in a loss of the participating Vκ gene segments from the final V-Jκ repertoire due to irreversible damage to the bRSS. In addition, deletion would remove multiple additional Vκ gene segments contained within the intervening sequences (Figure 5C).

The presence and activity of the Vκ RC in ΔJκ pre-B cells clearly indicate that the two RCs can function independently of each other. We have found that the probability of intra-Vκ cluster recombination was highest among V gene segments flanked by high RIC score bRSS. This would lead to a more balanced use of Vκ gene segments in the final V-Jκ repertoire. Our model adds a fundamentally different layer to the complex regulation of Vκ gene segment use, which includes the dynamic regulation of the 3D conformation of the Igκ locus (Lin et al., 2012) and the developmental activation of intra-Vκ enhancer elements, such as E88 (Barajas-Mora et al., 2019). Generally, intra-Vκ cluster recombination events occur in close proximity to the bait Vκ gene segment (Figure 2A), in accordance with the RAG scanning model of V(D)J recombination (Hu et al., 2015; Jain et al., 2018), which restricts recombination to gene segments located within the same chromatin loop domain. Based on the RAG scanning model, Vκ gene segments near to convergent CBE sites are predicted to be the most frequent targets for rearrangement controlled by the Jκ RC. The presence of an additional, possibly competing RC in the Vκ cluster could further increase the probability of other, CBE-distal V gene segments to gain access to RAG proteins anchored at the Jκ RC and undergo productive V-to-Jκ rearrangements.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell Culture and Mice
- METHOD DETAILS
  - END-seq
  - ChIP-Seq
  - HTGTS
  - Generation of Cas9/gRNA DSBs for HTGTS Libraries
  - Genome Alignment
  - Genomic annotations
  - Sequence Conservation
  - Calculating RAG mediated breakage from END-seq data
  - Heptamer/Nonamer/RSS scores
  - Estimating Probability of Deletion of Vκ segment
  - Data Visualization
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2019.11.088.

ACKNOWLEDGMENTS

We thank Barry P. Sleckman for discussions and the pre-B cells lines; and Frederick W. Alt for suggestions and HTGTS protocols. D.G.S. was supported by NIH grant R01 AI32524 and K.S. was supported by a fellowship from the Uehara Memorial Foundation. Sequencing was performed by NCI CCR Genomics Core and flow cytometry analyses and sorting were performed at the NCI CCR LGI Flow Cytometry Core. The A.N. laboratory is supported by the Intramural Research Program of the NIH; an Ellison Medical Foundation Senior Scholar in Aging Award (AG-SS-2633-11); the Department of Defense Idea Expansion (W81XWH-15-2-006) and Breakthrough (W81XWH-16-1-599)

awards; the Alex Lemonade Stand Foundation Award; and an NIH Intramural FLEX Award.

AUTHOR CONTRIBUTIONS

A.N., K.S., Y.M., and F.L. designed experiments. Y.M. performed bioinformatics analyses. K.S. and A.C. performed experiments. F.L. performed fluorescence-activated cell sorting (FACS). D.G.S. and F.L. provided expert advice, and A.N. supervised the project. K.S., Y.M., F.L., and A.N. wrote the manuscript with comments from the other authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: August 19, 2019
Revised: November 4, 2019
Accepted: November 21, 2019
Published: December 24, 2019

REFERENCES

Aoki-Ota, M., Torkamani, A., Ota, T., Schork, N., and Nemazee, D. (2012). Skewed primary Igκ repertoire and V-J joining in C57BL/6 mice: implications for recombination accessibility and receptor editing. J. Immunol. *188*, 2305–2315.

Barajas-Mora, E.M., Kleiman, E., Xu, J., Carrico, N.C., Lu, H., Oltz, E.M., Murre, C., and Feeney, A.J. (2019). A B-Cell-Specific Enhancer Orchestrates Nuclear Architecture to Generate a Diverse Antigen Receptor Repertoire. Mol. Cell *73*, 48–60.e45.

Bredemeyer, A.L., Helmink, B.A., Innes, C.L., Calderon, B., McGinnis, L.M., Mahowald, G.K., Gapud, E.J., Walker, L.M., Collins, J.B., Weaver, B.K., et al. (2008). DNA double strand breaks activate a multi-functional genetic program in developing lymphocytes. Nature *456*, 819–823.

Bredemeyer, A.L., Sharma, G.G., Huang, C.Y., Helmink, B.A., Walker, L.M., Khor, K.C., Nuskey, B., Sullivan, K.E., Pandita, T.K., Bassing, C.H., and Sleckman, B.P. (2006). ATM stabilizes DNA double-strand-break complexes during V(D)J recombination. Nature *442*, 466–470.

Canela, A., Sridharan, S., Sciascia, N., Tubbs, A., Meltzer, P., Sleckman, B.P., and Nussenzweig, A. (2016). DNA Breaks and End Resection Measured Genome-wide by End Sequencing. Mol. Cell *63*, 898–911.

Canela, A., Maman, Y., Huang, S.N., Wutz, G., Tang, W., Zagnoli-Vieira, G., Callen, E., Wong, N., Day, A., Peters, J.M., et al. (2019). Topoisomerase II-Induced Chromosome Breakage and Translocation Is Determined by Chromosome Architecture and Transcriptional Activity. Mol. Cell *75*, 252–266.

Canela, A., Maman, Y., Jung, S., Wong, N., Callen, E., Day, A., Kieffer-Kwon, K.R., Pekowska, A., Zhang, H., Rao, S.S.P., et al. (2017). Genome Organization Drives Chromosome Fragility. Cell *170*, 507–521.e518.

Carico, Z.M., Roy Choudhury, K., Zhang, B., Zhuang, Y., and Krangel, M.S. (2017). Tcrd Rearrangement Redirects a Processive Tcra Recombination Program to Expand the Tcra Repertoire. Cell Rep. *19*, 2157–2173.

Coster, G., Gold, A., Chen, D., Schatz, D.G., and Goldberg, M. (2012). A dual interaction between the DNA damage response protein MDC1 and the RAG1 subunit of the V(D)J recombinase. J. Biol. Chem. *287*, 36488–36498.

Cowell, L.G., Davila, M., Kepler, T.B., and Kelsoe, G. (2002). Identification and utilization of arbitrary correlations in models of recombination signal sequences. Genome Biol *3*, research0072.1–research0072.20.

Desiderio, S. (2010). Temporal and spatial regulatory functions of the V(D)J recombinase. Semin. Immunol. *22*, 362–369.

Durdik, J., Moore, M.W., and Selsing, E. (1984). Novel kappa light-chain gene rearrangements in mouse lambda light chain-producing B lymphocytes. Nature *307*, 749–752.

Feddersen, R.M., Martin, D.J., and Van Ness, B.G. (1990). Novel recombinations of the IG kappa-locus that result in allelic exclusion. J. Immunol. *145*, 745–750.

Helmink, B.A., and Sleckman, B.P. (2012). The response to and repair of RAG-mediated DNA double-strand breaks. Annu. Rev. Immunol. *30*, 175–202.

Hu, J., Zhang, Y., Zhao, L., Frock, R.L., Du, Z., Meyers, R.M., Meng, F.L., Schatz, D.G., and Alt, F.W. (2015). Chromosomal Loop Domains Direct the Recombination of Antigen Receptor Genes. Cell *163*, 947–959.

Hu, J., Meyers, R.M., Dong, J., Panchakshari, R.A., Alt, F.W., and Frock, R.L. (2016). Detecting DNA double-stranded breaks in mammalian genomes by linear amplification-mediated high-throughput genome-wide translocation sequencing. Nat. Protoc. *11*, 853–871.

Hunt, S.E., McLaren, W., Gil, L., Thormann, A., Schuilenburg, H., Sheppard, D., Parton, A., Armean, I.M., Trevanion, S.J., Flicek, P., and Cunningham, F. (2018). Ensembl variation resources. Database (Oxford) *2018*, bay119.

Jain, S., Ba, Z., Zhang, Y., Dai, H.Q., and Alt, F.W. (2018). CTCF-Binding Elements Mediate Accessibility of RAG Substrates During Chromatin Scanning. Cell *174*, 102–116.e114.

Ji, Y., Resch, W., Corbett, E., Yamane, A., Casellas, R., and Schatz, D.G. (2010). The *in vivo* pattern of binding of RAG1 and RAG2 to antigen receptor loci. Cell *141*, 419–431.

Karki, S., Kennedy, D.E., Mclean, K., Grzybowski, A.T., Maienschein-Cline, M., Banerjee, S., Xu, H., Davis, E., Mandal, M., Labno, C., et al. (2018). Regulated Capture of Vκ Gene Topologically Associating Domains by Transcription Factories. Cell Rep. *24*, 2443–2456.

Karolchik, D., Hinrichs, A.S., Furey, T.S., Roskin, K.M., Sugnet, C.W., Haussler, D., and Kent, W.J. (2004). The UCSC Table Browser data retrieval tool. Nucleic Acids Res. *32*, D493–D496.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res. *12*, 996–1006.

Kent, W.J., Zweig, A.S., Barber, G., Hinrichs, A.S., and Karolchik, D. (2010). BigWig and BigBed: enabling browsing of large distributed datasets. Bioinformatics *26*, 2204–2207.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Li, S., Liu, W., Li, Y., Zhao, S., Liu, C., Hu, M., Yue, W., Liu, Y., Wang, Y., Yang, R., et al. (2016). Contribution of secondary Igkappa rearrangement to primary immunoglobulin repertoire diversification. Mol. Immunol. *78*, 193–206.

Lieberman, A.E., Kuraoka, M., Davila, M., Kelsoe, G., and Cowell, L.G. (2009). Conserved cryptic recombination signals in Vkappa gene segments are cleaved in small pre-B cells. BMC Immunol. *10*, 37.

Lin, Y.C., Benner, C., Mansson, R., Heinz, S., Miyazaki, K., Miyazaki, M., Chandra, V., Bossen, C., Glass, C.K., and Murre, C. (2012). Global changes in the nuclear positioning of genes and intra- and interdomain genomic interactions that orchestrate B cell fate. Nat. Immunol. *13*, 1196–1204.

Lin, S.G., Ba, Z., Du, Z., Zhang, Y., Hu, J., and Alt, F.W. (2016). Highly sensitive and unbiased approach for elucidating antibody repertoires. Proc. Natl. Acad. Sci. USA *113*, 7846–7851.

Lin, S.G., Ba, Z., Alt, F.W., and Zhang, Y. (2018). RAG Chromatin Scanning During V(D)J Recombination and Chromatin Loop Extrusion are Related Processes. Adv. Immunol. *139*, 93–135.

Livak, F., and Schatz, D.G. (1996). T-cell receptor alpha locus V(D)J recombination by-products are abundant in thymocytes and mature T cells. Mol. Cell. Biol. *16*, 609–618.

Livak, F., Burtrum, D.B., Rowen, L., Schatz, D.G., and Petrie, H.T. (2000). Genetic modulation of T cell receptor gene segment usage during somatic recombination. J. Exp. Med. *192*, 1191–1196.

Matthews, A.G., and Oettinger, M.A. (2009). RAG: a recombinase diversified. Nat. Immunol. *10*, 817–821.

Merelli, I., Guffanti, A., Fabbri, M., Cocito, A., Furia, L., Grazini, U., Bonnal, R.J., Milanesi, L., and McBlane, F. (2010). RSSsite: a reference database and prediction tool for the identification of cryptic Recombination Signal Sequences in human and murine genomes. Nucleic Acids Res. *38*, W262–W267.

Muljo, S.A., and Schlissel, M.S. (2003). A small molecule Abl kinase inhibitor induces differentiation of Abelson virus-transformed pre-B cell lines. Nat. Immunol. *4*, 31–37.

Nemazee, D. (2017). Mechanisms of central tolerance for B cells. Nat. Rev. Immunol. *17*, 281–294.

Ortega-Molina, A., Boss, I.W., Canela, A., Pan, H., Jiang, Y., Zhao, C., Jiang, M., Hu, D., Agirre, X., Niesvizky, I., et al. (2015). The histone lysine methyltransferase KMT2D sustains a gene expression program that represses B cell lymphoma development. Nat. Med. *21*, 1199–1208.

Proudhon, C., Hao, B., Raviram, R., Chaumeil, J., and Skok, J.A. (2015). Long-Range Regulation of V(D)J Recombination. Adv. Immunol. *128*, 123–182.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Rubelt, F., Bolen, C.R., McGuire, H.M., Vander Heiden, J.A., Gadala-Maria, D., Levin, M., Euskirchen, G.M., Mamedov, M.R., Swan, G.E., Dekker, C.L., et al. (2016). Individual heritable differences result in unique cell lymphocyte receptor repertoires of naïve and antigen-experienced cells. Nat. Commun. *7*, 11112.

Sakano, H., Hüppi, K., Heinrich, G., and Tonegawa, S. (1979). Sequences at the somatic recombination sites of immunoglobulin light-chain genes. Nature *280*, 288–294.

Schatz, D.G., and Ji, Y. (2011). Recombination centres and the orchestration of V(D)J recombination. Nat. Rev. Immunol. *11*, 251–263.

Shimizu, T., and Yamagishi, H. (1992). Biased reading frames of pre-existing DH–JH coding joints and preferential nucleotide insertions at VH–DJH signal joints of excision products of immunoglobulin heavy chain gene rearrangements. EMBO J. *11*, 4869–4875.

Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S., et al. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res. *15*, 1034–1050.

Siminovitch, K.A., Moore, M.W., Durdik, J., and Selsing, E. (1987). The human kappa deleting element and the mouse recombining segment share DNA sequence homology. Nucleic Acids Res. *15*, 2699–2705.

Teng, G., and Schatz, D.G. (2015). Regulation and Evolution of the RAG Recombinase. Adv. Immunol. *128*, 1–39.

Teng, G., Maman, Y., Resch, W., Kim, M., Yamane, A., Qian, J., Kieffer-Kwon, K.R., Mandal, M., Ji, Y., Meffre, E., et al. (2015). RAG Represents a Widespread Threat to the Lymphocyte Genome. Cell *162*, 751–765.

Tonegawa, S. (1983). Somatic generation of antibody diversity. Nature *302*, 575–581.

van Gent, D.C., Ramsden, D.A., and Gellert, M. (1996). The RAG1 and RAG2 proteins establish the 12/23 rule in V(D)J recombination. Cell *85*, 107–113.

Victor, K.D., Vu, K., and Feeney, A.J. (1994). Limited junctional diversity in kappa light chains. Junctional sequences from CD43+B220+ early B cell progenitors resemble those from peripheral B cells. J. Immunol. *152*, 3467–3475.

Vinocur, J.M., Fesnak, A.D., Liu, Y., Charan, D., and Prak, E.T. (2009). Violations of the 12/23 rule at the mouse immunoglobulin kappa locus, including V kappa-V kappa rearrangement. Mol. Immunol. *46*, 2183–2189.

Wu, C., Bassing, C.H., Jung, D., Woodman, B.B., Foy, D., and Alt, F.W. (2003). Dramatically increased rearrangement and peripheral representation of Vbeta14 driven by the 3'Dbeta1 recombination signal sequence. Immunity *18*, 75–85.

Yu, K., Taghva, A., and Lieber, M.R. (2002). The cleavage efficiency of the human immunoglobulin heavy chain VH elements by the RAG complex: implications for the immune repertoire. J. Biol. Chem. *277*, 5040–5046.

Zhang, Y., Cheng, T.C., Huang, G., Lu, Q., Surleac, M.D., Mandell, J.D., Pontarotti, P., Petrescu, A.J., Xu, A., Xiong, Y., and Schatz, D.G. (2019). Transposon molecular domestication and the evolution of the RAG recombinase. Nature *569*, 79–84.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

Zhao, L., Frock, R.L., Du, Z., Hu, J., Chen, L., Krangel, M.S., and Alt, F.W. (2016). Orientation-specific RAG activity in chromosomal loop domains contributes to Tcrd V(D)J recombination during T cell development. J. Exp. Med. *213*, 1921–1936.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit monoclonal anti-RAG1 (mAb 23)	Coster et al., 2012	N/A
Rabbit monoclonal anti-RAG2 (mAb 39)	Coster et al., 2012	N/A
Rabbit polyclonal anti-H3K4me3	Millipore	Cat# 07-473; RRID: AB_1977252
eFluor 450 Rat monoclonal CD24 (M1/69)	Thermo Fisher Scientific	Cat# 48-0242-82; RRID: AB_1311169
APC Rat monoclonal CD43 (S11)	BioLegend	Cat# 143207; RRID: AB_11149489
APC/Cy7 Rat anti-Mouse/Human B220 (RA3-6B2)	BioLegend	Cat# 103224; RRID: AB_313007
FITC Rat anti-Mouse IgM (II/41)	BD Biosciences	Cat# 553437; RRID: AB_394857
Bacterial and Virus Strains		
Bacteria: TOP10 Chemically Competent *E. coli*	Thermo Fisher Scientific	Cat# C404006
Lentivirus: pKLV-U6gRNA-EF(BbsI)-PGKpuro2ABFP	Addgene	RRID:Addgene_62348
Mammalian expression: hCas9	Addgene	RRID:Addgene_41815
Mammalian expression: pcDNA3.1-Hygro-delta-hCD4	Gift from Eugene Oltz	N/A
Mammalian expression: pX330-U6-Chimeric_BB-CBh-hSpCas9	Addgene	RRID:Addgene_42230
Chemicals, Peptides, and Recombinant Proteins		
Dynabeads Protein A	Thermo Fisher Scientific	Cat# 10002D
CD43 microbeads (Ly-48)	Miltenyi Biotec	Cat# 130-049-801
Imatinib mesylate (STI-571)	Selleckchem	Cat# S1026
DAPI	Thermo Fisher Scientific	Cat# 62248
cOmplete, Mini Protease inhibitor cocktail	Roche Diagnostics	Cat# 11836153001
Puregene Proteinase K enzyme	QIAGEN	Cat# 158920
Puregene RNase A Solution	QIAGEN	Cat# 158924
T4 DNA Polymerase	NEB	Cat# M0203L
T4 Polynucleotide Kinase	NEB	Cat# M0201L
DNA Polymerase I, Large (Klenow) Fragment	NEB	Cat# M0210L
Exonuclease T (ExoT)	NEB	Cat# M0265L
Klenow Fragment (3′→5′ exo-)	NEB	Cat# M0212L
Quick Ligation Kit	NEB	Cat# M2200L
USER enzyme	NEB	Cat# M5505L
KAPA HiFi HotStart ReadyMix (2X)	KAPA Biosystems	Cat# KK2600
MyOne Streptavidin C1 Beads	ThermoFisher	Cat# 650-01
Agencourt AMPure XP beads	Beckman Coulter	Cat# A63881
Hexaamminecobalt(III) chloride	Sigma-Aldrich	Cat# 481521
T4 DNA Ligase	Promega	Cat# M1804
Phusion High-Fidelity PCR Master Mix with HF buffer	NEB	Cat# M0531L
Critical Commercial Assays		
KAPA Library Quantification Kit	Kapa Biosciences	Cat# KK4824
CHEF Mammalian Genomic DNA plug kit	Bio-Rad	Cat# 1703591
EasySep Human CD4 Positive Selection Kit II	STEMCELL technologies	Cat# 18052
EasySep Human B Cell Isolation Kit	STEMCELL technologies	Cat# 17954
Human B cell Nucleofector Kit	Lonza	Cat# VPA-1001
Deposited Data		
Raw and analyzed data	This paper	GSE140677
ChIP-seq for RAG1	Teng et al., 2015	GSE69478

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental Models: Cell Lines		
Pre-B cell lines	Bredemeyer et al., 2006	N/A
Pre-B cell lines: *ΔJk*	This paper	N/A
Pre-B cell lines: *ΔJk-RS*	This paper	N/A
Pre-B cell lines: *Rag1*$^{-/-}$	Bredemeyer et al., 2006	N/A
Pre-B cell lines: *Lig4*$^{-/-}$	Bredemeyer et al., 2008	N/A
Pre-B cell lines: *Artemis*$^{-/-}$ clone1	Bredemeyer et al., 2008	N/A
Pre-B cell lines: *Artemis*$^{-/-}$ clone2	Bredemeyer et al., 2008	N/A
Experimental Models: Organisms/Strains		
C57BL/6NCr mice	Charles River	Strain code# 027
Mouse: *Vavp-Bcl2* Tg	Ortega-Molina et al., 2015	N/A
Oligonucleotides		
sgRNA to generate ΔJk v-Abl cells: Sense 5′-AAGCAT GCGTGGAAGCGCTT-3′; Antisense 5′-GGGCTCATTA TCAGTTGACG-3′	This paper	N/A
sgRNA to generate ΔJk-RS v-Abl cells: Sense 5′-ATC ACACGTATAGAGTAAGC-3′; Antisense 5′-CCTGCCC ACACGACTCCTTC-3′	This paper	N/A
Primers for genotyping ΔJk allele: ΔJk-Fw, 5′-ACTAAC TGCTGAGCCACCTC-3′; ΔJk-Rv, 5′-GCAGTCAGACC CAGATCTCAA-3′; ΔJk-Intact-Rv, 5′-AGCCACAGACA TAGACAACGG-3′	This paper	N/A
Primers for genotyping ΔJk-RS allele: ΔJk-RS-Fw, 5′-ACC TGGGGAACAAAACTGGA-3′; ΔJk-RS-Rv, 5′-AATCTGCC TGTCTGAAGCCC-3′; ΔJk-RS-Intact-Rv, 5′-GGAAGACAA AGGAGGCCACG-3′	This paper	N/A
sgRNA to downstream of Vk1-117: Vk1-117 CRISPR-Cas9 5′-TTGCTACATATCTGGCACCG-3′	This paper	N/A
END-seq adaptor 1, 5′-phosphate -GATCGGAAGAGCGTC GTGTAGGGAAAGAGTGUU[Biotin-dT]U[Biotin-dT]UUAC ACTCTTTCCCTACACGACGCTCTTCCGATC*T-3′	Canela et al., 2016	N/A
END-seq adaptor 2, 5′-phosphate -GATCGGAAGAGCACA CGTCUUUUUUUUAGACGTGTGCTCTTCCGATC*T-3′	Canela et al., 2016	N/A
TruSeq barcoded primer p5, 5′-AATGATACGGCGAC CACCGAGATCTACACNNNNNNNNACACTCTTTCCC TACACGACGCTCTTCCGATC*T-3′ (N = barcode)	Canela et al., 2019	N/A
TruSeq barcoded primer p7, 5′-CAAGCAGAAGACGGC ATACGAGANNNNNNNGTGACTGGAGTTCAGACGTGT GCTCTTCCGATC*T- 3′ (N = barcode)	Canela et al., 2019	N/A
Illumina truncated adaptor 1, 5′-ACACTCTTTCCCTACA CGACGCTCTTCCGATC*T-3′	Canela et al., 2017	N/A
Illumina truncated adaptor 2, 5′-phosphate-GATCGGAA GAGCACACGTCT-3′	Canela et al., 2017	N/A
Adaptor-upper 5′-GGTACACGACGCTCTTCCGATCTN NNNNN/3AmMO/-3′	Canela et al., 2019	N/A
Adaptor-lower 5′-phosphate-AGATCGGAAGAGCGTCG TGTACC/3AmMO/-3′	Canela et al., 2019	N/A
I5-bridge 5′-AATGATACGGCGACCACCGAGATCTACA CTCTTTCCCTACACGACGCTCTTCCGATC*T-3′	Canela et al., 2019	N/A
P5-I5c 5′-AATGATACGGCGACCACCGAGATCTACAC TCTTT*C-3′	Canela et al., 2019	N/A
P7-I7c 5′-CAAGCAGAAGACGGCATACGAGATCGGTCTC GGCATTCCTGCTGAACCGCTCTT*C-3′	Canela et al., 2019	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Jk1-bio 5′-/Biosg/TTCCCAGCTTTGCTTACGGAG-3′	Lin et al., 2016	N/A
I7-Jk1-nested-barcode 5′-CTCGGCATTCCTGCTGAACC GCTCTTCCGATCTNNNNNNNNAGTGCCAGAATCTGGT TTCAGAG-3′ (N = barcode)	This paper	N/A
Vk1-117-bio 5′-/5Biosg/CAGAAGCCTTCAGTATGCACCA-3′	This paper	N/A
I7-Vk1-117-nested-barcode 5′-CTCGGCATTCCTGCTGAA CCGCTCTTCCGATCTNNNNNNNNCAGGGACAGATTTCA CACTCAAG-3′ (N = barcode)	This paper	N/A
Vk8-34-bio 5′-/5Biosg/CAGAAACCAGGACGATCTCCT-3′	This paper	N/A
I7-Vk8-34-nested-barcode 5′-CTCGGCATTCCTGCTGAACCGC TCTTCCGATCTNNNNNNNNACTAGGGTATCTGGAGTCCCTG-3′	This paper	N/A
bio-human-IGJK1 5′-/5Biosg/TCCCCAGGACATTTCTGAAG-3′	This paper	N/A
I7-human-IGJK1-barcode 5′-CTCGGCATTCCTGCTGAACCG CTCTTCCGATCNNNNNNNNGGGCTGATTGCAGAGTCACCT-3′	This paper	N/A
bio-human-IGVK2-28 5′-/5Biosg/TAACTTTGCAATTCATTAT TTCAGGA-3′	This paper	N/A
I7-human-IGVK2-28-barcode 5′-CTCGGCATTCCTGCTGAACCG CTCTTCCGATCNNNNNNNNCTGGATACAACTATTTGGATTGG-3′	This paper	N/A
bio-human-IGVK3-20 5′-/5Biosg/GCACCCTGTCTTTGTCTCCA-3′	This paper	N/A
I7-human-IGVK3-20-barcode 5′-CTCGGCATTCCTGCTGAACC GCTCTTCCGATCNNNNNNNNTATCAGCAGACTGGAGCCTGAA-3′	This paper	N/A
Software and Algorithms		
Prism 8	GraphPad	https://www.graphpad.com/
RStudio	RStudio Team	https://rstudio.com/
Bowtie 1.1.2	Langmead et al., 2009	https://sourceforge.net/projects/bowtie-bio/files/bowtie/1.1.2/
MACS 1.4.3	Zhang et al., 2008	https://pypi.org/pypi/MACS/1.4.3
UCSC database	Karolchik et al., 2004	https://genome.ucsc.edu
UCSC genome browser	Kent et al., 2002	https://genome.ucsc.edu
Bedtools	Quinlan and Hall, 2010	https://github.com/arq5x/bedtools2
R 3.6.1	R Core Team	https://www.r-project.org/
FlowJo (10.1)	FlowJo LLC	https://www.flowjo.com/
Other		
Amaxa Nucleofector II	Lonza	N/A
FACSAria II cell sorter	BD Biosciences	N/A

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Andre Nussenzweig (andre_nussenzweig@nih.gov). All unique/stable reagents generated in this study are available from the Lead Contact with a completed Materials Transfer Agreement.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell Culture and Mice

Abelson-transformed pre-B (v-Abl) cell lines WT, *Rag1*$^{-/-}$, *Lig4*$^{-/-}$, and *Artemis*$^{-/-}$ (Clone1 and Clone2) were obtained from Barry Sleckman (Bredemeyer et al., 2006). In all experiments, pre-B cells were arrested in G1 with 3 μM STI-571 (imatinib) for 48 hr (Bredemeyer et al., 2006). ΔJk and ΔJk-RS clones were generated by CRISPR/Cas9-mediated gene deletion. Ten million v-Abl cells were electroporated with hCas9 plasmid (Addgene 41815), expression plasmids for two guide RNA (gRNAs) targeting sequences that flank the region to be deleted, and a plasmid encoding hCD4 (pcDNA3.1-Hygro-delta-hCD4). hCD4^{+} cells were purified 24 hour post-transfection using magnetic beads (STEMCELL Technologies 18052), passaged for ~7 days, subcloned by single cell sorting, and screened for deletions using multiple independent primer pairs outside and inside of the gRNA target sites. The gRNA sequences

are shown in Table S1. gRNAs were cloned into pKLV-U6gRNA(BbsI)-PGKpuro2ABFP (Addgene 62348). PCR primers for screening the deletions are provided in Table S1. PCR products spanning deletion sites were purified and confirmed with Sanger sequencing.

Mature resting B cells were obtained from spleens of 6-12 week old wild-type C57BL/6 mice using magnetic depletion with anti-CD43 MicroBeads (Miltenyi Biotech). Primary mouse bone marrow pre-B cells for END-Seq analyses were obtained from 6-12 week old Vavp-Bcl2-transgenic mice (Ortega-Molina et al., 2015) and cultured in the presence of recombinant IL-7 (ThermoFisher) for 7-10 days at 2×10^6 cells/ml *in vitro*. To induce cell cycle arrest, cells were resuspended in media without IL-7 and maintained at 2×10^6 cells/ml for 48 hours and harvested for END-seq. Primary mouse bone marrow pre-B cells for HTGTS analyses were obtained from 6-12 week old wild-type, C57BL/6 mice with flow cytometry cell sorting on a FACSAria II cell sorter (BD Biosciences) using a combination of anti-B220, anti-CD43 (BioLegend), anti-CD24 (ThermoFisher/eBioscience) and anti-IgM antibodies (BD Biosciences) (Lin et al., 2016). Both male and female mice were indiscriminately used for these studies. Human peripheral blood B cells were isolated from deidentified, donated PBMC using human B cell isolation kit (STEMCELL Technologies 17954). All animal experimentation was approved by the NCI Animal Care and Use Committee.

METHOD DETAILS

END-seq

Single cell suspension of v-Abl cells (30 million), primary pre B cells (10 million) or thymocytes (70 million) were embedded in a single agarose plug (1% agarose final). Embeded cells were lysed and digested using Proteinase K (50°C, 1 hour then 37°C for 7 hours). Plugs were rinsed in TE buffer and treated with RNase A at 37°C for 1 hour and DNA ends were blunted. DNA was retained in agarose plugs to prevent shearing throughout the ssDNA blunting reactions. The first blunting reaction was performed using ExoT (NEB, M0265S) for 45 min, 24°C. After blunting, two washes were performed in NEB Buffer 2 (1x), followed by A-tailing to attach dA to the free 3′-OH (Klenow 3′- > 5′exo-, NEB, M0212S). The A-tailed product was ligated to "END-seq hairpin adaptor 1," listed in reagents section, using NEB Quick Ligation Kit (NEB, M2200S).

Agarose plugs were melted and dissolved, and DNA was sonicated to a median shear length of 170bp using a Covaris S220 sonicator for 4 min at 10% duty cycle, peak incident power 175, 200 cycles per burst at 4°C. DNA was ethanol-precipitated and dissolved in 70 μL TE buffer. 35 μL of Dynabeads were washed twice with 1 mL Binding and Wash Buffer (1xBWB) (10 mM Tris-HCl pH8.0, 1 mM EDTA, 1 M NaCl, 0.1% Tween20). Beads were recovered using a DynaMag-2 magnetic separator (12321D, Invitrogen). Supernatants were discarded. Washed beads were resuspended in 70 μL 2xBWB (10 mM Tris-HCl pH8.0, 2 mM EDTA, 2 M NaCl) combined with the 70 μL of sonicated DNA and incubated at 24°C for 30 min in a ThermoMixer C at 400 rpm.

Following 30 min mixing, the supernatant was removed and the bead-bound biotinylated DNA was washed 3 times with 1 mL 1xBWB, twice with 1 mL EB buffer, once with 1 mL T4 ligase reaction buffer (NEB) and then resuspended in 50 μL of end-repair reaction mix (0.4 mM of dNTPs, 2.7 U of T4 DNA polymerase (NEB), 9 U of T4 Polynucleotide Kinase (NEB) and 1 U of Klenow fragment (NEB). The end-repair reaction was incubated at 24°C for 30 min in a ThermoMixer C at 400 rpm (tubes were vortexed every 10 min). The supernatant was removed using a magnetic separator and beads were then washed once with 1 mL 1xBWB, twice with 1 mL EB buffer, once with 1 mL NEBNext dA-Tailing reaction buffer (NEB) and then resuspended in 50 μL ofA-tailing reaction with NEBNext dA-Tailing reaction buffer (NEB) and 20 U of Klenow fragment exo- (NEB). The A-tailing reaction was incubated at 37°C for 30 min in a ThermoMixer C at 400 rpm (tubes were vortexed every 10 min). The supernatant was removed using a magnetic separator and washed once with 1 mL NEBuffer 2 and then resuspended in 115 mL of Ligation reaction with Quick Ligase buffer (NEB), 6,000 U of Quick Ligase (NEB) and ligated to "END-seq hairpin adaptor 2" and incubated at 25°C for 30 min in a ThermoMixer C at 400 rpm. Ligation reactions were stopped by adding 50 mM of EDTA, then beads were washed 3 times with BWB and 3 times with EB, and eluted in 8 μL of EB. Hairpin adaptors were digested using USER enzyme (NEB, M5505S) at 37°C, 30 minutes. PCR amplification was performed in 50 μL reaction with 10 mM primers 5′-CAAGCAGAAGACGGCATACGA-GATNNNNNNGTGACTG GAGTTCAGACGTGTGCTCTTCCGATC*T-3′ and 5′-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATC*T-3′, and 2X Kapa HiFi HotStart Ready mix (Kapa Biosciences), where * represents a phosphothiorate bond and NNNNNN a Truseq index sequence. PCR amplification was performed at 98°C, 45 s; 15 cycles [98°C, 15 s; 63°C, 30 s; 72°C, 30 s]; 72°C, 5 min. PCR reactions were cleaned with AMPure XP beads, and 200-500 bp fragments were isolated after running on 2% agarose gel. Libraries were purified using QIA-quick Gel Extraction Kit (QIAGEN). Library concentration was determined with KAPA Library Quantification Kit for Illumina Platforms (Kapa Biosystems). Sequencing was performed on using Illumina NextSeq 500 or 550 (75bp single end reads). A more detailed END-seq protocol can be found in Canela et al. (2016), (2017).

ChIP-Seq

ChIP-seq was performed as described previously (Canela et al., 2017). Rabbit monoclonal antibodies for RAG1 (mAb 23) and RAG2 (mAb 39) (Coster et al., 2012) and a polyclonal rabbit antibody for H3K4me3 (Millipore) were used. Cells were fixed by adding 37% formaldehyde (F1635, Sigma) to a final concentration of 1% and incubated at 37°C for 10 min. Fixation was quenched by addition of 1M glycine (Sigma) in PBS at a final concentration of 125 mM. Twenty million fixed cells were washed twice with cold PBS and pellets were snap frozen in dry ice and stored at −80°C. Fixed cell pellets of 20 million cells were thawed on ice and resuspended in 2 mL of cold RIPA buffer (10 mM TrisHCl pH 7.5, 1 mM EDTA, 0.1% SDS, 0.1% sodium deoxycholate, 1% Triton X-100, 1 Complete Mini EDTA free proteinase inhibitor (Roche)). Sonication was performed using the Covaris S220 sonicator at duty cycle 20%, peak incident

power 175, cycle/burst 200 for 30 min at 4°C or using the Branson sonifier at amplitude 35%, 12 cycles of 20 s sonication and 30 s of pause at 4°C. Chromatin were clarified by centrifugation at 21,000 g at 4°C for 10 min and precleared with 80 ul prewashed Dynabeads protein A (ThermoFisher) for 30 min at 4°C. 40 μL prewashed Dynabeads protein A were incubated with 10 μg of each respective antibody in 100 μL of PBS for 10 min at room temperature in continuous mixing, washed twice in PBS for 5 min and added to 1 mL of chromatin followed by overnight incubation at 4°C on a rotator. Beads were then collected in a magnetic separator (DynaMag-2 Invitrogen), washed twice with cold RIPA buffer, twice with RIPA buffer containing 0.3M NaCl, twice with LiCl buffer (0.25 M LiCl, 0.5% Igepal-630, 0.5% sodium deoxycholate), once with TE (10 mM Tris pH 8.0, 1mM EDTA) plus 0.2% Triton X-100, and once with TE. Crosslinking was reversed by incubating the beads at 65°C for 4 hr in the presence of 0.3% SDS and 1mg/mL of Proteinase K (Ambion). DNA was purified using Zymo ChIP DNA clean and concentrator kit (Zymo Research) and eluted in 20 μl. The entire ChIP DNA was used to prepare Illumina sequencing libraries. End-repair was performed in 75 μL of T4 ligase reaction buffer, 0.4 mM of dNTPs, 4 U of T4 DNA polymerase (NEB), 13.5 U of T4 Polynucleotide Kinase (NEB) and 1.5 U of Klenow fragment (NEB) at 24°C for 30 min in a ThermoMixer C at 400 rpm. End-repair reaction was cleaned using 2X Agencourt AMPure XP beads and eluted in 15 μL of EB that was used for A-tailing reaction in 30 μL of NEBNext dA-Tailing reaction buffer (NEB) with 7.5 U of Klenow fragment exo- (NEB) at 37°C for 30 min. The 30 μL of the A-tailing reaction was mixed with Quick Ligase buffer 2X (NEB), 3,000 U of Quick ligase and 5 nM of annealed adaptor (Illumina truncated adaptor) in a volume of 75 μL and incubated at 25°C for 20 min. The adaptor was prepared by annealing the following HPLC oligos: 5′-phosphate/GATCGGAAGAGCACACGTCT-3′ and 5′-ACACTCTTTCCCTACACGACGCT CTTCCGATC*T-3′ (*phosphorothioate bond). Ligation was stopped by adding 50mM of EDTA and cleaned with 1.8X Agencourt AM- Pure XP beads and eluted in 15ul of EB that was used for PCR amplification in a 50 uL reaction with 1 μM primers TruSeq barcoded primer p5, AATGATACGGCGACCACCGAGATCTACACNNNNNNNNACACTCTTTCCCTACACGACGCTCTTCCGATC*T, TruSeq barcoded primer p7, CAAGCAGAAGACGGCATACGAGANNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T, (NNNNNNNN represents barcode and * a phosphothiorate bond), and 2X Kapa HiFi HotStart Ready mix (Kapa Biosciences). The temperature settings during the PCR amplification were 45 s at 98°C followed by 15 cycles of 15 s at 98°C, 30 s at 63°C, 30 s at 72°C and a final 5 min extension at 72°C . PCR reactions were cleaned with Agencourt AMPure XP beads (Beckman Coulter), run on a 2% agarose gel and a smear of 200-500bp was cut and gel purified using QIAquick Gel Extraction Kit (QIAGEN). Library concentration was determined with KAPA Library Quantification Kit for Illumina Platforms (Kapa Biosystems). Sequencing was performed on the Illumina Nextseq500 (75bp single-end reads).

HTGTS

HTGTS was performed as described (Hu et al., 2015). Genomic DNA was extracted from splenic B cells, isolated B cells from human PBMC, or v-Abl cells arrested in G1 for 3 days by treatment with 3 μM of STI-571. Briefly, 20-50 ug of DNA was fragmented via sonication on a. Diagenode bioruptor and subjected to linear PCR amplification with a biotinylated primer. Single-stranded PCR products were purified via Dynabeads MyONE C1 streptavidin beads (Life Technologies, 65002) and ligated to bridge adaptors. Adaptor-ligated products were amplified via nested PCR with indexed locus-specific primers and primer annealed to adaptor. The PCR products were further tagged with Illumina sequencing adaptor sequences, size-selected (fragment size 500 – 1,000 bp) via gel extraction and loaded to Illumina MiSeq for paired-end 250 bp sequencing. For translocation analysis, the standard LAM-HTGTS bioinformatic pipeline was used (Hu et al., 2015). Primer information can be found in Table S1.

Generation of Cas9/gRNA DSBs for HTGTS Libraries

Cas9/gRNA was design to generate DSBs 654 bp downstream of the Vκ1-117 bRSS breakage site. HTGTS primer was design that allowed 5′ broken ends of these DSBs to be used as a bait. WT v-Abl cells were treated with 3 μM imatinib at a concentration of $5x10^5$ cells/ml for 30 hours followed by nucleofection of pX330-Cas9-Vκ1-117-CRSPR plasmid (15 μg) into $20x10^6$ G1-arrested cells (2 reactions; $10x10^6$ cells each reaction) using X-001 program of Amaxa Nucleofector II (Lonza) with Human B cell Nucleofector kit (Lonza). Transfected cells were cultured in DMEM medium with 15% (v/v) FBS plus 3 μM imatinib for 3 more days before harvested for genomic DNA. HTGTS libraries were prepared and processed as described above. Primer information can be found in Table S1.

Genome Alignment

For END-seq and ChIP-seq, tags were aligned to the mouse (GRCm38p2/mm10) or human (GRCm37/hg19) genomes using Bowtie (version 1.1.2) (Langmead et al., 2009) with the options–best–all –strata–l 50. For ChIP-seq, we allow 2 mismatches and discarded tags with multiple alignments (-n 2 -m 1). For END-seq, we allowed 3 mismatches and kept the best strata for tags with multiple alignments (-n 3 –k 1). HTGTS reads, and the identification of sites that underwent chromosomal translocations to the bait site, were done using HTGTS pipeline, as described earlier (Hu et al., 2016)

Genomic annotations

bRSSs were identified at the flanks of each Ig/TCR segment as defined by GENCODE version M22 (Ensmbl 97) (Hunt et al., 2018). Ig/TCR CDR3 and FR3 were obtained from IMGT data base. In cases where more than one allele existed for a gene segment, all alleles were included. Vκ loop annotation was taken from published data (Karki et al., 2018).

Sequence Conservation

Sequence conservation around RSSs was calculated as the averaged PhastCons (60-vertebrates) (Siepel et al., 2005) conservation score over all sequences at a specific position.

Calculating RAG mediated breakage from END-seq data

Total RAG mediated breakage in a locus was calculated as the total breakage within the locus that occurs precisely at the 5′ of the minimal RAG target sequence – CAC/GTG. Signal was calculated as reads-per-million (RPM).

Heptamer/Nonamer/RSS scores

To estimate the strength of individual heptamer/nonamer, a position-weight-matrix (PWM) was first built based on Vκ bRSSs. Using this PWM, heptamer/nonamer score was defined as $\sum_{ij} PWM_{ij} S_{ij} E_j$, where $PWM_{i,j}$ is the probability to for nucleotide i at position j in Vκ RSS, $S_{i,j}$ is 1 if heptamer/nonamer S has nucleotide i at position j, and 0 otherwise, and E_j is the entropy at position j.

RSS strength quantification is also obtained by using the published online algorithm (Merelli et al., 2010), which calculates RIC scores (Cowell et al., 2002).

Estimating Probability of Deletion of Vκ segment

The probability of a Vκ segment to get deleted is obtained by combination of three factors.

The first is the sum of probabilities of all the possible 12/23 (bfRSS/flRSS) pairs (see Figure S4B for illustration). This probability, p(pair) was calculated as the product of the probability of the two RSS to be functional, which in turn is estimated as MinMax function of the RSS RIC score (p12 and p23 for 12 and 23RSS, respectively):

$$p(pair_{ij}) = p12_i * p23_j$$

The second is a binary factor, *r*, indicating whether RAG binds to at least one of the RSS pair mates based on the RAG ChIP seq data (Figures 4A and 4E).

The third factor, l, is a loop coefficient, giving higher pairing probability for RSS pairs that share the same loop (see Genomic Annotations section above).

Using the above factors, p(del) can be described as:

$$p(del)_k = \sum p(pair)_{ij} * r_{ij} * l_{ij}$$

where $k \geq \min(i,j)$ and $k \leq \max(i,j)$.

Data Visualization

Aligned-reads bed files were first converted to bedgraph files using *bedtools genomecov* (Quinlan and Hall, 2010) following by *bedGraphToBigWig* to make a bigwig file (Kent et al., 2010). Visualization of genomic profiles was done by the UCSC browser (Kent et al., 2002). Genome browser profiles were normalized to present RPM.

Heatmaps were produced using the R package *pheatmap*. For Seq logo representation, position-weight-matrices (PWMs) was first produced from the array of sequences. These PWMs were used as an input for seq logo using R package *seqLogo*.

QUANTIFICATION AND STATISTICAL ANALYSIS

Fisher test has been used for comparison of fraction of breakage between fRSS and other Vκ heptamers (Figure 1C), the CDR3 size of the different AgR loci (Figure 3C), and breakage fraction of bRSS and fRSS between Igκ and TCRα (Figure S3). t test, has been used to compare 12RSS and 23RSS nonamer scores between broken and not broken fRSS, and to compare RIC scores and deletion scores between different usage groups. Paired t test has been used to compare bRSS and fRSS (Figure 1H) breakage and RAG1 binding (Figure 4D) between WT and ΔJK cells. Finally, pearson correlation has been performed between usage and RIC score, and combination of RIC and deletion scores (Figure 5B). All statistical test above were performed using R version 3.6.1 (https://www.r-project.org/).

DATA AND CODE AVAILABILITY

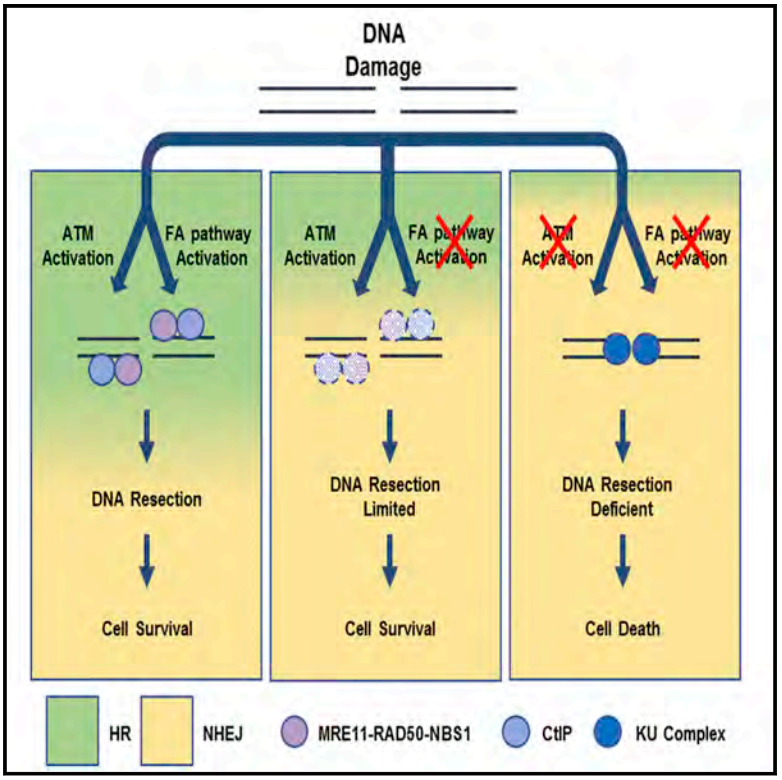
The accession number for the sequencing data reported in this paper is GEO: GSE140677.

Article

Cell Reports

Cooperation of the ATM and Fanconi Anemia/BRCA Pathways in Double-Strand Break End Resection

Graphical Abstract



Highlights

- Genome-wide CRISPR screen identifies determinants of ATM inhibitor sensitivity
- Loss of Fanconi anemia (FA) pathway and ATM is synthetic lethal
- ATM is required in FA-deficient cells to promote DNA end resection and HR repair
- Loss of FA pathway and ATM results in accumulation of toxic levels of NHEJ

Authors

Mu-Yan Cai, Connor E. Dunn, Wenxu Chen, ..., Kalindi Parmar, David Kozono, Alan D. D'Andrea

Correspondence

alan_dandrea@dfci.harvard.edu

In Brief

ATM inhibitors are currently in clinical development as anti-cancer agents. Using a genome-wide CRISPR screen, Cai et al. demonstrate that cancer cells with Fanconi anemia (FA) pathway deficiency are sensitive to ATM inhibition. The authors also show that synthetic lethality between ATM and the FA pathway is due to reduced DNA resection and increased toxic NHEJ.

Cai et al., 2020, Cell Reports *30*, 2402–2415
February 18, 2020 © 2020 The Authors.
https://doi.org/10.1016/j.celrep.2020.01.052

CellPress

OPEN ACCESS
CellPress

Cell Reports
Article

Cooperation of the ATM and Fanconi Anemia/BRCA Pathways in Double-Strand Break End Resection

Mu-Yan Cai,[1,2] Connor E. Dunn,[1,3] Wenxu Chen,[1,2] Bose S. Kochupurakkal,[1,3] Huy Nguyen,[1,3] Lisa A. Moreau,[1] Geoffrey I. Shapiro,[3,4] Kalindi Parmar,[1,3] David Kozono,[1] and Alan D. D'Andrea[1,3,5,*]

[1]Department of Radiation Oncology, Dana-Farber Cancer Institute, Boston, MA 02215, USA
[2]Collaborative Innovation Center for Cancer Medicine, State Key Laboratory of Oncology in South China, Sun Yat-sen University Cancer Center, Guangzhou 510060, China
[3]Center for DNA Damage and Repair, Dana-Farber Cancer Institute, Boston, MA 02215, USA
[4]Early Drug Development Center, Dana-Farber Cancer Institute, Boston, MA 02215, USA
[5]Lead Contact
*Correspondence: alan_dandrea@dfci.harvard.edu
https://doi.org/10.1016/j.celrep.2020.01.052

SUMMARY

Cells deficient in ataxia telangiectasia mutated (ATM) are hypersensitive to ionizing radiation and other anti-cancer agents that induce double-strand DNA breaks. ATM inhibitors may therefore sensitize cancer cells to these agents. Some cancers may also have underlying genetic defects predisposing them to an ATM inhibitor monotherapy response. We have conducted a genome-wide CRISPR screen to identify genetic vulnerabilities that sensitize lung cancer cells to ATM inhibitors. Knockout of genes in the Fanconi anemia (FA)/BRCA pathway results in hypersensitivity to the ATM inhibitor M3541. Knockdown of either an FA gene or of ATM results in reduced double-strand break end resection, enhanced non-homologous end joining (NHEJ) repair, and decreased homologous recombination repair. Knockout of both the FA/BRCA pathway and ATM strongly inhibits end resection and generates toxic levels of NHEJ, thereby elucidating a mechanism of cellular death by synthetic lethality. ATM inhibitors may therefore be useful for the treatment of tumors with a defective FA/BRCA pathway.

INTRODUCTION

The ataxia telangiectasia mutated (ATM) kinase is a master regulator of the DNA damage response (DDR) (Shiloh and Ziv, 2013), and multiple ATM substrates have been identified. Mutations in the ATM gene result in hypersensitivity to ionizing radiation (IR) and to other DNA damaging agents. ATM coordinates the activity of two DNA double-strand break (DSB) repair pathways, namely, the homologous recombination (HR) pathway and the non-homologous end joining (NHEJ) pathway. Mutations in the ATM gene result in discordant activity of these pathways, persistent DSBs, and genomic instability. Consistent with these cellular phenotypes, humans or mice with biallelic ATM mutations are predisposed to cancers (Swift and Su, 1999).

Since ATM plays such a central role in the DDR, pharmacologic inhibition of ATM is an attractive opportunity for cancer therapy (Batey et al., 2013; Hickson et al., 2004). Nonspecific drugs, such as wortmannin, inhibit ATM and cause radiosensitization; however, this agent also targets other members of the PIKK (phosphatidylinositol-3-kinase-like kinase) family, including DNA-PK (DNA-dependent protein kinase) and ATR (ataxia telangiectasia and Rad3-related protein). KU-559333, another ATM inhibitor (ATMi) (Batey et al., 2013), does not target other PIKKs at effective concentrations; however, early clinical trials have not progressed due to off-target side effects. A recent compound, M3541, is a highly specific and potent inhibitor of ATM (Zimmermann et al., 2018), and a National Cancer Institute (NCI) clinical trial was recently initiated with the drug in combination with radiotherapy (NCT03225105).

Human cancers with an underlying defect in one DNA repair pathway may be hyperdependent on a second pathway. For example, BRCA1/2-deficient tumors with a defect in the HR pathway are hyperdependent on a polyADP ribose polymerase (PARP1)-dependent repair pathway. Accordingly, PARP1 inhibition has proven clinical efficacy in BRCA1/2 mutant ovarian and breast cancer (Bryant et al., 2005; Farmer et al., 2005), and four PARP1 inhibitors have now been granted limited FDA approval. Whether tumors with a deficiency of the ATM pathway also have specific synthetic lethal relationships with other DNA repair pathways remains largely unknown.

In the current study, we performed a genome-wide clustered regularly interspaced short palindromic repeats (CRISPR)-Cas9 loss-of-function genetic screen to identify enhancers of cell killing by the ATMi M3541 (Zimmermann et al., 2018). Notably, single guide RNAs (sgRNAs) directed against multiple genes in the Fanconi anemia (FA)/BRCA pathway (Ceccaldi et al., 2016b; Rodríguez and D'Andrea, 2017)—including FANCA, FANCF, and FANCG—and FA-like genes—including FAAP24, FAAP100, and FANCM—were synthetic lethal with ATM. sgRNAs directed against additional DNA repair genes including PTEN, Rad18, and STAG2 also conferred ATMi hypersensitivity. These genes are not FA genes per se, but their knockdown has previously been shown to confer DNA crosslinker sensitivity. Concurrent knockdown of both the FA/BRCA pathway and the ATM pathway resulted in reduced DSB end resection, hyperactivation of NHEJ repair, and enhanced chromosome aberrations, thereby

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

elucidating a mechanism of the synthetic lethality. Inhibition of NHEJ activity in tumor cells with a dual disruption of the FA/BRCA and ATM pathways restored cell survival. Taken together, our results suggest that the FA/BRCA pathway and the ATM pathway play redundant functions in enhancing DSB resection and suppressing NHEJ. We hypothesize that human tumors with an underlying defect in the FA/BRCA pathway may be hypersensitive to ATMis. Moreover, understanding the potential sensitivity and resistance mechanisms of ATMis may be useful in predicting patient responses, designing ATMi therapy combinations, and implementing mechanisms of re-sensitization.

RESULTS

Identification of sgRNAs That Selectively Enhance or Suppress Cell Killing by ATMi M3541

Initially, we evaluated A549 and NCI-H460 (H460) non-small cell lung cancer cell (NSCLC) lines, which stably express the Cas9 endonuclease. Using a clonogenic assay, we analyzed the cytotoxicity resulting from a combination of M3541—or of another ATMi, KU59333—with IR in these cells (Figures S1A and S1B). In accordance with previous studies, minimal toxicity was observed from ATMi exposure alone, with concentrations ranging up to 1.0 μM of drug. Enhanced cytotoxicity, however, was observed upon subsequent exposure to 2–6 Gy doses of IR. IR is known to rapidly activate the phosphorylation of ATM at Ser1981 (Bakkenist and Kastan, 2003). Low doses of M3541 or KU59333, in the micromolar range, resulted in a decrease in pS1981 over 2 h post-drug exposure in both the A549 and H460 cell lines (Figure S1C).

We next devised a genome-wide CRISPR screen to identify genes that, when disrupted, enhance or suppress cell killing by the ATMi M3541 (Figure 1). We reasoned that sgRNAs directed against specific genes would either enhance or suppress cell killing by M3541. The screen was conducted in the A549 and H460 cell lines, in the presence or absence of the ATMi M3541. The A549 and H460 cell lines are wild type (WT) for both ATM and the pro-apoptotic gene, p53.

In order to identify sgRNAs that target genes involved in both sensitivity and resistance to M3541, we chose a moderate concentration of the drug that inhibits the phosphorylation of ATM at Ser1981 at a low concentration (1 μM). Cells were infected with the Brunello v2 lentiviral sgRNA library and grown for 19 days in media treated with DMSO or the ATMi M3541 (Figure 1A). Fresh media, with freshly dissolved drug, were added every 3 days. Cells were collected pre- and post-drug selection, genomic DNA was extracted, and sgRNA sequences were PCR amplified and Illumina sequenced. Top hits among gene knockouts that conferred drug sensitivity or resistance were identified.

To identify highly significant hits, we utilized a STAR score at $\geq$4 when analyzing a single cell line. We included hits with a STAR score at $\geq$2 when analyzing the two cell lines together. The screen identified 62 genes in the A549 cell line and 38 genes in the H460 cell line that modulated cell killing by M3541. Out of these hits, we identified 30 genes that were hits in both cell lines. Many hits corresponded to specific signaling pathways (Figure 1B), and the top hits were DNA repair pathways. Gene Ontology analysis of 70 gene sets revealed strong enrichment for biological processes linked to the FA/BRCA pathway, a DNA crosslink repair pathway composed of 23 known FA genes (Ceccaldi et al., 2016b; Rodríguez and D'Andrea, 2017). Mapping the 70-gene set on the HumanMine protein-protein interaction database created a highly connected network consisting of multiple DNA damage repair genes, including most of the members of the FA/BRCA pathway (Figure 1C).

Many of the same genes that modulate M3541-mediated cell killing were identified in both lung cancer cell lines (see Figures 1D [scatterplot] and S1D [volcano plots] for the A549 and H460 cell lines). Top hits included FAAP24, FAAP100, FANCF, FANCE, FANCI, FANCA, FANCG, FANCB, FANCC, and FANCD2. Other relevant top hits included RAD18, an E3 ligase known to monoubiquitinate and activate FANCD2 (Williams et al., 2011); the tumor suppressor gene PTEN (McCabe et al., 2015); and the cohesin protein STAG2 (Watrin and Peters, 2009). These genes have also previously been shown to regulate DNA repair processes.

Several genes that regulate DNA DSB end resection were also hits in the screen. Knockdown of CtIP, a known phosphorylation target of ATM (Duquette et al., 2012), and knockdown of EXO1 were synthetic lethal with the ATMi (Figure 1D, lower left quadrant). While CtIP is not an FA gene, per se, it is known to bind to FANCD2 (Murina et al., 2014; Unno et al., 2014; Yeo et al., 2014). Knockdown of either FA/BRCA genes or of nucleases involved in DSB end resection, such as MRE11, conferred cellular hypersensitivity to the ATMi. Accordingly, our results implied that the FA/BRCA pathway and the ATM pathway may directly or indirectly regulate the DSB end resection through interactions with resection nucleases. Moreover, knockdowns of components of the ATM activating MRN complex (Wang et al., 2013), including Rad50 and NBS1, were also hits in the screen.

Interestingly, sgRNAs directed against components of the BRCA1-A complex (including BABAM1, BABAM2, and ABRAXAS1) were shown to increase the growth of both A549 and H460 cells grown in the presence of the ATMi (Figures 1D [upper right quadrant] and S1D). This functional interaction of ATM with subunits of the BRCA1-A complex has been recently demonstrated (Balmus et al., 2019). The BRCA1-A complex is known to suppress HR repair (Balmus et al., 2019; Harris and Khanna, 2011; Sobhian et al., 2007). The loss of this pathway may enhance the HR and increase resistance to ATM inhibition. This finding further supports an underlying role of ATM in enhancing HR repair. Finally, sgRNA directed against ATM itself protected cells treated with M3541. This result further validates the screen and provides further evidence that the ATMi is specific and is acting through its specific target. Loss of a drug target is a common mechanism of acquired drug resistance (Pettitt et al., 2018).

Concomitant Loss of the FA/BRCA Pathway and ATM Function Is Toxic to Cells

We previously demonstrated that a small interfering RNA (siRNA) directed against the FA protein FANCG is synthetically lethal with ATM loss (Kennedy et al., 2007). Whether ATM knockdown is also synthetically lethal with the loss of other genes in the FA/BRCA pathway is unknown. We next validated the CRISPR screen results by CRISPR knockout of individual FA genes and of FA-associated genes FAAP24 (Figures S2A and 2A), FAAP100 (Figures S2A and 2B), and FANCF (Figures S2A and 2C) in A549 and H460 cells. Protein knockdown by CRISPR

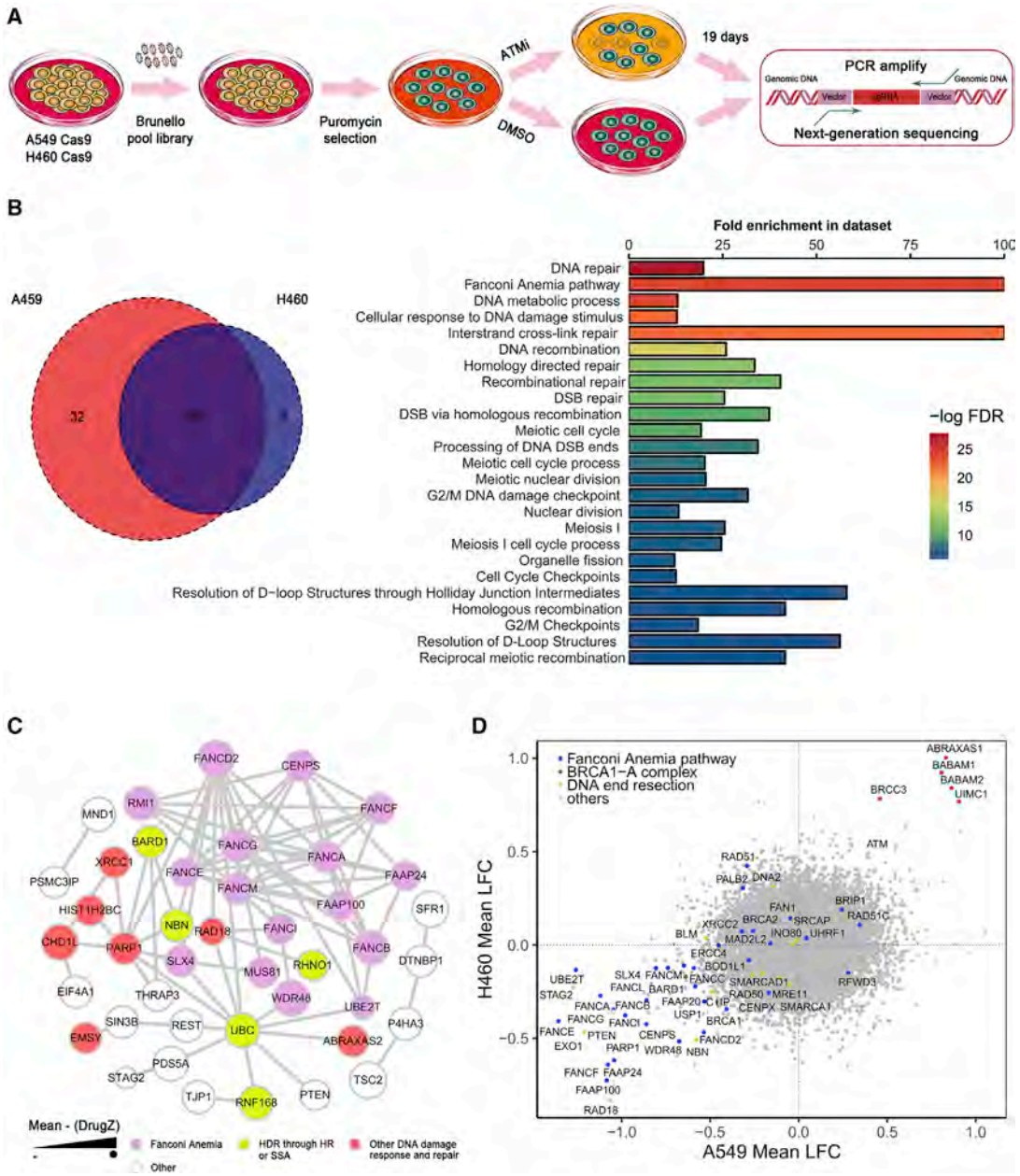


Figure 1. Whole Genome-wide CRISPR Screens Identify Determinants of ATMi Sensitivity

(A) Schematic of the genome-wide CRISPR-Cas9 screens in lung cancer lines.

(B) Venn diagram of all high-confidence hits in the two cell lines (left panel: STAR score at $\geq$ 4 in one line and STAR score at $\geq$ 2 in both lines). Gene Ontology (GO) terms significantly enriched in hits common to the two-cell lines (right panel).

(C) esyN network analysis of interactions between hits common to the two cell lines.

(D) Gene-centric visualization of average log2 fold change (LFC) in ATMi (M3541)- versus DMSO-treated in both cell lines.

See also Figure S1.

knockout was confirmed by immunoblotting with the corresponding antibodies. As expected, knockout of these individual genes resulted in enhanced sensitivity to the ATMi. We next validated synthetic lethality relationships between the ATM and these genes along with RAD18 and PTEN (Figures S2B and S2C). The siRNA-mediated knockdown of FAAP24, FAAP100, FANCF, RAD18, or PTEN in A549 and NCI-H460 cells resulted in cellular sensitivity to the ATMi, as predicted by the CRISPR screen results. We next determined whether the synthetic lethal relationship of FA genes and ATM could be extended to other cell types. Indeed, knockout of the FA genes FANCA and FANCD2 in HeLa and U2OS cells, respectively, conferred

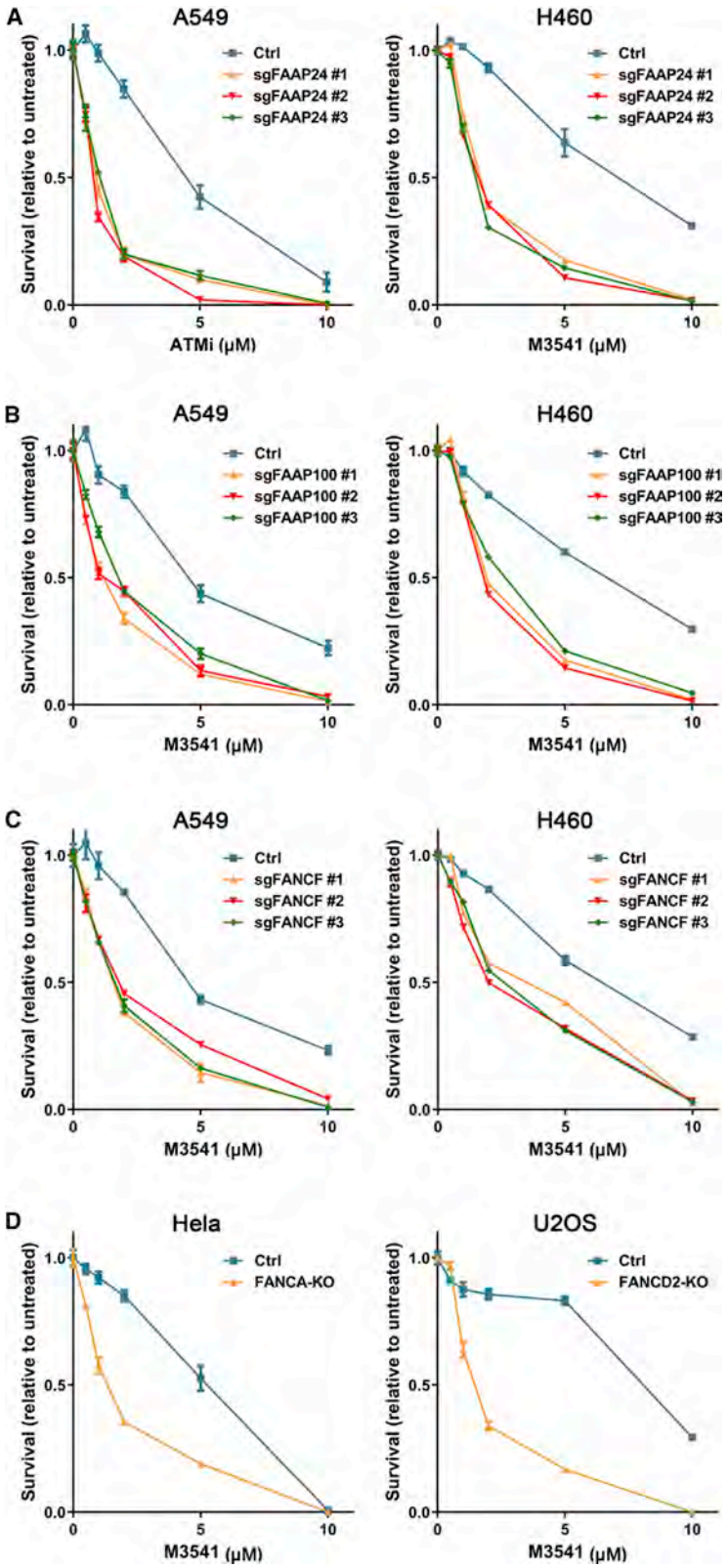


Figure 2. Loss of FA Pathway Confers Sensitivity to ATM Inhibition in Cancer Cells

(A–C) Clonogenic assays of control and knockout (KO) cell pools after treatment with ATM inhibitor (M3541) following transduction with FAAP24- (A), FAAP100- (B), and FANCF-targeting (C) sgRNAs (3 sgRNAs per gene). Survival assays in A549 cells (left panels) and H460 cells (right panels). #1, #2 and #3 indicate three independent sgRNAs for targeting each gene.

(D) Survival assays of HeLa and U2OS cells after treatment with M3541 in stable clones of FANCA and FANCD2 knockout via CRIPSR-Cas9. Clonogenic survival assays in HeLa cells (left panel) and U2OS cells (right panel). Data are shown as mean ± SEM from three independent experiments.

See also Figure S2.

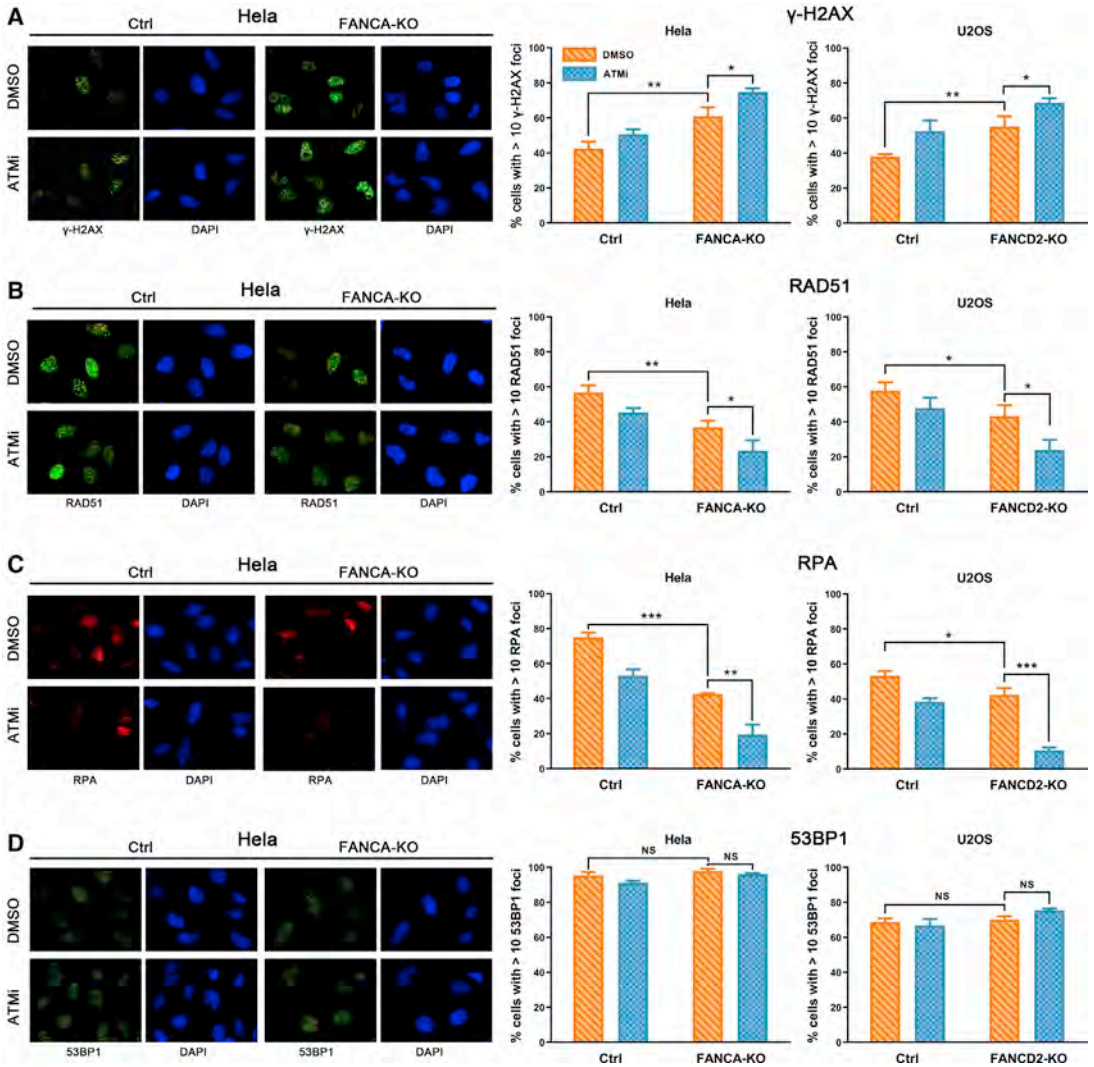


Figure 3. Inhibition of ATM Impairs RAD51 and RPA Focus Formation and Results in DNA Damage in FA-Deficient Cells

Immunofluorescence and quantification of the percentage of cells with more than 10 γ-H2AX (A), RAD51 (B), RPA (C), and 53BP1 (D) foci 3 h after treatment with irradiation (5 Gy) following 24 h exposure to DMSO or M3541 (1 μM) in control and FANCA-KO cells: representative images (left panels) and graphical quantitation of foci in HeLa cells (middle panels) and U2OS cells (right panels). One hundred nuclei were randomly counted. Data are shown as mean ± SEM from three independent experiments. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ on unpaired two-tailed Student's t test. See also Figure S3.

sensitivity to the ATMi, further validating the CRISPR screen results and extending the data (Figures S2A and 2D).

Mechanism of Synthetic Lethality between the FA Pathway and ATM Pathway

To determine the mechanism of synthetic lethality in cells with a dual knockdown of the FA pathway and ATM, we examined various DNA repair biomarkers (Figure 3). Initially, we examined HeLa cells or U2OS cells that have a CRISPR-engineered knockout of either FANCA or FANCD2, respectively. Knockout of FANCA or FANCD2 in either cell line resulted in increased DNA DSB, as measured by IR-induced H2AX phosphorylation (Figure 3A). Breakage was enhanced by exposure to ATMi, perhaps resulting from slower DNA repair kinetics. In a comet assay, we also observed longer tails, reflecting increased DNA damage following treatment with the ATMi in the FA-deficient HeLa and U2OS cell lines (Figures S3A and S3B).

We next determined whether the loss of a gene in the FA pathway could confer a reduction in HR repair in these cell lines. Previous studies have indicated that FA gene knockdown can result in a mild reduction in both HR activity and single-strand annealing (SSA) (Nakanishi et al., 2005). Indeed, FANCA or FANCD2 knockout in either HeLa cells or U2OS cells resulted in decreased IR-induced RAD51 foci (Figure 3B). The level of RAD51 focus formation is a measure of BRCA2/RAD51 loading and HR competence. Interestingly, exposure of FANCA-deficient or FANCD2-deficient cells to the ATMi M3541 resulted in a further loss of RAD51 foci (Figure 3B). Taken together, these

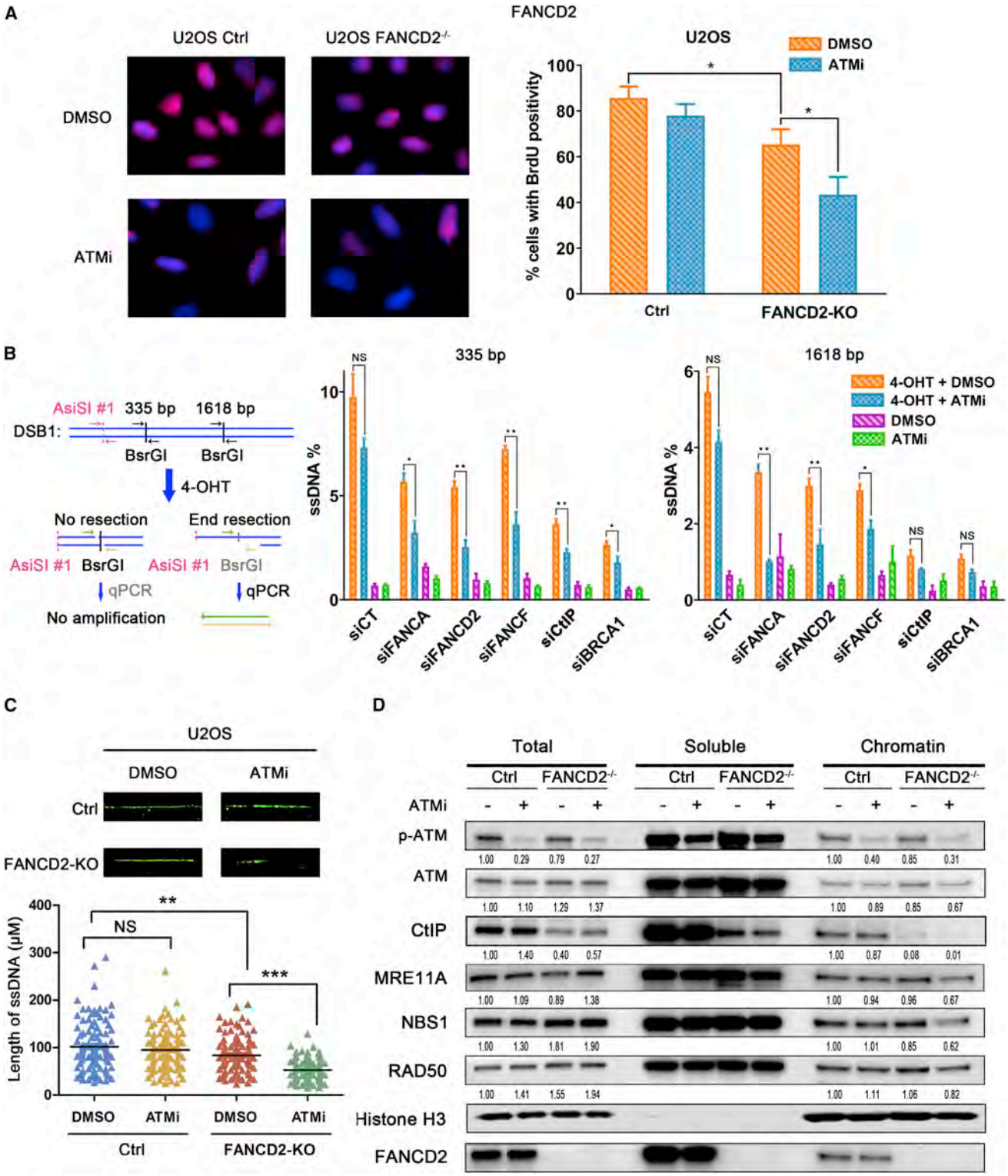


Figure 4. Inhibition of ATM Decreases DNA End Resection Machinery in FA-Pathway-Deficient Cells

(A) Immunofluorescence and quantification of the percentage of cells with BrdU-positive cells 6 h after treatment with irradiation (5 Gy) following 24 h exposure to DMSO or M3541 (1 μM) in control and FANCD2-KO U2OS cells: representative images (left panel) and graphical quantitation of BrDU-positive cells (right panel). One hundred nuclei were randomly counted. *$p < 0.05$, unpaired two-tailed Student's t test. Data are shown as mean ± SD from three independent experiments.

(legend continued on next page)

results indicate that the FA pathway and the ATM pathway each contribute positively to HR repair. We hypothesized that the loss of both pathways would cause a profound loss of HR repair, further contributing to the lethal phenotypes observed when both pathways are compromised.

Knockdown of FANCA or FANCD2 resulted in decreased DNA DSB resection and decreased single-stranded DNA (ssDNA) generation, as demonstrated by decreased IR-induced replication factor A (RPA) foci (Figure 3C). The decrease in RPA foci following FA gene knockdown suggests that the FA pathway—through its interaction with various resection proteins, such as CtIP and EXO1—can positively regulate the process of end resection and HR repair. In contrast, changes in IR-induced 53BP1 foci were not observed when comparing vehicle control and ATMi treated FANCA$^{-/-}$ or FANCD2$^{-/-}$ cells (Figure 3D).

FANCA-deficient cells exhibited enhanced ATM phosphorylation (Figure S3C), consistent with the synthetic lethality observed when both pathways are disrupted. This increased pATM level suggests that in the absence of the FA/BRCA pathway, cells are hyperdependent on activation of the ATM pathway for survival. Similarly, the loss of ATM resulted in hyperactivation of the FA/BRCA pathway, as detected by elevated FANCD2 monoubiquitination (Figure S3D). Again, these results further support a synthetic lethal relationship of these two pathways. When one pathway is knocked down, the other pathway is upregulated.

Knockout of the FA and ATM Pathways Reduces DSB End Resection, Increases End Joining, and Increases NHEJ

Recent studies demonstrate that the cellular choice between HR repair and NHEJ repair of DSBs is determined by the process of DSB end resection (Ceccaldi et al., 2016a; Symington, 2016). Increased resection favors HR, while the blockage of resection favors NHEJ. We next evaluated the role of the FA and ATM pathways in end resection by using several independent assays.

End resection of a DSB by nucleolytic degradation of the 5′ strand is the first step in initiating the HR pathway. The exposed 3′ single-strand DNA binds to RPA, which in turn is displaced by the recombinase, RAD51. Previous studies have shown that the ATM kinase can promote DSB end resection by phosphorylating and activating the protein CtIP (Duquette et al., 2012). Other studies have independently shown that monoubiquitinated FANCD2, the downstream regulator in the FA pathway, can recruit CtIP to chromatin and further promote DSB end resection (Murina et al., 2014; Unno et al., 2014; Yeo et al., 2014).

To directly determine the role of FA and ATM in DSB resection, we labeled genomic DNA with 5-bromo-2′-deoxyuridine (BrdU) and used an antibody that detects ssDNA resulting from end resection. Knockout of FANCD2 decreased BrdU-positive cells, and inhibition of ATM with M3541 further impaired IR-induced BrdU focus formation in FANCD2$^{-/-}$ cells (Figure 4A).

To further quantify the extent of resection, we introduced the AsiSI enzyme fused with a hormone-binding domain in U2OS cells with either FANCD2 or FANCF knockdown. A quantitative PCR (qPCR)-based method was used to examine ssDNA (Figure 4B). A decrease in the amount of ssDNA was found after the depletion of FANCF or of FANCD2. A further decrease in the level of ssDNA was observed following the depletion of both the FA/BRCA pathway and the ATM pathway.

Next, we performed a SMART (single-molecule analysis of resection tracks) assay to further evaluate the effect of FA pathway disruption and/or ATM inhibition on DSB end resection (Figures 4C and S4A). This assay quantifies the amount of ssDNA generated by end resection at a defined DSB (Huertas and Cruz-García, 2018). Disruption of the FANCD2 gene resulted in a minor reduction in DSB end resection, as did exposure to the ATMi. Interestingly, dual disruption of the FA/BRCA and ATM pathways resulted in a further decrease in DSB end resection, consistent with the enhancement of toxic NHEJ activity observed when both pathways are knocked down. Moreover, complementation with FANCD2 in FANCD2-KO cells rescued this phenotype (Figure S4B). Consistent with previous studies, these results indicate that the FA/BRCA pathway normally promotes DSB end resection, leading to a suppression of NHEJ. We next determined whether the FA pathway is required for the recruitment of CtIP to chromatin and whether ATM further promotes this recruitment (Figures 4D and S4C). Interestingly, knockout of FANCD2 (Figure 4D) or of FANCA (Figure S4C) reduced CtIP expression (lanes 3 and 4) and chromatin recruitment of CtIP (lanes 11 and 12), and ATMi further reduced CtIP recruitment. CtIP knockdown resulted in reduced survival of WT U2OS cells. CtIP knockdown resulted in increased toxicity, however, in FANCD2 knockout U2OS cells, suggesting a synthetic lethality between CtIP and FANCD2. Overexpression of CtIP rescued the lethality in both the WT and FANCD2 knockout U2OS cells (Figure S4D). Taken together, these results further demonstrate that the FA and ATM pathways cooperate in enhancing CtIP mediated end resection.

Knockdown of the FA Pathway Results in Increased NHEJ Activity and Decreased HR Activity

Previous studies have suggested that NHEJ is upregulated in cells deficient in the FA pathway (Adamo et al., 2010; Eccles et al., 2018; Pace et al., 2010). Knockdown of components of the NHEJ pathway, such as DNA-PK (Adamo et al., 2010) or DNA ligase 4 (Maruyama et al., 2015), reduces the interstrand crosslinker (ICL) sensitivity of FA cells. Recent studies also indicate that NHEJ is upregulated in ATM-deficient cells (Balmus et al., 2019), suggesting that a principal cellular function of

(B) The schematic of *in vitro* end-resection assay in U2OS cells (left panel). ssDNA was quantified by qPCR at 335 bp or 1,618 bp downstream of the AsiSI-induced break site after indicated siRNA transfection and DSB induction (with 4-hydroxytamoxifen) or no DSB induction (without 4-hydroxytamoxifen) (right panel). Data are shown as mean ± SEM compiled from three independent experiments. *$p < 0.05$, **$p < 0.01$.

(C) Change in ssDNA length of control and FANCD2$^{-/-}$ U2OS cells 3 h after treatment with irradiation (5 Gy) following overnight exposure to 1 μM ATMi. Representative images of SMART assay (top panel) and the quantification (relative to untreated, bottom panel). In total, 100 DNA fibers were assessed per condition. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, unpaired two-tailed Student's t test. NS, not significant (bottom panel).

(D) Immunoblots of the indicated proteins in subcellular fractions of control and FANCD2$^{-/-}$ U2OS cells after treatment with 1 μM ATMi for 6 h. Quantification of the ratio of total and chromatin bound CtIP in control versus FANCD2$^{-/-}$ U2OS cells is shown. See also Figure S4.

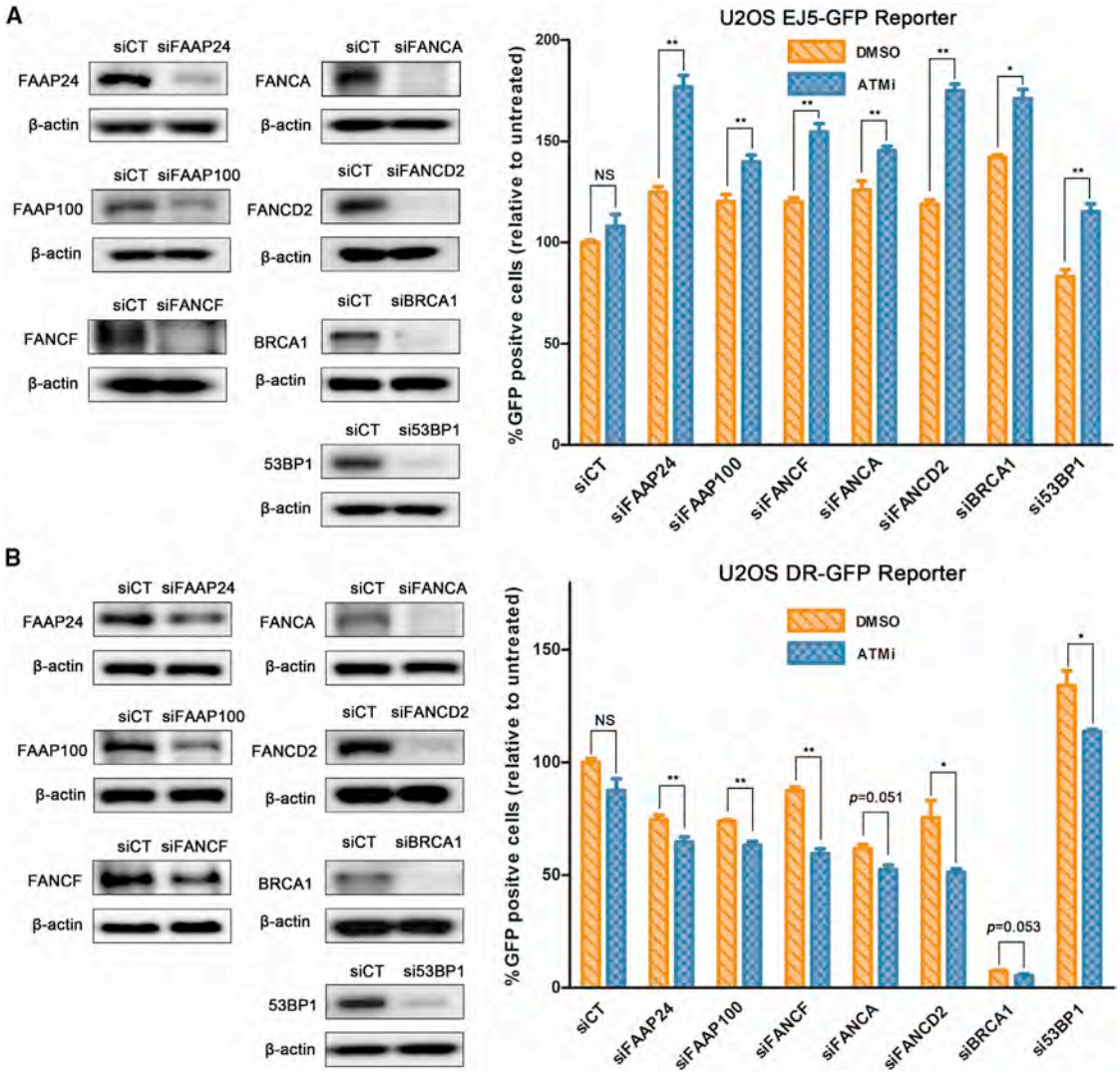


Figure 5. Inhibition of ATM Results in Increased NHEJ and Decreased HR in FA-Pathway-Deficient Cells

U2OS cells stably carrying the EJ5-GFP (A) and DR-GFP (B) reporter were transfected with the indicated siRNAs and were treated with 1 μM ATMi overnight 36 h after siRNA transfection. Subsequently, the cells were exposed to I-SceI adenovirus for 6 h and then cultured with fresh medium. GFP-positive cells were quantified 72 h later by FACS. siRNA knockdown efficiency was assessed by western blotting assay (left panels). Quantitation for the HR activity and NHEJ activity as % of GFP positive cells is shown (right panels). Data are shown as mean ± SEM from three independent experiments. *$p < 0.05$ and **$p < 0.01$ on unpaired two-way ANOVA test. See also Figure S5.

ATM is to suppress the NHEJ pathway and to promote DSB end resection and HR. Knockdown of the HR-suppressing BRCA1-A complex can enhance resistance to ATMi (Balmus et al., 2019).

Based on these studies and our evidence of synthetic lethality between FA and ATM, we reasoned that the mechanism of synthetic lethality is a decrease in DSB end resection, an increase in end joining, and an increase in toxic NHEJ following FA and ATM knockdown. Previous studies have indicated that either the FA pathway or ATM is required to remove NHEJ proteins, such as 53BP1, from DSBs that have been generated in the process of DNA crosslink repair or by nuclease processing of reversed forks. We reasoned that loss of both the FA pathway and ATM would further inhibit DSB resection, increase end joining, and further upregulate NHEJ.

To further examine whether the ATMi promotes NHEJ, we used an NHEJ reporter template (Figure 5; Howard et al., 2015). Knockdown of components of the FA/BRCA pathway, including BRCA1, FAAP24, or FANCD2, in U2OS cells increased NHEJ activity. Efficient knockdown of the FA proteins was confirmed by immunoblot. Knockdown of FA protein expression resulted in an enhancement of NHEJ activity. Cellular exposure to the ATMi further increased NHEJ activity in these cells (Figure 5A).

Interestingly, knockdown of the FA pathway and ATM resulted in a decrease in HR, as measured by the DR-GFP assay (Figure 5B). As further evidence to support this mechanism, knockout of FANCA in HeLa cells (Figure S5A), or knockout of FANCD2 in U2OS cells (Figure S5B), resulted in increased phosphorylation of DNA-PK (pDNA-PK), a biomarker of the activated

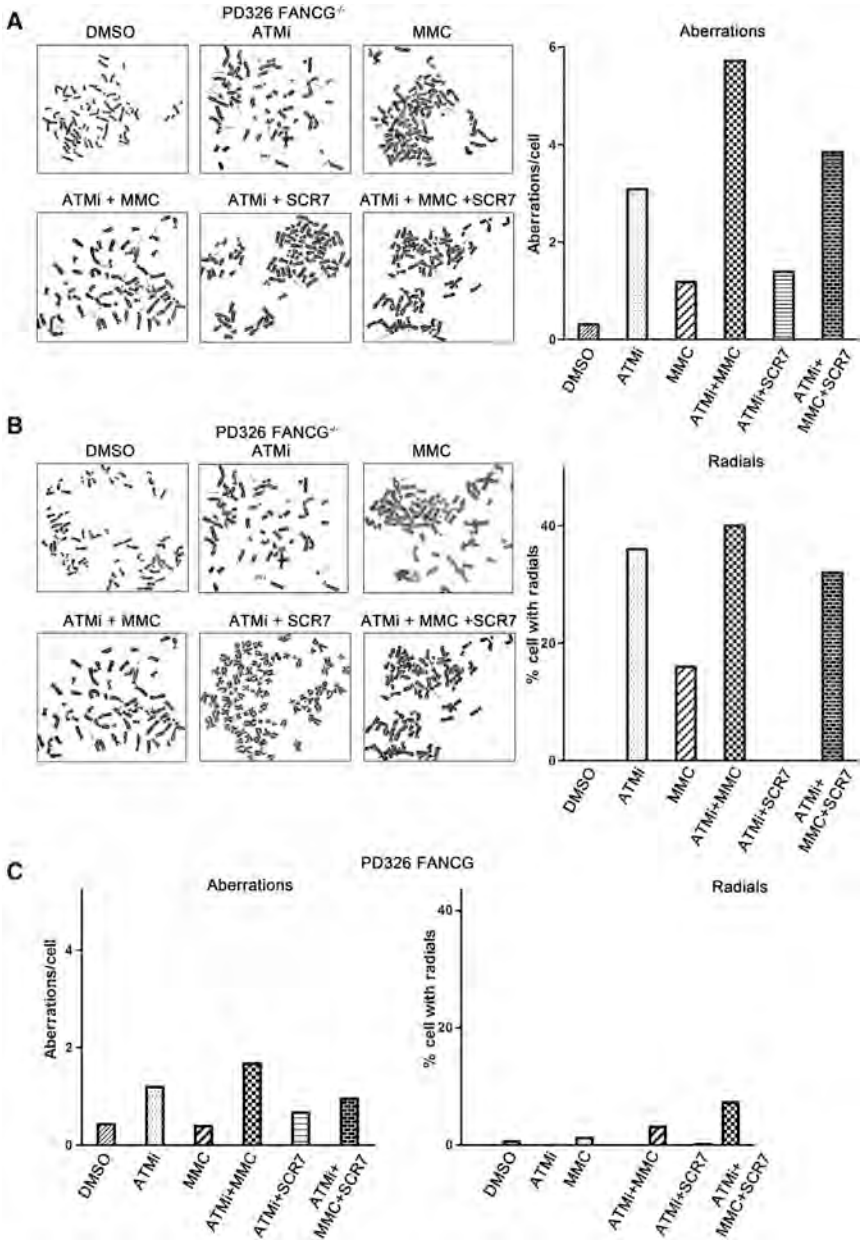


Figure 6. Inhibition of ATM Results in Chromosomal Aberrations and Radials in FA-Pathway-Deficient Cells

(A and B) Representative images (left panels) and quantification (right panels) of chromosomal aberrations (A) and radial chromosomes (B, indicated by arrows) per cell on metaphase spreads of PD326 (FANCG-deficient) cells 48 h after treatment with 5 ng/ml Mitomycin C (MMC) following 24 h exposure to 1 μM DMSO, 1 μM ATMi, or 1 μM ATMi plus 5 μM Lig4 inhibitor (SCR7). At least 50 cells were counted in each experiment, and the average of 50 cells is shown.

(C) Quantification of chromosomal aberrations (left panel) and radial chromosomes (right panel) per cell as a measure on metaphase spreads 48 h after treatment with 5 ng/ml MMC following 24 h exposure to 1 μM DMSO, 1 μM ATMi, or 1 μM ATMi plus 5 μM Lig4 inhibitor (SCR7) in PD326 FANCG-corrected cells. See also Figure S6.

NHEJ pathway. The addition of ATMi further increased pDNA-PK and DNA damage, as measured by pH2AX (Figures S5A and S5B). Taken together, these results indicate that the FA pathway and the ATM pathway are non-epistatic, and each suppresses NHEJ activity. Knocking down both pathways resulted in excessive end-joining activity and cellular death, providing additional mechanistic evidence for synthetic lethality.

ATM Is Required for Homologous Recombination Repair in FA/BRCA-Pathway-Deficient Cells

FA-pathway-deficient cells and ATM-deficient cells have a common cellular phenotype, namely, the accumulation of chromosome radial forms. Indeed, the diagnostic test for FA is the DEB (diepoxybutane) or mitomycin C (MMC)-induced increase in chromosome radials. Chromosome radials result from hyperactive non-homologous pairing of chromosomes, indirectly resulting from increased NHEJ activity (Newell et al., 2004). BRCA1-deficient cells exposed to PARP inhibitors exhibit elevated radial chromosomes (Bunting and Nussenzweig, 2010). Knockout of multiple components of the NHEJ pathway, such as the 53BP1 or SHLD complex subunits, in BRCA1-deficient cells results in a reduction of radial chromosomes and in enhanced cell survival (Mirman et al., 2018).

We next determined whether a disruption of ATM activity would further increase NHEJ activity, chromosome aberrations, and radial formation in FA cells (Figure 6). For this purpose, we examined the effect of ATMi exposure on the FANCG-deficient lymphoblast line, PD326. M3541 increased chromosome radials in PD326 cells, consistent with an enhancement of toxic NHEJ activity. ATMi plus the crosslinker, MMC, further enhanced chromosome radial production. Interestingly, an inhibitor of DNA ligase IV, a downstream enzyme in the NHEJ pathway, reduced chromosome radials, further indicating that the increased radial production resulted from decreased end resection and increased end joining activity (Figures 6A and 6B). As a control, ATMi or MMC exposure did not induce chromosome aberrations or radials in the corrected isogenic cell line, PD326 plus FANCG (Figure 6C). The inhibitor of DNA ligase IV also enhanced clonogenic survival of other FA-pathway-deficient cells, including the FANCA-deficient and FANCD2-deficient cells, when they were exposed to ATMi (Figures S6A and S6B).

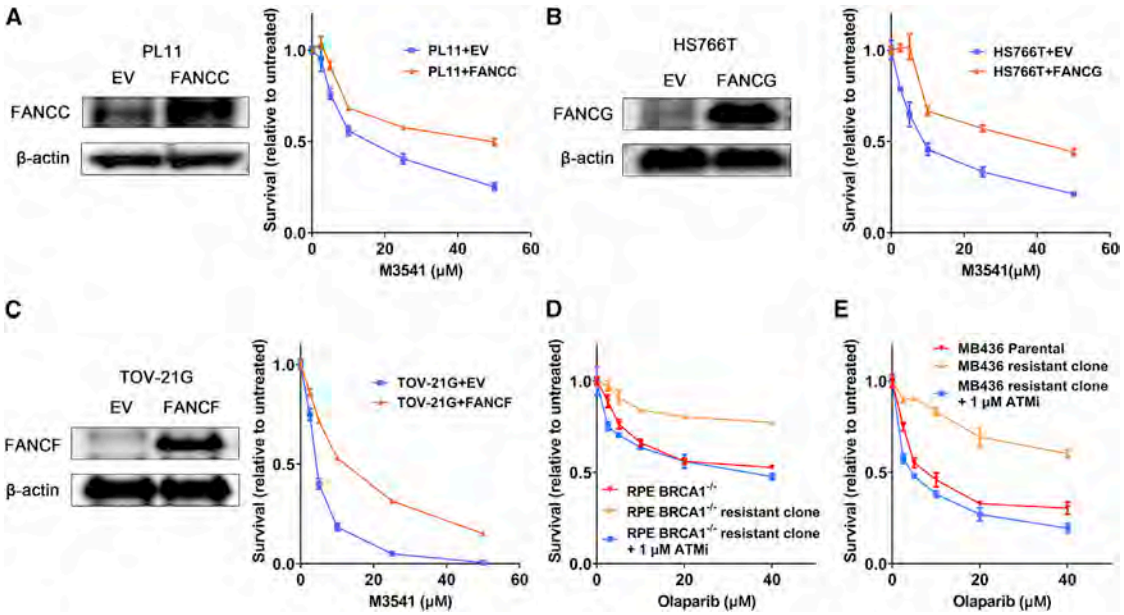


Figure 7. ATM Inhibition Shows Greater Cytotoxicity with FA Pathway Deficiency

(A) Western blot showing FANCC expression (left panel) and quantitation of CellTiter-Glo cell viability assay (right panel) in FANCC mutant PL11 cells and the isogenic FANCC-corrected cells 72 h after M3541 treatment.

(B) Western blot showing FANCG expression (left panel) and quantitation of CellTiter-Glo cell viability assay (right panel) in FANCG mutant HS766T cells and the isogenic FANCG-corrected cells 72 h after M3541 treatment.

(C) Western blot showing FANCF expression (left panel) and quantitation of CellTiter-Glo cell viability assay (right panel) in FANCF mutant TOV-21G cells and the isogenic FANCF-corrected cells 72 h after M3541 treatment.

(D and E) Quantitation of the CellTiter-Glo cell viability assay in BRCA1-KO RPE (D) and MB436 (E) (parental and olaparib-resistant) cells exposed to graded concentrations of the PARP inhibitor olaparib in the presence or absence of the ATMi M3541 (1 μM). See also Figure S7.

Synthetic Lethality of FA Pathway and ATM in Human Cancer Cell Lines

FA pathway disruption is observed in many human cancers, including pancreatic cancer (van der Heijden et al., 2005) and ovarian cancer (Taniguchi et al., 2003). The FA-pathway-deficient pancreatic cancer cell lines PL11 and HS766T were sensitive to the ATMi. Decreased expression of the FANCC protein in PL11 cells and decreased expression of the FANCG protein in the HS766T cell line were confirmed by immunoblotting with the corresponding antibodies. The cDNA-complemented versions of these cell lines exhibited decreased sensitivity to the ATMi (Figures 7A and 7B). The TOV21G cell line is an ovarian cancer line exhibiting methylation and silencing of the FANCF gene (Taniguchi et al., 2003). An isogenic line TOV21G+FANCF has an intact FA pathway. As predicted, the TOV21G cells were more sensitive to ATMi exposure (Figure 7C) than were the complemented cells. Finally, BRCA1-deficient cells that had acquired resistance to the PARP inhibitor, olaparib, were sensitized by the addition of the ATMi (Figures 7D and 7E).

Validation of Synthetic Lethality of FA and ATM Pathways Using Computational Approaches

Analysis of publicly available tumor and cell line databases was conducted to further validate the synthetic lethal relationship between the FA and ATM pathways. We initially focused on genes that, when lost or mutated, sensitized cells to the ATMi (M3541) in our CRISPR screen. The gene hits with an average $-\log_{10}(P) < 2.5$ and an average $\log_2$ fold change of at least -0.5, in both A549 and H460 cells, were RAD18, FAAP24, FAAP100, and FANCF (Figure S1D). Analysis of a cancer-dependency study revealed that cell lines with copy number loss of any one of these four genes, but not cell lines with normal copy number of these genes, are hyperdependent on ATM for survival. (Figure S7A). Moreover, cell lines with mutations in genes encoding proteins in the FA core complex proteins are more dependent on ATM for survival than are the cell lines that are the WT for these genes (Figure S7B).

MMC is a chemotherapeutic agent known to cause DNA crosslinks. Crosslinks are repaired, in part, by the FA pathway. Cells lacking a functional FA pathway are deficient in crosslink repair and are hypersensitive to MMC. As described above, cells lacking the FA pathway are also sensitive to ATMi. We therefore investigated a possible correlation between cellular sensitivity to MMC and ATMi using a drug sensitivity database. Indeed, the cell lines that are sensitive to MMC tend to be more sensitive to the ATMi KU-55933, corroborating the notion that a loss of the FA pathway results in sensitivity to both an ATMi and MMC (Figure S7C).

The synthetic lethal relationship between the FA pathway and ATM described here suggests that cells with defects in both the FA pathway and ATM may not be viable. Tumors with co-occurrence of FA and ATM mutations are therefore hypothesized to

have reduced viability. In order to test this hypothesis, we analyzed clinical data from the PanCancer MSK-IMPACT Clinical Sequencing Cohort. Interestingly, the patients with tumors harboring mutations in both the FA pathway and ATM showed significantly improved survival by a log-rank test, compared to the patients with tumors that are WT for these genes (Figure S7D). Taken together, computational analysis of publicly available databases supports the synthetic lethality between the FA pathway and ATMi observed in the CRISPR screen.

DISCUSSION

Cancer cells have multiple DNA repair pathways for coping with intrinsic DNA damage and replication stress. These cells often lose one or more of these pathways through mutation, deletion, or epigenetic silencing, resulting in genomic instability, a mutator phenotype, and a more rapidly evolving malignancy. Cells with one dysfunctional DNA repair pathway may become hyperdependent on an alternative pathway for survival. For instance, tumor cells with a loss of BRCA1 and BRCA2, required for HR repair, are hyperdependent on PARP1-dependent DNA repair mechanisms such as base excision repair (BER) (Lord and Ashworth, 2016) or alternative end joining (Alt-EJ) (Ceccaldi et al., 2015; Mateos-Gomez et al., 2015). Accordingly, inhibitors of PARP or other components of the Alt-EJ pathway, such as POLQ (Higgins and Boulton, 2018), may be cytotoxic to HR-deficient tumor cells.

In the current study, we sought to identify genetic vulnerabilities in cancer cell lines that are predisposed to cell death by ATMi monotherapy. We demonstrated that a disruption of the FA/BRCA pathway enhances cell killing by M3541. Further disruption of the terminal components of cNHEJ can reduce the cell killing. Moreover, we demonstrate that one mechanism of the synthetic lethality resulting from a dual disruption of ATM and FA pathway genes is elevated (toxic) cNHEJ activity and decreased HR activity.

Previous studies have suggested that elevated NHEJ activity may account for the cellular features of FA cells. FA cells have elevated ICL sensitivity and an elevation in ICL-induced chromosome radial forms. These radial forms result from enhanced non-homologous pairing of chromosomes, as evident in metaphase chromosome spreads (Newell et al., 2004). Previous studies have also indicated that FA proteins or ATM can function, at least in part, by excluding or removing NHEJ proteins, such as 53BP1, from the sites of DSBs. These breaks may occur during ICL repair or during nuclease processing of reversed replication forks. Loss of both the FA pathway and ATM results in further retention of NHEJ factors at DSBs and persistence of NHEJ activity, extending even into the S phase of the cell cycle. This toxic NHEJ can cause the inappropriate fusion of DNA breaks on different chromosomes. Indeed, the radial chromosomes resulting from fused segments of non-homologous chromosomes resemble the genomic rearrangements found in primary cells from FA patients. Interestingly, this increase in non-homologous pairing correlates with increased NHEJ activity (Adamo et al., 2010; Eccles et al., 2018; Pace et al., 2010). Consistent with this model, a knockdown of NHEJ activity, resulting from cellular exposure to a DNA-PK inhibitor or inhibition of ligase 4, can reverse this phenotype and reduce the ICL sensitivity of FA cells.

ATM is known to promote DSB end resection and suppress NHEJ. The enhancement of DSB end resection results, at least in part, from the ability of ATM to inactivate the 53BP1/RIF1/REV7/SHLD complex, thereby enabling DSB end resection, RAD51 loading, and homologous recombination. Our results indicate that the FA pathway also promotes DSB end resection, although the mechanism of enhanced resection remains unknown. Monoubiquitinated FANCD2, a downstream product of the FA pathway, may bind and stimulate the end resection activity of CtIP. Recent studies also indicate that the FANCA protein can promote the process of SSA (Benitez et al., 2018) and may thereby positively regulate end resection. Accordingly, inactivation of both the ATM and FA pathways appears to further reduce DSB end resection. Inactivation of both pathways simultaneously results in a further decrease in end resection, accounting for the enhancement of NHEJ activity and further contributing to synthetic lethality.

Our results indicate that ATM and the FA pathway cooperate in the regulation of DSB end resection. We previously demonstrated a convergence of this pathway in DNA crosslink repair (Taniguchi et al., 2002). The current study supports a model for the regulation of DSB end resection (Figure S7E). The FA pathway promotes the monoubiquitination and chromatin recruitment of FANCD2. FANCD2-Ub binds to and promotes the chromatin recruitment of CtIP to the site of DSB end resection. ATM further phosphorylates and activates CtIP. Activated CtIP next promotes DSB end resection and suppresses NHEJ. This model provides one of the mechanisms for the synthetic lethal relationship of FA genes and ATM. Loss of both pathways leads to a profound reduction in DSB end resection, leading to toxic NHEJ, as indicated by radial chromosome accumulation.

Our results also suggest that tumor cells may become resistant to an ATMi by finding an alternate mechanism for enhancing DSB end resection, such as by knocking down the BRCA1-A complex or by reducing NHEJ levels. FA cells corresponding to some FA complementation groups also have reduced HR activity and increased PARPi sensitivity. In our study, simultaneous disruption of the FA pathway and ATM resulted in a profound decrease in RAD51 focus formation and cell death, thereby providing a second mechanism of synthetic lethality. Interestingly, the loss of the BRCA1-A complex, a complex known to suppress HR repair, can rescue RAD51 foci, restore HR repair, and rescue cell viability, based on our CRISPR screen results. Accordingly, our CRISPR screen identified components of the BRCA1-A complex, which, when knocked out, enhanced the viability of A549 or H460 cells exposed to the ATMi.

In addition to bona fide members of the FA/BRCA pathway, we identified other genes whose loss is synthetically lethal with ATMi. RAD18 is a ubiquitin E3 ligase known to monoubiquitinate PCNA and to activate TLS (translesion synthesis) repair. Interestingly, RAD18 may also play a role in FA/BRCA pathway function. In one study, RAD18 was shown to promote the FA core complex-dependent monoubiquitination of FANCD2 in a manner that is secondary to its ability to monoubiquitinate PCNA (Song et al., 2010). Moreover, RAD18 was shown to bind to FANCD2 and to promote the monoubiquitination of the FANCD2/FANCI complex (Williams et al., 2011).

The tumor suppressor protein PTEN has also been reported to have a role in the DDR (Gupta et al., 2009; Shen et al., 2007). Loss of PTEN function underlies some cancers, such as prostate cancer and glioblastoma (Li et al., 1997). In addition to its cytoplasmic role in the PI3K pathway, PTEN functions as a nuclear protein and appears to have a direct function in DNA repair. Cells with a deficiency in nuclear targeting of PTEN are defective in HR repair and, like FA cells, are sensitive to DNA crosslinking agents such as cisplatin (Bassi et al., 2013). A recent report further supports a model in which PTEN loss is synthetically lethal with loss of the ATM kinase (McCabe et al., 2015).

STAG2 is a core component of the cohesin complex, which also includes SMC1, SMC2, and Rad21 (Watrin and Peters, 2009). The cohesin complex is required for optimal DNA repair, including DSB repair, in human cells (Terret et al., 2009; Tittel-Elmer et al., 2012). The cohesin complex is known to organize chromatin loops at DNA replication factories, to mediate replication stress tolerance, and to enhance the restart of stalled forks (Wu and Yu, 2012). Somatic mutation of STAG2 in some cancers accounts for crosslinker-dependent chromosome breakage and cisplatin sensitivity, a phenotype similar to cancers with FA/BRCA pathway deficiency (Stoepker et al., 2015). Interestingly, STAG2 also plays a role in replication fork progression and has a synthetic lethal relationship with PARP and ATR loss in DSB repair (Mondal et al., 2019). Taken together, these results are consistent with the synthetic lethal relationship of STAG2 and ATM described herein.

In summary, tumor cells that have an underlying genetic or epigenetic defect in the FA pathway—or in RAD18, PTEN, or STAG2—may be hypersensitive to the ATMi M3541. Accordingly, patients bearing tumors with these defects may be more likely to respond to the ATMi, and these patients may be more likely to benefit from it.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHODS DETAILS
  - Drug selection CRISPR screens
  - Genomic DNA preparation, sequencing and identification of top hits
  - Western blotting
  - CRISPR cas9 Knockout
  - siRNA transfection
  - Clonogenic assay
  - CellTiter-Glo assay
  - Subcellular Fractionation
  - Comet assays
  - Reporter assays
  - Immunofluorescence
  - Cytogenetics
  - SMART assay
  - *In vitro* DNA end resection
  - Computational Analysis
- QUANTITATION AND STATISTICAL ANALYSIS
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2020.01.052.

ACKNOWLEDGMENTS

We thank Dr. Gaëlle Legube for generously providing the HA-ER-AsiSI system. We thank members of the D'Andrea laboratory for their helpful suggestions and comments. This research was supported by grants from the U.S. NIH (R01HL052725, P01HL048546), the U.S. Department of Defense (BM110181, BC151331P1), the Breast Cancer Research Foundation, and the Fanconi Anemia Research Fund (to A.D.D.). This work was also supported by Sponsored Research Agreement from EMD Serono (to A.D.D. and G.I.S.) and the grants from the National Nature Science Foundation of China (81672407) and Excellent Young Talents in Sun Yat-sen University Cancer Center (to M.-Y.C.).

AUTHOR CONTRIBUTIONS

M.-Y.C. conceived the study, performed experiments, analyzed data, and wrote the manuscript. C.E.D. performed the experiments and computational analysis, analyzed the data, and helped write the manuscript. W.C. and L.A.M. performed the experiments and analyzed the data. B.S.K. performed the computational analysis, analyzed the data, and helped write the manuscript. H.N. performed the computational analysis and analyzed the data. G.I.S. provided critical advice on the study. K.P. contributed to writing the manuscript. D.K. helped with the design of the CRISPR screen and performed bioinformatics analysis. A.D.D. conceived the study and wrote the manuscript.

DECLARATION OF INTERESTS

A.D.D. is a consultant/advisory board member for Lilly Oncology, Merck-EMD Serono, Intellia Therapeutics, Sierra Oncology, Cyteir Therapeutics, Third Rock Ventures, AstraZeneca, Ideaya Inc., and Cedilla Therapeutics Inc.; a stockholder in Ideaya Inc., Cedilla Therapeutics Inc., and Cyteir; and reports receiving commercial research grants from Lilly Oncology and Merck-EMD Serono. G.I.S. is a consultant/advisory board member for Lilly, Sierra Oncology, Merck-EMD Serono, Pfizer, Astex, Almac, Roche, Bicycle Therapeutics, Fusion Pharmaceuticals, G1 Therapeutics, Bayer, Ipsen, Cybrexa Therapeutics, Angiex, Daiichi Sankyo, and Seattle Genetics; and reports receiving commercial research grants from Lilly, Sierra Oncology, Merck-EMD Serono, and Merck & Co.

Received: September 13, 2019
Revised: December 11, 2019
Accepted: January 15, 2020
Published: February 18, 2020

REFERENCES

Adamo, A., Collis, S.J., Adelman, C.A., Silva, N., Horejsi, Z., Ward, J.D., Martinez-Perez, E., Boulton, S.J., and La Volpe, A. (2010). Preventing nonhomologous end joining suppresses DNA repair defects of Fanconi anemia. Mol. Cell *39*, 25–35.

Bakkenist, C.J., and Kastan, M.B. (2003). DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. Nature *421*, 499–506.

Balmus, G., Pilger, D., Coates, J., Demir, M., Sczaniecka-Clift, M., Barros, A.C., Woods, M., Fu, B., Yang, F., Chen, E., et al. (2019). ATM orchestrates

the DNA-damage response to counter toxic non-homologous end-joining at broken replication forks. Nat. Commun. *10*, 87.

Bassi, C., Ho, J., Srikumar, T., Dowling, R.J., Gorrini, C., Miller, S.J., Mak, T.W., Neel, B.G., Raught, B., and Stambolic, V. (2013). Nuclear PTEN controls DNA repair and sensitivity to genotoxic stress. Science *341*, 395–399.

Batey, M.A., Zhao, Y., Kyle, S., Richardson, C., Slade, A., Martin, N.M., Lau, A., Newell, D.R., and Curtin, N.J. (2013). Preclinical evaluation of a novel ATM inhibitor, KU59403, in vitro and in vivo in p53 functional and dysfunctional models of human cancer. Mol. Cancer Ther. *12*, 959–967.

Benitez, A., Liu, W., Palovcak, A., Wang, G., Moon, J., An, K., Kim, A., Zheng, K., Zhang, Y., Bai, F., et al. (2018). FANCA Promotes DNA Double-Strand Break Repair by Catalyzing Single-Strand Annealing and Strand Exchange. Mol. Cell *71*, 621–628.e624.

Bryant, H.E., Schultz, N., Thomas, H.D., Parker, K.M., Flower, D., Lopez, E., Kyle, S., Meuth, M., Curtin, N.J., and Helleday, T. (2005). Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. Nature *434*, 913–917.

Bunting, S.F., and Nussenzweig, A. (2010). Dangerous liaisons: Fanconi anemia and toxic nonhomologous end joining in DNA crosslink repair. Mol. Cell *39*, 164–166.

Cancer Cell Line Encyclopedia Consortium; Genomics of Drug Sensitivity in Cancer Consortium (2015). Pharmacogenomic agreement between two cancer cell line data sets. Nature *528*, 84–87.

Ceccaldi, R., Liu, J.C., Amunugama, R., Hajdu, I., Primack, B., Petalcorin, M.I., O'Connor, K.W., Konstantinopoulos, P.A., Elledge, S.J., Boulton, S.J., et al. (2015). Homologous-recombination-deficient tumours are dependent on Polθ-mediated repair. Nature *518*, 258–262.

Ceccaldi, R., Rondinelli, B., and D'Andrea, A.D. (2016a). Repair Pathway Choices and Consequences at the Double-Strand Break. Trends Cell Biol. *26*, 52–64.

Ceccaldi, R., Sarangi, P., and D'Andrea, A.D. (2016b). The Fanconi anaemia pathway: new players and new functions. Nat. Rev. Mol. Cell Biol. *17*, 337–349.

Chanut, P., Britton, S., Coates, J., Jackson, S.P., and Calsou, P. (2016). Coordinated nuclease activities counteract Ku at single-ended DNA double-strand breaks. Nat. Commun. *7*, 12889.

Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat. Biotechnol. *34*, 184–191.

Duquette, M.L., Zhu, Q., Taylor, E.R., Tsay, A.J., Shi, L.Z., Berns, M.W., and McGowan, C.H. (2012). CtIP is required to initiate replication-dependent interstrand crosslink repair. PLoS Genet. *8*, e1003050.

Eccles, L.J., Bell, A.C., and Powell, S.N. (2018). Inhibition of non-homologous end joining in Fanconi Anemia cells results in rescue of survival after interstrand crosslinks but sensitization to replication associated double-strand breaks. DNA Repair (Amst.) *64*, 1–9.

Farmer, H., McCabe, N., Lord, C.J., Tutt, A.N., Johnson, D.A., Richardson, T.B., Santarosa, M., Dillon, K.J., Hickson, I., Knights, C., et al. (2005). Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. Nature *434*, 917–921.

Gupta, A., Yang, Q., Pandita, R.K., Hunt, C.R., Xiang, T., Misri, S., Zeng, S., Pagan, J., Jeffery, J., Puc, J., et al. (2009). Cell cycle checkpoint defects contribute to genomic instability in PTEN deficient cells independent of DNA DSB repair. Cell Cycle *8*, 2198–2210.

Harris, J.L., and Khanna, K.K. (2011). BRCA1 A-complex fine tunes repair functions of BRCA1. Aging (Albany N.Y.) *3*, 461–463.

Hickson, I., Zhao, Y., Richardson, C.J., Green, S.J., Martin, N.M., Orr, A.I., Reaper, P.M., Jackson, S.P., Curtin, N.J., and Smith, G.C. (2004). Identification and characterization of a novel and specific inhibitor of the ataxia-telangiectasia mutated kinase ATM. Cancer Res. *64*, 9152–9159.

Higgins, G.S., and Boulton, S.J. (2018). Beyond PARP-POLθ as an anticancer target. Science *359*, 1217–1218.

Howard, S.M., Yanez, D.A., and Stark, J.M. (2015). DNA damage response factors from diverse pathways, including DNA crosslink repair, mediate alternative end joining. PLoS Genet. *11*, e1004943.

Huertas, P., and Cruz-García, A. (2018). Single Molecule Analysis of Resection Tracks. Methods Mol. Biol. *1672*, 147–154.

Kennedy, R.D., Chen, C.C., Stuckert, P., Archila, E.M., De la Vega, M.A., Moreau, L.A., Shimamura, A., and D'Andrea, A.D. (2007). Fanconi anemia pathway-deficient tumor cells are hypersensitive to inhibition of ataxia telangiectasia mutated. J. Clin. Invest. *117*, 1440–1449.

Li, J., Yen, C., Liaw, D., Podsypanina, K., Bose, S., Wang, S.I., Puc, J., Miliaresis, C., Rodgers, L., McCombie, R., et al. (1997). PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. Science *275*, 1943–1947.

Lord, C.J., and Ashworth, A. (2016). BRCAness revisited. Nat. Rev. Cancer *16*, 110–120.

Maruyama, T., Dougan, S.K., Truttmann, M.C., Bilate, A.M., Ingram, J.R., and Ploegh, H.L. (2015). Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. Nat. Biotechnol. *33*, 538–542.

Mateos-Gomez, P.A., Gong, F., Nair, N., Miller, K.M., Lazzerini-Denchi, E., and Sfeir, A. (2015). Mammalian polymerase θ promotes alternative NHEJ and suppresses recombination. Nature *518*, 254–257.

McCabe, N., Hanna, C., Walker, S.M., Gonda, D., Li, J., Wikstrom, K., Savage, K.I., Butterworth, K.T., Chen, C., Harkin, D.P., et al. (2015). Mechanistic Rationale to Target PTEN-Deficient Tumor Cells with Inhibitors of the DNA Damage Response Kinase ATM. Cancer Res. *75*, 2159–2165.

Meyers, R.M., Bryan, J.G., McFarland, J.M., Weir, B.A., Sizemore, A.E., Xu, H., Dharia, N.V., Montgomery, P.G., Cowley, G.S., Pantel, S., et al. (2017). Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. Nat. Genet. *49*, 1779–1784.

Mirman, Z., Lottersberger, F., Takai, H., Kibe, T., Gong, Y., Takai, K., Bianchi, A., Zimmermann, M., Durocher, D., and de Lange, T. (2018). 53BP1-RIF1-shieldin counteracts DSB resection through CST- and Polα-dependent fill-in. Nature *560*, 112–116.

Mondal, G., Stevers, M., Goode, B., Ashworth, A., and Solomon, D.A. (2019). A requirement for STAG2 in replication fork progression creates a targetable synthetic lethality in cohesin-mutant cancers. Nat. Commun. *10*, 1686.

Murina, O., von Aesch, C., Karakus, U., Ferretti, L.P., Bolck, H.A., Hänggi, K., and Sartori, A.A. (2014). FANCD2 and CtIP cooperate to repair DNA interstrand crosslinks. Cell Rep. *7*, 1030–1038.

Nakanishi, K., Yang, Y.G., Pierce, A.J., Taniguchi, T., Digweed, M., D'Andrea, A.D., Wang, Z.Q., and Jasin, M. (2005). Human Fanconi anemia monoubiquitination pathway promotes homologous DNA repair. Proc. Natl. Acad. Sci. USA *102*, 1110–1115.

Newell, A.E., Akkari, Y.M., Torimaru, Y., Rosenthal, A., Reifsteck, C.A., Cox, B., Grompe, M., and Olson, S.B. (2004). Interstrand crosslink-induced radials form between non-homologous chromosomes, but are absent in sex chromosomes. DNA Repair (Amst.) *3*, 535–542.

Pace, P., Mosedale, G., Hodskinson, M.R., Rosado, I.V., Sivasubramaniam, M., and Patel, K.J. (2010). Ku70 corrupts DNA repair in the absence of the Fanconi anemia pathway. Science *329*, 219–223.

Pettitt, S.J., Krastev, D.B., Brandsma, I., Dréan, A., Song, F., Aleksandrov, R., Harrell, M.I., Menon, M., Brough, R., Campbell, J., et al. (2018). Genome-wide and high-density CRISPR-Cas9 screens identify point mutations in PARP1 causing PARP inhibitor resistance. Nat. Commun. *9*, 1849.

Rodríguez, A., and D'Andrea, A. (2017). Fanconi anemia pathway. Curr. Biol. *27*, R986–R988.

Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelson, T., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G., and Zhang, F. (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. Science *343*, 84–87.

Shen, W.H., Balajee, A.S., Wang, J., Wu, H., Eng, C., Pandolfi, P.P., and Yin, Y. (2007). Essential role for nuclear PTEN in maintaining chromosomal integrity. Cell *128*, 157–170.

Shiloh, Y., and Ziv, Y. (2013). The ATM protein kinase: regulating the cellular response to genotoxic stress, and more. Nat. Rev. Mol. Cell Biol. *14*, 197–210.

Sobhian, B., Shao, G., Lilli, D.R., Culhane, A.C., Moreau, L.A., Xia, B., Livingston, D.M., and Greenberg, R.A. (2007). RAP80 targets BRCA1 to specific ubiquitin structures at DNA damage sites. Science *316*, 1198–1202.

Song, I.Y., Palle, K., Gurkar, A., Tateishi, S., Kupfer, G.M., and Vaziri, C. (2010). Rad18-mediated translesion synthesis of bulky DNA adducts is coupled to activation of the Fanconi anemia DNA repair pathway. J. Biol. Chem. *285*, 31525–31536.

Stoepker, C., Ameziane, N., van der Lelij, P., Kooi, I.E., Oostra, A.B., Rooimans, M.A., van Mil, S.E., Brink, A., Dietrich, R., Balk, J.A., et al. (2015). Defects in the Fanconi Anemia Pathway and Chromatid Cohesion in Head and Neck Cancer. Cancer Res. *75*, 3543–3553.

Swift, M., and Su, Y. (1999). Link between breast cancer and ATM gene is strong. BMJ *318*, 400.

Symington, L.S. (2016). Mechanism and regulation of DNA end resection in eukaryotes. Crit. Rev. Biochem. Mol. Biol. *51*, 195–212.

Taniguchi, T., Garcia-Higuera, I., Xu, B., Andreassen, P.R., Gregory, R.C., Kim, S.T., Lane, W.S., Kastan, M.B., and D'Andrea, A.D. (2002). Convergence of the fanconi anemia and ataxia telangiectasia signaling pathways. Cell *109*, 459–472.

Taniguchi, T., Tischkowitz, M., Ameziane, N., Hodgson, S.V., Mathew, C.G., Joenje, H., Mok, S.C., and D'Andrea, A.D. (2003). Disruption of the Fanconi anemia-BRCA pathway in cisplatin-sensitive ovarian tumors. Nat. Med. *9*, 568–574.

Terret, M.E., Sherwood, R., Rahman, S., Qin, J., and Jallepalli, P.V. (2009). Cohesin acetylation speeds the replication fork. Nature *462*, 231–234.

Tittel-Elmer, M., Lengronne, A., Davidson, M.B., Bacal, J., François, P., Hohl, M., Petrini, J.H.J., Pasero, P., and Cobb, J.A. (2012). Cohesin association to replication sites depends on rad50 and promotes fork restart. Mol. Cell *48*, 98–108.

Unno, J., Itaya, A., Taoka, M., Sato, K., Tomida, J., Sakai, W., Sugasawa, K., Ishiai, M., Ikura, T., Isobe, T., et al. (2014). FANCD2 binds CtIP and regulates DNA-end resection during DNA interstrand crosslink repair. Cell Rep. *7*, 1039–1047.

van der Heijden, M.S., Brody, J.R., Dezentje, D.A., Gallmeier, E., Cunningham, S.C., Swartz, M.J., DeMarzo, A.M., Offerhaus, G.J., Isacoff, W.H., Hruban, R.H., and Kern, S.E. (2005). In vivo therapeutic responses contingent on Fanconi anemia/BRCA2 status of the tumor. Clin. Cancer Res. *11*, 7508–7515.

Wang, H., Shi, L.Z., Wong, C.C., Han, X., Hwang, P.Y., Truong, L.N., Zhu, Q., Shao, Z., Chen, D.J., Berns, M.W., et al. (2013). The interaction of CtIP and Nbs1 connects CDK and ATM to regulate HR-mediated double-strand break repair. PLoS Genet. *9*, e1003277.

Watrin, E., and Peters, J.M. (2009). The cohesin complex is required for the DNA damage-induced G2/M checkpoint in mammalian cells. EMBO J. *28*, 2625–2635.

Williams, S.A., Longerich, S., Sung, P., Vaziri, C., and Kupfer, G.M. (2011). The E3 ubiquitin ligase RAD18 regulates ubiquitylation and chromatin loading of FANCD2 and FANCI. Blood *117*, 5078–5087.

Wu, N., and Yu, H. (2012). The Smc complexes in DNA damage response. Cell Biosci. *2*, 5.

Yang, W., Soares, J., Greninger, P., Edelman, E.J., Lightfoot, H., Forbes, S., Bindal, N., Beare, D., Smith, J.A., Thompson, I.R., et al. (2013). Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res. *41*, D955–D961.

Yeo, J.E., Lee, E.H., Hendrickson, E.A., and Sobeck, A. (2014). CtIP mediates replication fork recovery in a FANCD2-regulated manner. Hum. Mol. Genet. *23*, 3695–3705.

Zehir, A., Benayed, R., Shah, R.H., Syed, A., Middha, S., Kim, H.R., Srinivasan, P., Gao, J., Chakravarty, D., Devlin, S.M., et al. (2017). Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. Nat. Med. *23*, 703–713.

Zhou, Y., Caron, P., Legube, G., and Paull, T.T. (2014). Quantitation of DNA double-strand break resection intermediates in human cells. Nucleic Acids Res. *42*, e19.

Zimmermann, A., Zenke, F., Dahmen, H., Sirrenberg, C., Grombacher, T., Vassilev, L.T., Fuchss, T., and Blaukat, A. (2018). Abstract 338: A new investigational ATM Inhibitor, M3541, synergistically potentiates fractionated radiotherapy and chemotherapy in cancer cells and animal models. Proceedings of the AACR Annual Meeting 2018 *78*, 338.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Cas9 (7A9-3A3)	Cell Signaling Technology	Cat# 14697S; RRID:AB_2750916
ATM	Cell Signaling Technology	Cat# 2873S; RRID:AB_2062659
pATM (s1981)	Cell Signaling Technology	Cat# 13050S
β-actin	Cell Signaling Technology	Cat# 4967S; RRID:AB_330288
alpha-Tubulin	Cell Signaling Technology	Cat# 2144S; RRID:AB_2210548
GAPDH	Cell Signaling Technology	Cat# 5174S; RRID:AB_10622025
FAAP24	Fisher Scientific	Cat# PA559464; RRID:AB_2641194
FAAP100	Fisher Scientific	Cat# NBP215635
FANCF	Everest Biotech	Cat# EB10140
RAD18	Cell Signaling Technology	Cat# 9040S; RRID:AB_2756446
PTEN	Cell Signaling Technology	Cat# 9188; RRID:AB_2253290
FANCA	Cell Signaling Technology	Cat# 14657S; RRID:AB_2798558
FANCD2	Novus Biologicals	Cat# NB100-182; RRID:AB_10002867
FANCD2	Santa Cruz Biotechnology	Cat# SC-20022; RRID:AB_2278211
γ-H2AX	Cell Signaling Technology	Cat# 2577S; RRID:AB_2118010
RAD51	Cell Signaling Technology	Cat# 8875S; RRID:AB_2721109
RPA	Cell Signaling Technology	Cat# 2208S; RRID:AB_2238543
53BP1	Cell Signaling Technology	Cat# 4937S; RRID:AB_10694558
FANCG	Fisher Scientific	Cat# NB1002566; RRID:AB_921259
BRCA1	Cell Signaling Technology	Cat# 9010S; RRID:AB_2228244
BrdU	Cell Signaling Technology	Cat# 5292S; RRID:AB_10548898
BrdU	Abcam	Cat# ab6326; RRID:AB_2313786
FANCC	Abcam	Cat# ab54631; RRID:AB_2101904
Histone H3	Cell Signaling Technology	Cat# 9715S; RRID:AB_331563
Ku80	Cell Signaling Technology	Cat# 2180S; RRID:AB_1612531
p-DNAPK	Abcam	Cat# ab18192; RRID:AB_869495
DNAPK	Santa Cruz Biotechnology	Cat# SC-390849
XRCC4	Life Technologies	Cat# MA524383; RRID:AB_2633355
PAXX	Cell Signaling Technology	Cat# 92448; RRID:AB_2800186
XLF	Cell Signaling Technology	Cat# 2854S; RRID:AB_2152954
Ligase IV	Cell Signaling Technology	Cat# 14649S; RRID:AB_2750871
CtIP	Cell Signaling Technology	Cat# 9201S; RRID:AB_10828593
MRE11	Cell Signaling Technology	Cat# 4895S; RRID:AB_2145100
NBS1	Cell Signaling Technology	Cat# 14956S; RRID:AB_2798660
RAD50	Cell Signaling Technology	Cat# 3427S; RRID:AB_2176936
BLM	Cell Signaling Technology	Cat# 2742S; RRID:AB_2064649
EXO1	Sigma-Aldrich	Cat# SAB4503568; RRID:AB_10761635
DNA2	Abcam	Cat# ab220883
Chemicals, Peptides, and Recombinant Proteins		
M3541	EMD Serono/Merck Company	N/A
KU-59333	Selleck Chemicals	Cat# S1092
M3814	Chemgood	Cat# C-1358

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Olaparib (AZD2281)	Selleck Chemicals	Cat# S1060
Mitomycin C	Sigma-Aldrich	Cat# M0503
(Z)-4-Hydroxytamoxifen	Sigma-Aldrich	Cat# H7904
5-Bromo-2′-deoxyuridine	Sigma-Aldrich	Cat# B9285
SCR7	Selleck Chemicals	Cat# S7742
Critical Commercial Assays		
CellTiter-Glo Luminescent Viability Assay	Promega	Cat# G7570
DNeasy Blood & Tissue Kit	QIAGEN	Cat# 69504
Subcellular Protein Fractionation Kit	Life Technologies	Cat# 78840
PowerUp SYBR Green Master Mix	Life Technologies	Cat# A25741
Experimental Models: Cell Lines		
A549-CAS9 Expressing Cell Line	Broad Institute	N/A
NCI-H460-CAS9 Expressing Cell Line	Broad Institute	N/A
HeLa Cell Line	ATCC	N/A
293T Cell Line	ATCC	N/A
U2OS Cell Line	ATCC	N/A
RPE Cell Line	ATCC	N/A
MDA-MB-436 Cell Line	Gift from Dr. Geoffrey Shapiro	N/A
MDA-MB-436-olaparib-resistant Cell Line	Gift from Dr. Geoffrey Shapiro	N/A
MDA-MB-436+BRCA1 Cell Line	Gift from Dr. Geoffrey Shapiro	N/A
PD326 Cell Line	Kennedy et al., 2007	N/A
GM6914 Cell Line	Kennedy et al., 2007	N/A
TOV21G Cell Line	Kennedy et al., 2007	N/A
HS766T Cell Line	Kennedy et al., 2007	N/A
PL-11 Cell Line	Kennedy et al., 2007	N/A
Oligonucleotides		
siRNA targeting sequence: FAAP24: GGGTTAGAAATTCCAATAA	Life Technologies	Cat# s40660
siRNA targeting sequence: FAAP100: CCTGCCCGATTATCTGGAG	Life Technologies	Cat# n268651
siRNA targeting sequence: FANCF: GCTATAGAGAGAACCCAAA	Life Technologies	Cat# s5016
siRNA targeting sequence: RAD18: CATATTAGATGAACTGGT	Life Technologies	Cat# s32295
siRNA targeting sequence: PTEN: GCATACGATTTTAAGCGGA	Life Technologies	Cat# s325
siRNA targeting sequence: FANCA: TATTTTTCCCTCTTGACCCTT	QIAGEN	N/A
siRNA targeting sequence: FANCD2: CTGGCTCAGGATTCTAATGTA	QIAGEN	N/A
siRNA targeting sequence: BRCA1: GAATGAGCATGATTTTGAA	Life Technologies	Cat# s224682
siRNA targeting sequence: 53BP1: AGAACGAGGAGACGGTAATAG	QIAGEN	N/A
siRNA targeting sequence: negative Control: AATTCTCCGAACGTGTCACGT	QIAGEN	N/A
See Table S3 for remainder of oligonucleotides		

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Alan D. D'Andrea (Alan_Dandrea@dfci.harvard.edu). The knockout cell lines generated in this study are available from the Lead Contact without restriction.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

A549 and NCI-H460 cells stably expressing the Cas9 endonuclease were obtained from the Genetic Perturbation Platform at the Broad Institute and were grown at 37°C in a humidified 5% CO_2 incubator in RPMI 1640 (GIBCO) supplemented with 10% fetal bovine serum (FBS, Sigma), 1% penicillin-streptomycin (GIBCO) and 100 μg/mL Normocin (InvivoGen, San Diego, CA). HeLa and 293T cells were purchased from ATCC and grown in DMEM (GIBCO) supplemented with 10% FBS and 1% penicillin-streptomycin (GIBCO). U2OS and RPE cells were purchased from ATCC and grown in DMEM/F12 (GIBCO) supplemented with 10% FBS and 1% penicillin-streptomycin (GIBCO). MDA-MB-436, MDA-MB-436-olaparib-resistant, MDA-MB-436+BRCA1 cells were a kind gift from Dr. Geoffrey Shapiro (Dana-Farber Cancer Institute) and grown in RPMI 1640 (GIBCO) supplemented with 10% FBS and 1% penicillin-streptomycin (GIBCO). Generation and maintenance of the PD326, GM6914, TOV-21G, HS766T and PL-11-derived cell lines has been previously described (Kennedy et al., 2007).

METHODS DETAILS

Drug selection CRISPR screens

NSCLC cell lines A549 and NCI-H460 stably expressing the Cas9 endonuclease (pXPR_311) were obtained from the Genetic Perturbation Platform at the Broad Institute. The lines were grown at 37°C in a humidified 5% CO_2 incubator in GIBCO RPMI 1640 (Life Technologies, Grand Island, NY) supplemented with 10% fetal bovine serum (FBS, Sigma-Aldrich, St. Louis, MO) and 100 μg/ml Normocin (InvivoGen, San Diego, CA). The pXPR_311 plasmid contains a blasticidin resistance tag to allow for selection of infected cells. For each line, the selected Cas9 positive cells were infected in four 12-well tissue culture plates with the Brunello whole genome sgRNA pooled lentiviral library consisting of 76,441 sgRNAs targeting 19,114 different genes (Doench et al., 2016). For adequate representation of each sgRNA, e.g., an average of 500 cells/sgRNA, a total of 135 million cells divided into wells each containing 3 million cells were infected at a multiplicity of infection (MOI) of 0.3 in media containing 8 μg/ml polybrene. Plates were centrifuged at 2000 rpm for 2 hours at 37°C, and then 2 hours after centrifugation, media was replaced with fresh media without polybrene and cells were incubated overnight. Infected cells were transferred to 15 cm tissue culture dishes and selected with 2 μg/mL puromycin for 4 days prior to drug treatment. After selection, 40 million cells (8 plates each containing 5 million cells) were plated for each biological replicate, and an additional 40 million cells per replicate were collected and frozen for subsequent genomic DNA (gDNA) extraction. A549 and NCI-H460 cells were then treated with 1 μM M3541 versus DMSO control, for 19 days. Cells were re-plated every 3-4 days and maintained in drug or DMSO-containing media, each time plating a minimum of 40 million cells per replicate. At the end of the screen, the cells were harvested and frozen for gDNA extraction.

Genomic DNA preparation, sequencing and identification of top hits

gDNA was isolated using QIAGEN QIAamp DNA blood midi and maxi kits, according to manufacturer's protocol. PCR of gDNA was performed to attach sequencing adaptors and barcode samples for sgRNA sequencing, as previously described (Doench et al., 2016). Samples were sequenced on an Illumina HiSeq2000. Reads per million were $\log_2$-transformed by first adding one to all values, which was necessary to take the log of sgRNAs with zero reads. For analysis, these normalized data were used to determine the $\log_2$-fold change of each sgRNA for the drug treatment group relative to DMSO control group (Shalem et al., 2014). The STARS gene-ranking algorithm was used to determine the top hits (https://portals.broadinstitute.org/gpp/public/software/stars); at least two sgRNAs targeting a gene had to rank within the top 10% of sgRNAs by $\log_2$-fold change in one direction or another (Doench et al., 2016). The hit genes were scored using the STARS score and the false discovery rate. Additionally, the normalized data were also analyzed with the hypergeometric distribution method (https://github.com/mhegde/volcano_plots). The volcano plots of these results were created using R (https://www.r-project.org).

Western blotting

Cells were lysed with lysis buffer [300 mM NaCl, 50 mM Tris-Cl (pH 7.4), 1 mM EDTA, 0.5% NP-40], lysates were resolved on denaturing Nu-page (Invitrogen) polyacrylamide gels and transferred onto nitrocellulose membranes. Membranes were blocked with 5% milk in TBST, and probed with primary and secondary antibodies respectively, then detected with chemiluminescence (Western Lightning, Perkin Elmer).

CRISPR cas9 Knockout

To generate A549, NCI-H460, HeLa and U2OS cells expressing the indicated sgRNAs, the oligonucleotides described in Table S1 were cloned into the pLentiCRISPR V2 viral vector.

Oligonucleotides used for sgRNA cloning are shown in the supplementary table (Table S3).

siRNA transfection

For siRNA transfection, cells were seeded at about 30% confluence into 6-well plates and transfected with a specific siRNA duplex using Lipofectamine RNAiMAX Reagent Agent (Life Technologies, #13778150) according to the manufacturer's instructions for 48 hr. The sequences of the siRNAs used in this study are shown in the supplementary table (Key Resources Table and Table S3). SiRNA resistant CtIP plasmids are described previously (Chanut et al., 2016).

Clonogenic assay

Cells were seeded at a concentration of 1000-2500 cells per well into 6-well plates. After 24 hr, cells were continuously treated with indicated compounds for about 2 weeks. Colonies were fixed with fixation solution (5 parts methanol + 1 part acetic acid) at room temperature for 20 min and then stained with a solution of 0.5% crystal violet in methanol for 1 hour.

CellTiter-Glo assay

Cells were seeded into 96-well plates at a cell density of 3,000 cells per well, incubated overnight and treated with indicated compounds. The ATP-based, cell viability assay was carried out after incubation for 4 days (CellTiter-Glo, Promega). Survival curves were calculated by best-fit analysis of the log of the drug concentration to fold change of treated cells over vehicle-treated cells. All survival assays included technical triplicates.

Subcellular Fractionation

Subcellular fractions were isolated using a subcellular protein fractionation kit (Thermo Fisher Scientific) according to the manufacturer's instructions, and the resultant fractions were analyzed with western blotting. Antibodies used in this study are listed in the Key Resources Table.

Comet assays

Alkaline comet assays were performed with manufacturer's instruction (Trevigen). Cells were trypsinized and suspended in cold PBS at a concentration of 2.0×10^5 cells/ml. The cells were then mixed with low melting agarose (Trevigen) at a ratio of 1:10 (v/v) and immediately plated onto Cometslide (Trevigen). Alkaline electrophoresis was run at 25 V for 30 minutes in the electrophoresis system. Data were analyzed with a fluorescence microscope (Zeiss).

Reporter assays

NHEJ and HR efficiencies were measured using the EJ5-GFP and DR-GFP reporter assay, respectively (Ceccaldi et al., 2015). Briefly, 24 h before infection with I-Sce-I adenovirus, U2OS cells with either GFP expression cassette were transfected with the indicated siRNA and/or treated with ATMi (1 μM). The HR and NHEJ activity was determined by flow cytometric quantification of viable GFP-positive cells 72 h after I-Sce-I adenovirus infection.

Immunofluorescence

Cells plated on coverslips were washed once with PBS and fixed with 4% paraformaldehyde for 10 minutes on ice. 0.5% Triton X-100 was added to cells for 30 min for extraction. Coverslips were then blocked with 3% BSA for 30 min and incubated with primary antibody (diluted in 3% BSA) overnight. Following overnight incubation, coverslips were washed three times with PBST, then incubated with anti-rabbit secondary antibody (Alexa Fluor, Life Technologies) for 1 hour. Coverslips were then mounted with DAPI medium and visualized under the fluorescence microscope.

Cytogenetics

Cells were transfected with siRNAs twice for 48 h and incubated for 48 h in the presence or absence of 20 ng/ml MMC. Cells were exposed for 2 hours to colcemid (0.1μg/ml) and harvested using a 0.075M KCl hypotonic solution and fixed with 3:1 methanol: acetic acid. Slides were stained with Wright's stain and when possible, 50 metaphase spreads were scored for aberrations. Metaphase spreads were observed using a Zeiss Axio Imager microscope and captured using CytoVision software from Applied Imaging.

SMART assay

Fiber comb assay was performed using the FiberComb machine (Genomic Vision). In brief, cells were treated with bromodeoxyuridine (BrdU, Sigma) for 24 hours. After 24-hour treatment with 1 μM ATMi, cells were then trypsinized, counted and embedded in low melting point agarose plugs, then treated with proteinase K overnight. Agarose plugs were then washed and digested with agarase. Agarase-treated samples were then poured into FiberComb wells and combed onto silanized coverslips. Coverslips were probed

with rat anti-BrdU antibody (Abcam) and visualized by fluorescence microscopy. Pictures were taken of at least 100 fibers per condition. DNA fibers were measured with ImageJ and graphed. Each experiment was repeated at least three times unless otherwise stated.

In vitro DNA end resection

U2OS cells stably expressing HA-ER-AsiSI were created as described previously (Zhou et al., 2014). ssDNAs were quantitated by the HA-ER-AsiSI system in which AsiSI enzyme is fused to the estrogen receptor. The cells were treated with or without 300 nM 4-OHT for four hours to induce DSBs at specific AsiSI sites. Subsequently, genomic DNA was extracted with the QIAGEN DNeasy Blood & Tissue Kit and ssDNA at the specific AsiSI site in chromosome 1 was measured by qPCR with the primers as describing in Table S3.

Computational Analysis

Data from the Broad Institute Cancer Dependency Map Achilles CRISPR Avana 17Q2 dataset (Meyers et al., 2017), the Cancer Cell Line Encyclopedia (Cancer Cell Line Encyclopedia and Genomics of Drug Sensitivity in Cancer, 2015), the Genomics of Drug Sensitivity in Cancer (GDSC) (https://www.cancerrxgene.org/) (Yang et al., 2013), and the PanCancer MSK-IMPACT Clinical Sequencing Cohort (Zehir et al., 2017) were extracted for analysis. The MSK-IMPACT Clinical Sequencing Cohort and Cancer Cell Line Encyclopedia datasets were accessed through the public portal (http://www.cbioportal.org). For dependency analysis, copy number loss of genes was indicated by a copy number variation of −2 or −1 relative to diploid. Normal copy number was indicated by 0 variation relative to diploid. Copy number was chosen as a proxy for gene depletion because RAD18, FAAP24, and FAAP100 mutations are not screened for in cell lines. Cell lines that were in the first quartile of IC_{50} to a drug were categorized as sensitive and cell lines that were in the fourth quartile of IC_{50} were categorized as resistant. The cell lines ± one quartile from the median were not used in our analysis. For analysis of correlation between patient survival and mutations in ATM or FA genes, patient survival data was matched with FANCA, FANCC, and ATM mutation status using the PanCancer MSK-IMPACT Clinical Sequencing Cohort. The study only screened FANCA and FANCC mutations, commonly mutated FA genes. Due to the significant difference in timeline between groups for survival data, all data before 27.25 months (the last patient in the FA and ATM mutated group), was utilized and analyzed.

QUANTITATION AND STATISTICAL ANALYSIS

All quantitative data were analyzed and graphed using Prism 7 GraphPad software. All data are represented as mean ± SEM calculated using the Prism software, unless indicated otherwise. Significance was tested using the non-parametric Mann-Whitney U Test unless indicated otherwise. Clinical data were analyzed using the Kaplan-Meier method and the log-rank (Mantel-Cox) regression test for significance.

DATA AND CODE AVAILABILITY

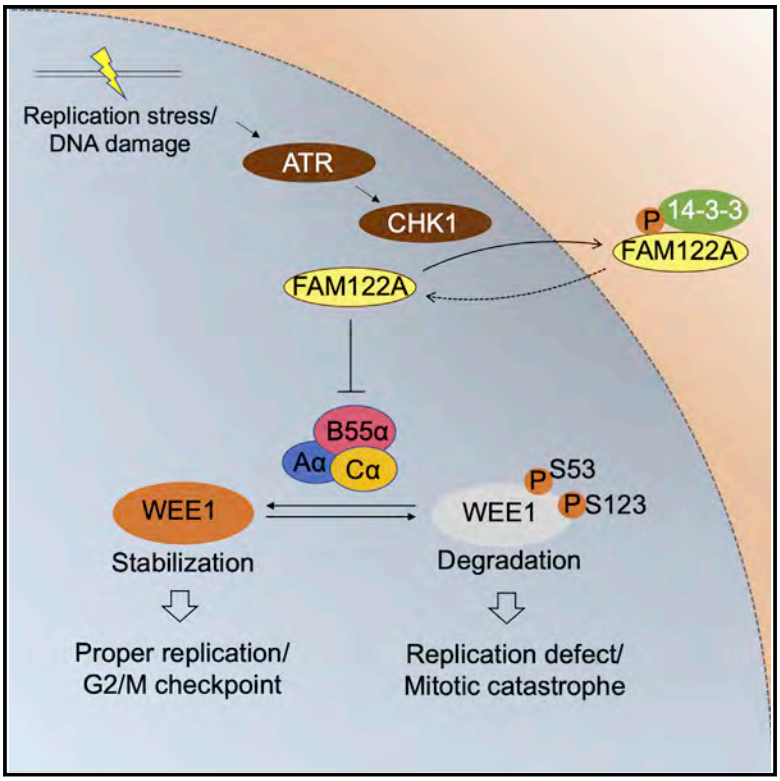
The data for $\log_2$-fold change of each sgRNA for the drug treatment group relative to DMSO control group in A549 and H460 cells in the genome-wide CRISPR screens are available in Table S1. The STARS analysis for each gene is available in Table S2.

Molecular Cell

Article

CHK1 Inhibitor Blocks Phosphorylation of FAM122A and Promotes Replication Stress

Graphical Abstract



Highlights

- CRISPR knockout of FAM122A confers resistance to inhibitors of CHK1 and ATR
- CHK1 phosphorylates FAM122A, thereby activating the phosphatase PP2A
- Activated PP2A dephosphorylates WEE1 and promotes WEE1 stabilization
- Tumor cells acquire drug resistance by deletion or reducing expression of FAM122A

Authors

Feng Li, David Kozono, Peter Deraska, ..., Geoffrey I. Shapiro, Dipanjan Chowdhury, Alan D. D'Andrea

Correspondence

alan_dandrea@dfci.harvard.edu

In Brief

Li and colleagues demonstrate that the cell-cycle checkpoint kinase CHK1 phosphorylates and inactivates FAM122A, an inhibitor of the phosphatase PP2A. Activated PP2A in turn dephosphorylates WEE1, prevents its ubiquitin-mediated proteolysis, increases WEE1 protein levels, and promotes the G2/M checkpoint. Loss of FAM122A expression is a mechanism of acquired CHK1 inhibitor resistance.

Li et al., 2020, Molecular Cell *80*, 410–422
November 5, 2020 © 2020 Elsevier Inc.
https://doi.org/10.1016/j.molcel.2020.10.008

Cell Press

Cell Press

Molecular Cell

Article

CHK1 Inhibitor Blocks Phosphorylation of FAM122A and Promotes Replication Stress

Feng Li,[1] David Kozono,[1] Peter Deraska,[1,2] Timothy Branigan,[3] Connor Dunn,[1,2] Xiao-Feng Zheng,[1] Kalindi Parmar,[1,2] Huy Nguyen,[1,2] James DeCaprio,[3] Geoffrey I. Shapiro,[2,3,4] Dipanjan Chowdhury,[1] and Alan D. D'Andrea[1,2,5,*]

[1]Department of Radiation Oncology, Dana-Farber Cancer Institute, Boston, MA 02215, USA
[2]Center for DNA Damage and Repair, Dana-Farber Cancer Institute, Boston, MA 02215, USA
[3]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 01115
[4]Early Drug Development Center, Dana-Farber Cancer Institute, Boston, MA 02215
[5]Lead Contact
*Correspondence: alan_dandrea@dfci.harvard.edu
https://doi.org/10.1016/j.molcel.2020.10.008

SUMMARY

While effective anti-cancer drugs targeting the CHK1 kinase are advancing in the clinic, drug resistance is rapidly emerging. Here, we demonstrate that CRISPR-mediated knockout of the little-known gene FAM122A/PABIR1 confers cellular resistance to CHK1 inhibitors (CHK1is) and cross-resistance to ATR inhibitors. Knockout of FAM122A results in activation of PP2A-B55α, a phosphatase that dephosphorylates the WEE1 protein and rescues WEE1 from ubiquitin-mediated degradation. The resulting increase in WEE1 protein expression reduces replication stress, activates the G2/M checkpoint, and confers cellular resistance to CHK1is. Interestingly, in tumor cells with oncogene-driven replication stress, CHK1 can directly phosphorylate FAM122A, leading to activation of the PP2A-B55α phosphatase and increased WEE1 expression. A combination of a CHK1i plus a WEE1 inhibitor can overcome CHK1i resistance of these tumor cells, thereby enhancing anti-cancer activity. The FAM122A expression level in a tumor cell can serve as a useful biomarker for predicting CHK1i sensitivity or resistance.

INTRODUCTION

Oncogene-driven cancer cells exhibit replication stress and are hyperdependent on the ATR-CHK1-WEE1 pathway for survival. In this pathway, the accumulation of single-stranded DNA at stalled replication forks activates the ATR kinase, which in turn phosphorylates and activates the CHK1 kinase (Paulsen and Cimprich, 2007; Zhao and Piwnica-Worms, 2001). CHK1 subsequently phosphorylates and inactivates the CDC25 phosphatases, preventing dephosphorylation of T14 and Y15 of CDK1, and phosphorylates WEE1, which phosphorylates T14 and Y15 of CDK1 (Patil et al., 2013). Phosphorylated CDK1 has reduced kinase activity, thus resulting in the activation of the G2/M checkpoint. Reduced entry to mitosis allows cells to tolerate replication stress. Whether CHK1 affects WEE1 expression levels or its cellular function in other ways remains unknown.

Among the kinase inhibitors blocking the ATR-CHK1-WEE1 pathway, new CHK1 inhibitors (CHK1is) have demonstrated therapeutic potential in a variety of tumor types (King et al., 2015). Phase 1 clinical trials of the CHK1i prexasertib (LY2606368) have been completed in non-small cell lung cancer (NSCLC) among others (Hong et al., 2016; Hong et al., 2018), and several phase 1 and 2 trials are currently actively recruiting patients. By inhibiting the G2/M- and S-phase checkpoints, CHK1is cause increased replication stress, DNA damage, and premature entry of cancer cells into mitosis, resulting in mitotic catastrophe and cancer cell death. CHK1is also have enhanced activity when combined with an ATR inhibitor (ATRi) through a mechanism of synthetic lethality (Sanjiv et al., 2016). While these inhibitors provide a therapeutic benefit *in vivo*, acquired resistance can also result. Little is known about how cancer cells become resistant to cell-cycle checkpoint kinase inhibitors. Understanding potential resistance mechanisms may allow the prediction of patient responses, the design of CHK1 therapy combinations, and the implementation of mechanisms for resensitization.

Recent studies indicate that, in addition to WEE1 kinase activity, an increase in WEE1 protein expression level may positively regulate the G2/M checkpoint. WEE1 protein levels are regulated, at least in part, by protein degradation (Watanabe et al., 2004, 2005). The PLK1 kinase and CDC2 kinase phosphorylate WEE1 on two sites (S53 and S123), resulting in Skp1, Cullins, F-box proteins (SCF) E3-ligase-mediated and ubiquitination/protein dependent degradation. Accordingly, downregulation of either WEE1 phosphorylation or of SCF E3 ligase activity can upregulate WEE1 protein expression, thereby enhancing the G2/M checkpoint. Other recent studies suggest that the multisubunit phosphatase PP2A regulates the dephosphorylation of WEE1 and may rescue WEE1 from degradation. Indeed, PP2A

has multiple functions in controlling the G2/M checkpoint (Forester et al., 2007; Wlodarchak and Xing, 2016), such as controlling the phosphorylation state of CDC25A itself. How PP2A activity is regulated during the cell cycle is incompletely known. There are substantially more kinases in the human genome compared to phosphatase genes (Duong-Ly and Peterson, 2013). Target specificity of the phosphatases including PP2A, a heterotrimeric complex consisting of a scaffolding A subunit, a regulatory B subunit, and a catalytic C subunit, is largely conferred by the incorporation of one of a multiple possible B subunits (Sangodkar et al., 2016).

Here, we performed a genome-wide clustered regularly interspersed short palindromic repeat (CRISPR)-Cas9 loss-of-function genetic screen to identify single guide RNAs (sgRNAs) that can either increase or decrease cell killing by the CHK1i prexasertib (King et al., 2015). sgRNAs directed against the protein FAM122A constituted a top hit in the screen, and FAM122A knockout in multiple cell lines resulted in resistance to CHK1is. Since FAM122A was recently shown to be a suppressor of PP2A-B55α (Fan et al., 2016), we reasoned that downregulation of FAM122A might enhance PP2A-B55α-mediated stabilization of WEE1 levels. Indeed, FAM122A knockout resulted in increased PP2A activity, increased WEE1 protein levels, reduced replication stress, and restoration of the G2/M checkpoint in cells exposed to the CHK1i. Knockdown of the B55α regulatory subunit, but not other B subunits, including B56α or B56δ, restored prexasertib sensitivity in FAM122A knockout cells, in keeping with the specificity of FAM122A against PP2A-B55α. Interestingly, CHK1 directly phosphorylated FAM122A on a highly conserved site, S37, leading to its release from the PP2A complex and an upregulation in PP2A activity. Phosphorylated FAM122A was sequestered by 14-3-3 proteins in the cytoplasm and failed to translocate to the nucleus, where it otherwise inhibits PP2A. CHK1i or small interfering RNA (siRNA) knockdown of CHK1 resulted in accumulation of FAM122A in the nucleus. Moreover, FAM122A knockout promoted cellular resistance to not only CHK1is but also ATR inhibitors, which are also currently undergoing clinical development (Qiu et al., 2018; Weber and Ryan, 2015). Taken together, our results indicate a novel process by which CHK1 reduces replication stress and activates the G2/M checkpoint, namely by phosphorylating and inactivating FAM122A and promoting the PP2A-mediated dephosphorylation and stabilization of WEE1 levels and activity. Loss of CHK1-mediated phosphorylation of FAM122A provides a useful pharmacodynamic biomarker for the development of new CHK1is. Moreover, the decreased cellular expression of FAM122A and the increased level of WEE1 protein may serve as useful biomarkers for predicting CHK1i resistance.

RESULTS

Identification of sgRNAs that Selectively Suppress Cell Killing by the CHK1i Prexasertib

Initially, we used the A549 and H460 (NCI-H460) NSCLC lines to conduct a CRISPR screen for sgRNAs that can either increase or decrease cell killing by the CHK1i prexasertib (Figure 1A). These cells were transduced to stably express the Cas9 endonuclease, and they were recently used to conduct a similar screen to identify cellular enhancers and inhibitors of an ataxia telangiectasia mutated (ATM) inhibitor (Cai et al., 2020). The A549 and H460 cell lines are wild type (WT) for both ATM and the pro-apoptotic gene p53.

In order to identify sgRNAs that target genes involved in resistance to the CHK1i, we chose a moderate concentration of the drug that results in no net cellular proliferation over a 2-week period. A549 and H460 cells exhibit an IC_{50} for prexasertib in the 10–100 nM range, similar to the drug sensitivity of several other cancer cell lines (Figures S1A and S1B). For the CRISPR screen, we chose a prexasertib concentration of 100 nM for A549 cells and a concentration of 50 nM for H460 cells (Figures S1C and S1D). Cells were infected with the Brunello v2 lentiviral sgRNA library and grown for 19 days in media treated with the CHK1i. Top hits among gene knockouts that conferred drug resistance were identified via the STARS gene-ranking algorithm (Doench et al., 2016).

FAM122A scored as a top hit in both lung cancer cell lines (Figure 1B; Tables S1 and S3). FAM122A is a little-known but highly conserved protein that has recently been shown to inhibit the PP2A multisubunit phosphatase with the B55α regulatory subunit (PP2A-B55α) (Fan et al., 2016). sgRNA-mediated knockout of FAM122A (Figure 1C), or siRNA-mediated knockdown of FAM122A, in A549 cells or H460 cells (Figures S1E–S1J) increased resistance to prexasertib, thereby validating the CRISPR screen results. sgRNA knockout of FAM122A also resulted in cellular resistance to another known CHK1i, SRA737 (Figure 1D) and, interestingly, generated cross-resistance to the ATR inhibitor VE822 (Figure 1E). siRNA knockdown of FAM122A also conferred resistance to the ATR inhibitor in another cell line, U2OS (Figure 1F).

Knockout of FAM122A Reduces Replication Stress

CHK1is are known to activate replication stress and disrupt the G2/M checkpoint of cancer cells, and these activities appear to be their major anti-cancer mechanism of action (Parmar et al., 2019). As expected, cellular exposure to CHK1is resulted in a dose-dependent increase in replication stress, as measured by enhanced phosphorylation of KAP1 and CHK1, and enhanced DNA double-strand break (DSB) generation (Figure 2A, lanes 1–4). CHKi exposure also resulted in an increase in phosphorylation of CHK1, consistent with the loss of a negative feedback loop and the increase in ATR kinase activity toward CHK1. DSB generation was verified by γH2AX staining (Figure 2B). Interestingly, sgRNA knockout of FAM122A resulted in reduced replication stress, providing a mechanism of increased CHK1 drug resistance (Figure 2A, lanes 5–8). CHK1i also had a strong impact on DNA replication and the cell cycle. Specifically, CHK1i resulted in an increase of unreplicated DNA and accumulation of G2/M cells that was reversed by the sgRNA knockout of FAM122A (Figures 2C and S2A–S2C).

As previously described (Parmar et al., 2019), CHK1i also caused reduced replication fork speed (Figures 2D and 2E); increased DNA damage, as measured by the comet assay (Figures 2F and 2G); and increased chromosome breakage (Figures S2D and S2E). In each case, sgRNA to FAM122A reversed these events, again correlating with the ability of FAM122A knockout to confer cellular resistance to CHK1i.

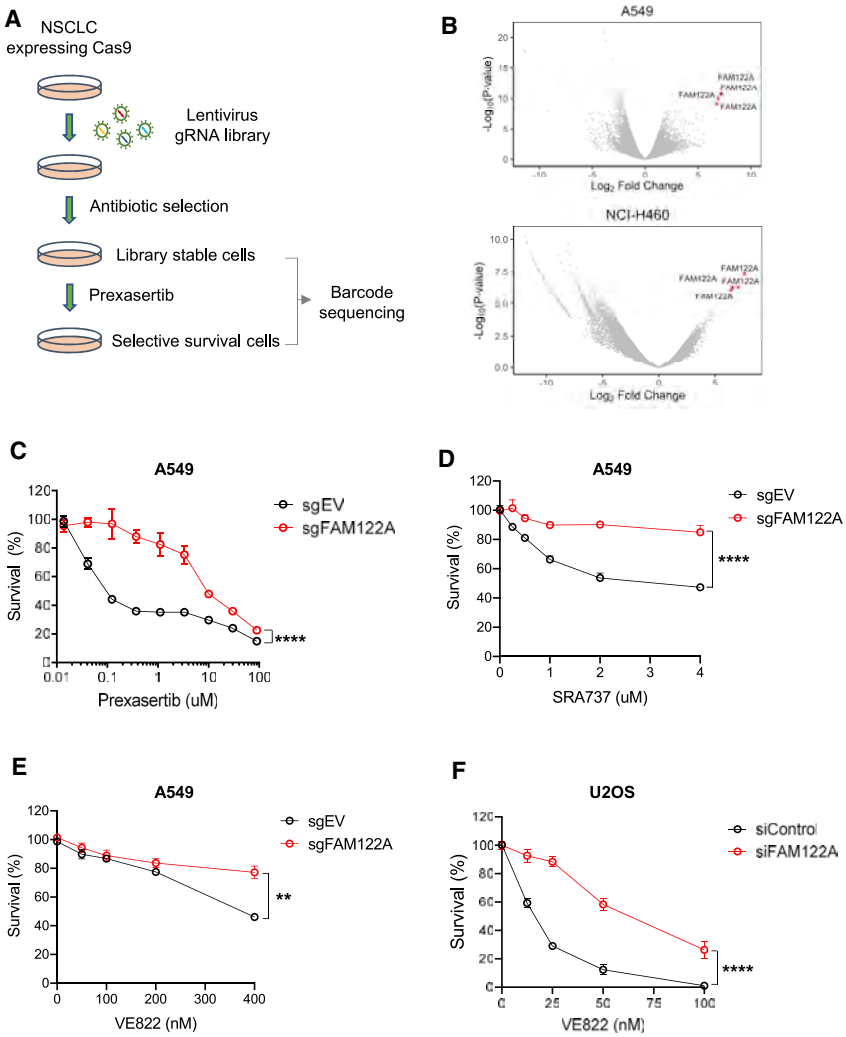


Figure 1. Genome-wide CRISPR Screen Reveals that FAM122A Loss Results in Resistance to CHK1is and ATRi in Lung Cancer Cells

(A) Schematic of CRISPR-based screen for resistance to CHK1i (prexasertib) in lung cancer cells. Cells were infected with the Brunello sgRNA library (containing ~74,000 different sgRNAs) with a multiplicity of infection of 0.3. Following puromycin selection, cells were passaged for three weeks in the presence of prexasertib.

(B) Volcano plots showing genes targeted by sgRNAs that were differentially depleted or enriched following prexasertib selection in A549 and NCI-H460 cells. Note that all four sgRNAs directed to FAM122A were found to be significantly enriched following prexasertib selection in A549 and NCI-H460 cells.

(C) Survival plots of control and FAM122A knockout (FAM122A-KO) A549 cells after exposure to CHK1i (prexasertib) for 5 days. Data are shown as mean ± SD from three independent experiments. ****$p < 0.0001$, statistical analysis was performed using two-way ANOVA.

(D) Survival plots of control and FAM122A-KO A549 cells after exposure to CHK1i (SRA737) for 5 days. Data are shown as mean ± SD from three independent experiments. ****$p < 0.0001$; statistical analysis was performed using two-way ANOVA.

(E) Survival plots of control and FAM122A-KO A549 cells after exposure to ATRi (VE822) for 5days. Data are shown as mean ± SD from three independent experiments. **$p < 0.01$; statistical analysis was performed using two-way ANOVA.

(F) Survival plots of siControl and siFAM122A U2OS cells after exposure to ATRi (VE822) for 10 days. Data are shown as mean ± SD from three independent experiments. ****$p < 0.0001$; statistical analysis was performed using two-way ANOVA.

Depletion of FAM122A Promotes WEE1 Protein Stability

CHK1 is known to directly phosphorylate WEE1, thereby activating a G2/M checkpoint (Leung-Pineda et al., 2006). Whether CHK1 also directly or indirectly regulates WEE1 expression levels has not been previously demonstrated. We initially observed that CHK1 inhibition in A549 cells resulted in reduced levels of WEE1 protein (Figure 3A, lanes 1 and 2). Interestingly, sgRNA-mediated knockout of FAM122A prevented the CHK1i-activated decline in WEE1 protein level and increased the expression of WEE1 protein (Figure 3A, lanes 3 and 4). Consistent with this, knockout FAM122A also increased resistance to the WEE1 inhibitor AZD1775 (Figure S3B). Again, knockout of FAM122A correlated with reduced γH2AX levels after cellular exposure to CHK1i (Figure 3A, lanes 3 and 4).

The increase in WEE1 protein expression following FAM122A knockout could either result from increased production or decreased degradation of WEE1. Previous studies have indicated that the WEE1 protein is degraded and that this process is regulated by protein phosphorylation and ubiquitin-mediated and proteasome proteolysis (Watanabe et al., 2004, 2005). Specifically, WEE1 is phosphorylated on S53 and S123 by the PLK1 and CDC2 kinases, respectively, thereby creating phosphodegrons and resulting in WEE1 degradation of by SCF E3 ligase complex.

To further analyze the role of FAM122A in WEE1 protein degradation, we evaluated the regulation of WEE1 protein degradation in cells with a knockout of FAM122A (Figure 3B). Again, the CHK1i promoted the reduction of WEE1 protein levels (Figure 3B, lanes 2 and 4), and knockout of FAM122A resulted in increased WEE1 protein expression (Figure 3B, lanes 3 and 4). Knockout of FAM122A failed to increase the expression of the WEE1-pS53 or WEE1-pS123 polypeptides. Furthermore, while CHK1i promoted the degradation of WEE1 protein, this degradation was prevented by cellular exposure to the proteasome inhibitor MG132 (Figure S3A). These results are consistent with the known ubiquitin-mediated degradation of the phosphorylated isoforms of WEE1. The stabilization of WEE1 protein, resulting from FAM122A knockout, was confirmed following cellular exposure to a cycloheximide challenge (Figure 3C). Taken together, these results demonstrate that FAM122A knockout regulates

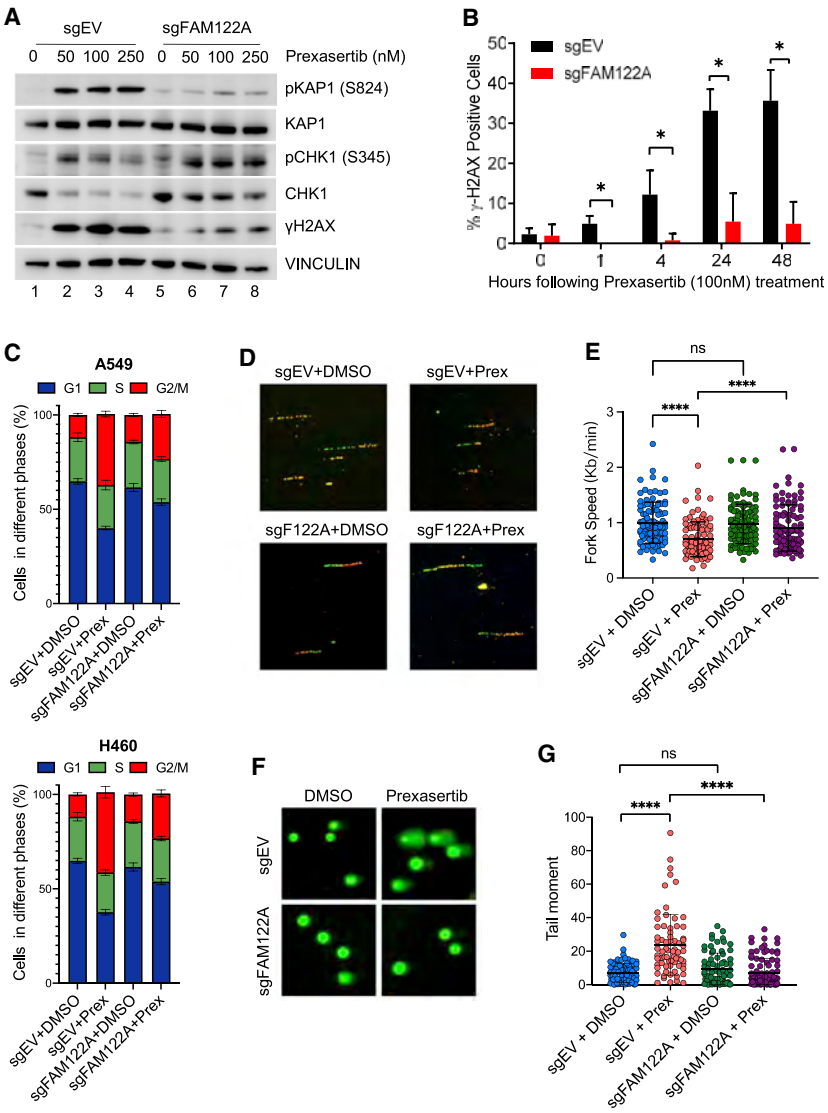


Figure 2. FAM122A Loss Rescues Lung Cancer Cells from Replication Stress, DNA Damage, and G2/M Arrest Caused by CHK1is

(A) Western blots of the lysates from control and FAM122A-KO A549 cells after treatment with prexasertib for 24 h.

(B) Quantification of γH2AX foci in control and FAM122A-KO A549 cells after treatment with prexasertib for up to 48 h. Error bars indicate standard errors, and p values were calculated using Student's t test (n = 3, *p < 0.001).

(C) Cell-cycle distribution of phosphatidylinositol (PI)-stained control and FAM122A-KO A549 (upper panels) and H460 (lower panels) cells treated with 250 nM prexasertib for 24 h. Data are shown as mean ± SD from three independent experiments.

(D) Representative images of the DNA fibers in control and FAM122A-KO A549 cells after treatment with DMSO or prexasertib (100 nM) for 24 h. Cells were sequentially labeled with IdU and CldU after exposure to prexasertib for 24 h, and the DNA fiber assay was performed to determine the replication fork speed.

(E) Quantitation of the replication fork speed in control and FAM122A-KO A549 cells after treatment with DMSO or prexasertib (100 nM) for 24 h. DNA fibers in (D) were analyzed from >100 ongoing replication forks in each condition. Error bars indicate standard errors, and p values were calculated using Student's t test (****p < 0.0001).

(F) Representative images of alkaline comets in control and FAM122A-KO A549 cells exposed to prexasertib (250 nM) for 24 h.

(G) Quantitation of the comet tail moment in (F) from more than 100 cells in each condition. Error bars indicate standard errors, and p values were calculated using Student's t test (****p < 0.0001).

the stability of the WEE1 protein, and its effect appears to be mediated via the phosphorylation state of WEE1.

It is well known that inhibitors of ATR, CHK1, or WEE1 can cause replication stress (Matheson et al., 2016; Qiu et al., 2018; Weber and Ryan, 2015), increase phosphorylation of the pKAP1 protein, and increase γH2AX levels. As expected, the knockout of FAM122A reduced the replication stress and DNA damage induced by the CHK1i prexasertib. Interestingly, we observed increased replication stress resulting from the combination of prexasertib plus the WEE1 inhibitor (WEE1i) AZD1775 combination (Figure 3D). Accordingly, while loss of FAM122A conferred a strong level of resistance to prexasertib, these cells were partially resensitized by the combination of prexasertib and WEE1 inhibitor in the media (Figures 3E and S3C). The combination of prexasertib and AZD1775 also increased the replication stress and G2/M accumulation in FAM122A knockout A549 cells (Figures S3D and S3E).

FAM122A Regulates WEE1 Protein Levels through Inhibition of the PP2A Phosphatase

A recent study indicated that the FAM122A protein binds and inhibits the PP2A phosphatase with the B55α regulatory subunit (Fan et al., 2016), although the functional impact on replication stress and cell-cycle transition was not evaluated. PP2A is a ubiquitously expressed, heterotrimeric serine/threonine (S/T) phosphatase consisting of a scaffold (A) subunit, a regulatory (B) subunit, and a catalytic (C) subunit (Janssens and Goris, 2001; Janssens et al., 2008). Based on the assortment and use of these three subunits, there may be as many as eighty distinct trimeric PP2A complexes, some with distinct cell-specific expression and substrate preference. PP2A is known to regulate many cell functions, including several cell-cycle transitions (Harvey et al., 2011; Johnson and Kornbluth, 2012; Wlodarchak and Xing, 2016). Target specificity of PP2A is directed by the variable regulatory (B) subunit. PP2A dephosphorylates several substrates involved in transformation and cancer, including MYC, Wnt, and components of the phosphatidylinositol 3-kinase (PI3K)/AKT pathway. By dephosphorylating pCDC25, PP2A also plays

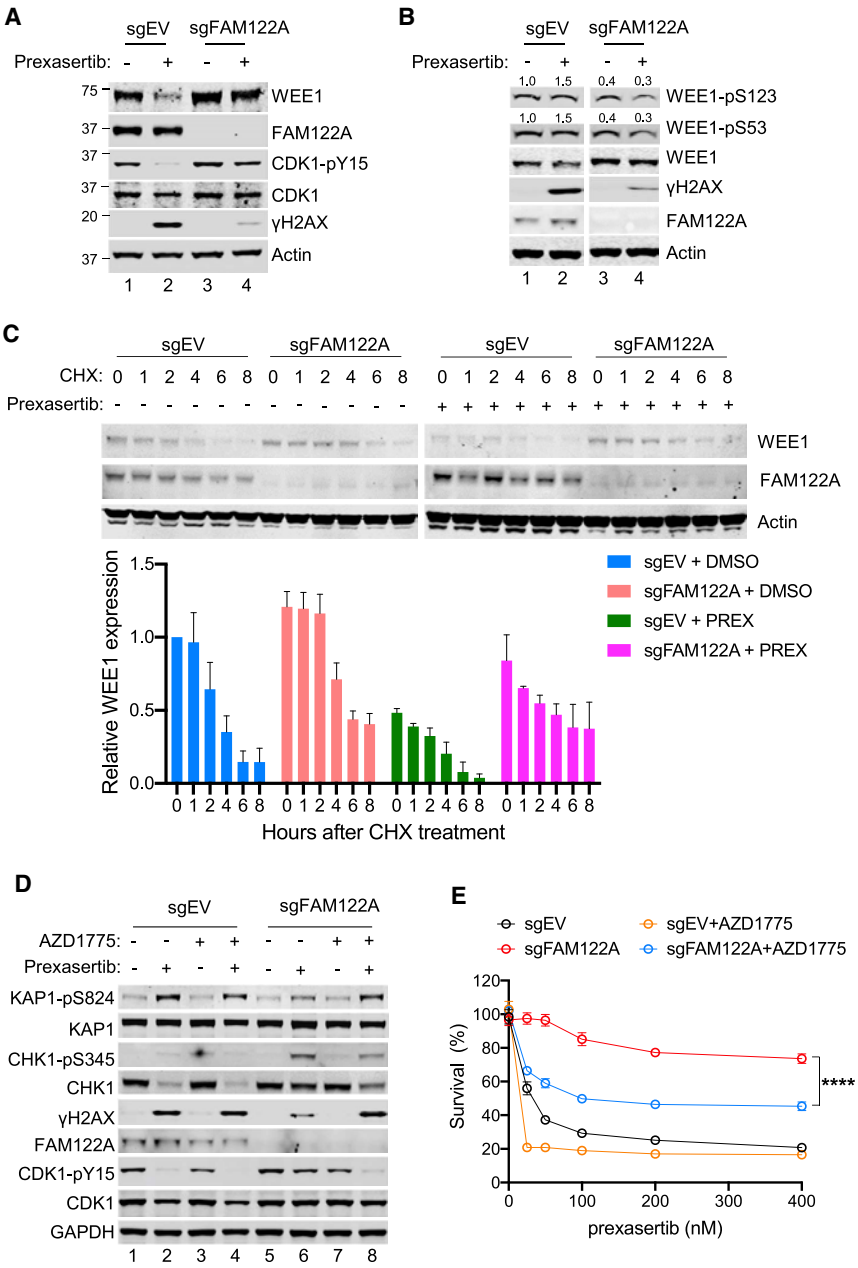


Figure 3. Depletion of FAM122A Promotes WEE1 Stability in A549 Lung Cancer Cells

(A) Western blots of the lysates from control and FAM122A-KO A549 cells after treatment with prexasertib (100 nM) for 24 h.

(B) Western blots of the lysates from control and FAM122A-KO A549 cells treated with/without prexasertib (100 nM) for 16 h. The relative levels of phosphorylated WEE1 compare to WEE1 were quantified as indicated.

(C) Top: Western blots of the lysates from control and FAM122A-KO A549 cells treated with DMSO or prexasertib (100 nM) for 24 h, followed by treatment with cycloheximide (CHX) for the indicated duration hours. Bottom: quantification of WEE1 protein levels. Error bars indicate standard errors (n = 3).

(D) Western blots of the lysates from control and FAM122A-KO A549 cells treated with prexasertib (100 nM) or AZD1775 (50 nM) for 24 h.

(E) Survival plots of control and FAM122A-KO A549 cells treated with graded concentrations of prexasertib along with 50 nM WEE1i (AZD1775). Note that WEE1i sensitizes resistant FAM122A-KO A549 cells to prexasertib. Data are shown as mean ± SD from three independent experiments. sgFAM122A versus sgFAM122A+AZD1775, ****p < 0.0001; statistical analysis was performed using two-way ANOVA.

a critical role in the G2/M cell-cycle transition (Forester et al., 2007; Yan et al., 2010).

While loss of function of PP2A is critical for tumor cell growth, mutations in the PP2A subunits are rare, suggesting that other mechanisms can modulate PP2A activity. Indeed, endogenous inhibitors of PP2A, such as SET and CIP2A, have been identified. These inhibitors have abnormal expression patterns in cancer cells and are candidate anti-cancer targets (Janghorban et al., 2014; Junttila et al., 2007; Khanna et al., 2013). FAM122A was recently identified as an endogenous inhibitor of the Aα/B55α/Cα trimeric complex of PP2A, and it has direct binding interactions with the B55α and Cα subunits of this complex. FAM122A binding inhibits the catalytic activity of this specific PP2A complex and promotes the ubiquitin-dependent degradation of the Cα subunit (Fan et al., 2016).

We reasoned that the PP2A complex, and specifically the Aα/B55α/Cα trimeric complex, might stabilize WEE1 protein by dephosphorylating WEE1 at the S53 and S123 sites. Furthermore, given its function as an inhibitor of this PP2A-B55α complex (Fan et al., 2016), the FAM122A protein might play a normal cellular role in inhibiting PP2A-B55α, blocking WEE1 dephosphorylation, and promoting WEE1 degradation, thereby releasing the G2/M checkpoint. To test this hypothesis, we further analyzed the data collected from our CRISPR screen in the A549 and H460 cells. While sgRNAs directed against FAM122A conferred CHK1i resistance, sgRNAs directed against the B55α regulatory subunit (PPP2R2A) of the PP2A phosphatase complex conferred sensitivity to CHK1is (Figures 4A and 4B; Tables S1, S2, and S3). This observation supported a model in which FAM122A loss leads to PP2A-B55α dephosphorylation of WEE1, resulting in WEE1 protein stabilization, reduction of replication stress, and an enhanced G2/M checkpoint.

We next further examined the role of specific subunits of the trimeric PP2A complex in the dephosphorylation of WEE1 (Figure 4C). A549 cells were sensitive to the CHK1i prexasertib, and knockout of FAM122A resulted in prexasertib resistance. siRNA knockdown of the B55α subunit of PP2A resensitized

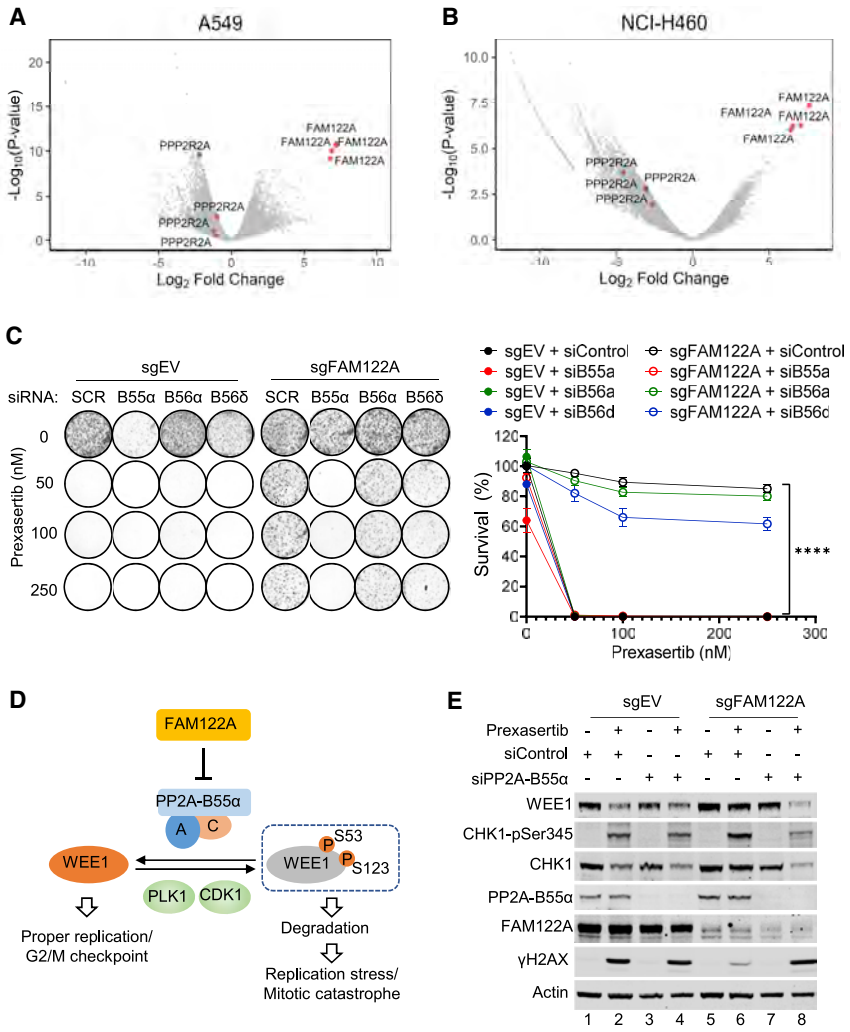


Figure 4. FAM122A Regulates WEE1 Stability by Targeting PP2A-B55α in Lung Cancer Cells

(A) Volcano plots showing genes targeted by sgRNAs that were differentially depleted or enriched following prexasertib selection in A549 cells. Note that all four sgRNAs directed to the PP2A B55α subunit PPP2R2A were found to be decreased following prexasertib selection in A549 cells.

(B) Volcano plots showing genes targeted by sgRNAs that were differentially depleted or enriched following prexasertib selection in H460 (NCI-H460) cells. Note that all four sgRNAs directed to the PP2A B55α subunit PPP2R2A were found to be decreased following prexasertib selection in H460 cells.

(C) Clonogenic assay of control and FAM122A-KO A549 cells treated with indicated siRNAs against PP2A subunits, followed by exposure to prexasertib. Note that knockdown of the B55α subunit of PP2A reverses CHK1i prexasertib resistance in FAM122A-KO A549 cells. In the right panel, siB55α and siB56α are referred as siB55a and siB56a, respectively. SiB56δ is referred as siB56d. Data are shown as mean ± SD from three independent experiments. sgFAM122A+siControl versus sgFAM122A+siPP2A-B55α, ****$p < 0.0001$; statistical analysis was performed using two-way ANOVA.

(D) Schematic of the model for WEE1 regulation by the FAM122A-PP2A pathway.

(E) Western blots of the lysates from control and FAM122A-KO A549 cells treated with prexasertib (100 nM) for 24 h after treatment with siControl or siPP2A-B55α.

FAM122A knockout cells to the drug, demonstrating that B55α is the critical subunit of PP2A required for dephosphorylation of WEE1. Accordingly, loss of B55α resulted in increased WEE1 degradation and increased prexasertib sensitivity. In contrast, siRNA knockdown of other B subunits of PP2A, such as the B56α and B56δ subunits, failed to resensitize the FAM122A knockout cells to prexasertib. Taken together, these results support a model in which PP2A, containing the B55α subunit, is the critical trimeric phosphatase required for FAM122A-regulated dephosphorylation of WEE1 (Figure 4D). To further confirm this model, we determined the effect of PP2A-B55α knockdown on the WEE1 protein expression level (Figure 4E). Again, knockout of FAM122A in A549 cells resulted in an increased WEE1 protein expression level, which was reversed by knockdown of the B55α subunit (Figure 4E, lane 8).

Moreover, knockdown of another PP2A subunit, the Cα catalytic subunit, also reduced the level of WEE1 protein (Figure S4A). Also, a broad-spectrum inhibitor of PP2A activity, okadaic acid (OA), reduced WEE1 protein levels and was able to override the effect of the FAM122A knockout (Figure S4B). The PP2A inhibitor LB-100 was synergistic with prexasertib, and this synergism was increased by sgRNA knockout of FAM122A (Figure S4C). These results suggest that PP2Ai and CHK1i may both contribute to the degradation of WEE1. Thus, a combination of a PP2A inhibitor plus a CHK1i may be synergistic in an anti-cancer clinical trial. siRNA knockdown of α-endosulfine (ENSA), another endogenous inhibitor of PP2A, but one that targets the B55δ regulatory subunit (Mochida et al., 2010), did not result in a reduction of prexasertib-mediated cell death (Figure S4D), suggesting a specific role of FAM122A in the regulation of CHK1 pathway. Taken together, these data support a model in which a trimeric PP2A complex, composed of the Aα/B55α/Cα subunits, can dephosphorylate WEE1. Consistent with these results, FAM122A is a known inhibitor of this PP2A-B55α complex (Fan et al., 2016). These findings also suggest that a PP2A inhibitor, or a WEE1 inhibitor, may be useful in the treatment of CHK1i-resistant tumor cells.

CHK1 Phosphorylates FAM122A on S37, Resulting in Activation of PP2A

Inspection of the primary amino acid sequence of FAM122A revealed a conserved potential CHK1 phosphorylation site at S37 (Figure S5A). The amino acid consensus sequence for a CHK1 phosphorylation site is RXXS. Moreover, phosphoproteomic databases demonstrated that FAM122A is indeed

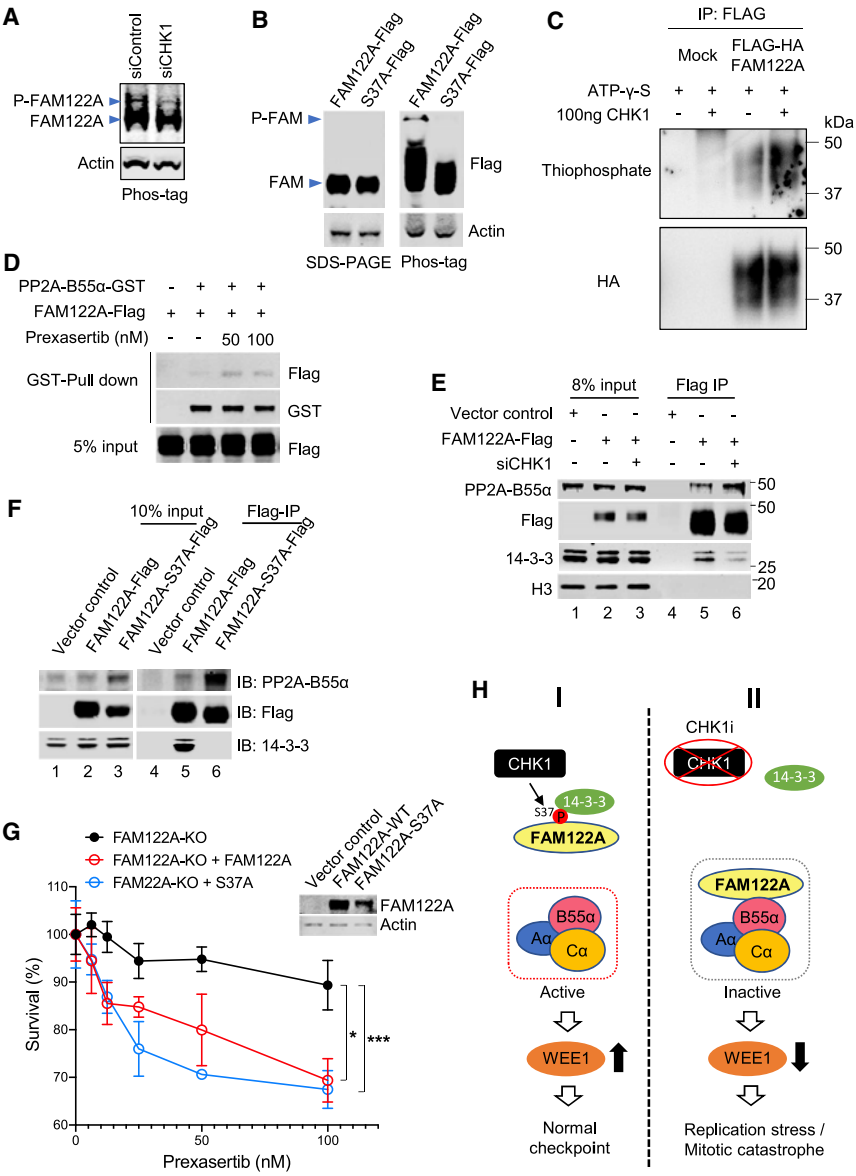


Figure 5. Chk1 Phosphorylates FAM122A at S37

(A) Western blots of the lysates from A549 cells treated with siControl or siCHK1. Samples were separated on a 12% Phos-tag-SDS-PAGE gel. Phosphorylated and unphosphorylated FAM122A are labeled.

(B) Western blots of the lysates from 293T cells expressing FLAG-tagged FAM122A (wild-type [WT] and S37A). Samples were separated on a normal SDS-PAGE gel (left) and a Phos-tag-labeled SDS-PAGE gel (right). Phosphorylated and unphosphorylated FAM122A are labeled.

(C) Kinase assays of 293T cells expressing FLAG-FAM122A. 293T cells were transiently transfected with hemagglutinin (HA)-FLAG-tagged FAM122A, followed by immunoprecipitation with ant-Flag antibody. HA-FLAG-FAM122A was incubated with ATP-γ-S and with or without purified CHK1. Phosphorylation of FAM122A was detected by thiophosphate.

(D) Anti-glutathione S-transferase (anti-GST) immunoprecipitation assays followed by western blots of stably expressing FAM122A-FLAG 293T cells. FAM122A-FLAG 293T cells were transiently transfected with PP2A-B55α-GST and then treated with prexasertib at indicated concentrations for 24 h, followed by GST-pull-down assay and western blots.

(E) Anti-FLAG immunoprecipitation assays followed by western blots of FAM122A-FLAG 293T cells. Stably expressing FAM122A-FLAG or control 293T cells were treated with siCHK1, followed by FLAG-pull-down assay, and western blots using anti-FLAG, anti-PP2A-B55α, anti-14-3-3, and anti-H3 antibodies.

(F) Anti-FLAG immunoprecipitation (IP) assays followed by western blots of FAM122A-FLAG 293T cells. 293T cells expressing FLAG-tagged FAM122A (WT and S37A mutant) were used for FLAG-IP and western blots.

(G) Survival plots of FAM122A-KO A549 cells transiently transfected with FAM122A-WT and -S37A mutant, followed by treatment with prexasertib for 3 days. Data are shown as mean ± SD from three independent experiments. sgFAM122A versus sgFAM122A+FAM122A, *p < 0.05; sgFAM122A versus sgFAM122A+FAM122A-S37A, ***p < 0.001. Statistical analysis was performed using two-way ANOVA. The expression of the WT or mutant FAM122A protein is shown on the western blot (see inset).

(H) Illustration of the CHK1-FAM122A-PP2A-B55α pathway.

a phosphoprotein. (https://www.phosphosite.org/siteGroupAction.action?id=3208890&protOrg=9790&showAllSites=true&showHTPRefsOnly=true#4), though the functional importance of its phosphorylation is unknown. We reasoned that the CHK1-mediated phosphorylation of FAM122A might negatively regulate the suppressive activity of FAM122A, resulting in increased PP2A-mediated phosphatase activity directed toward the WEE1 substrate. We therefore tested whether the S37 residue is a bona fide CHK1-mediated phosphorylated site of FAM122A.

Using a phosphotag reagent (Kinoshita et al., 2015), we demonstrated that WT FAM122A is indeed a phosphoprotein (Figures 5A and 5B). siRNA-mediated knockdown of CHK1 reduced FAM122A phosphorylation (Figure 5A). In contrast, a mutant version of FAM122A, containing the S37A mutation, failed to bind the phosphotag reagent (Figure 5B). Consistent with this result, WT FAM122A was also phosphorylated by CHK1 in an *in vitro* kinase assay (Figure 5C), demonstrating that FAM122A is a direct substrate of CHK1. The CHK1i prexasertib reduced the phosphorylation of WT-FAM122A (Figures S5B and S5C). The effect of siRNA-mediated knockdown of CHK1 (Figure 5A) was comparable to the effect of cellular exposure to the CHK1i (Figures S5B and S5C). Also, hydroxyurea (HU), a known stimulant of CHK1 kinase activity, activated the phosphorylation of FAM122A, and prexasertib inhibited this phosphorylation (Figure S5D).

We next performed a GST-PP2A pull down of cellular FAM122A. GST-PP2A-B55α could weakly bind to WT FAM122A, but the binding activity was significantly enhanced when prexasertib was used to inhibit the CHK1-mediated phosphorylation of FAM122A (Figure 5D). Taken together, these results further confirm that CHK1 phosphorylates FAM122A on S37 and that phosphorylation dissociates FAM122A from the PP2A complex, resulting in regulated PP2A activation.

Phosphorylated FAM122A Is Released from PP2A Complex and Binds to 14-3-3 Proteins

Previous studies have indicated that 14-3-3 proteins bind to phosphoproteins in the cytoplasm and prevent the nuclear transport of these proteins (Madeira et al., 2015; Margolis et al., 2006). To determine whether pS37-FAM122A binds to 14-3-3 protein, we next pulled down FLAG-FAM122A and immunoblotted for 14-3-3 (Figure 5E). As predicted, siRNA knockdown of CHK1 resulted in reduced coimmunoprecipitation of FAM122A and 14-3-3 (Figure 5E, lane 6), consistent with the reduced level of FAM122A phosphorylation in these cells.

We next examined the cellular localization of 14-3-3 and FAM122A in the absence or presence of CHK1 knockdown or inhibition. 14-3-3 and FAM122A were localized primarily to the cytoplasmic fraction of A549 cells (Figures S5E and S5F). Interestingly, an siRNA knockdown of CHK1 (Figure S5E) or cellular exposure to the CHK1i (Figure S5F) resulted in an increased accumulation of FAM122A in the nucleus. These results are consistent with a model in which FAM122A normally inhibits the PP2A complex in the nucleus. Activation of CHK1 results in the release of FAM122A from PP2A and the sequestration of pFAM122A by 14-3-3 in the cytoplasm. However, knockdown of CHK1 or CHK1 inhibition enhances the binding of unphosphorylated FAM122A to PP2A in the nucleus, thereby blocking PP2A-mediated dephosphorylation of WEE1 and promoting WEE1 degradation.

We next evaluated the ability of FAM122A WT versus FAM122A mutant proteins to differentially bind and coimmunoprecipitate with PP2A and 14-3-3 protein. The WT-FAM122A protein bound constitutively to PP2A or 14-3-3. Interestingly, the S37A-FAM122A mutant protein had enhanced binding to the PP2A complex but no detectable binding to the 14-3-3 protein (Figure 5F). As predicted, WT FAM122A bound to 14-3-3 (Figure 5F, lane 5), but not to PP2A. The converse was observed for the FAM122A-S37A protein (Figure 5F, lane 6). It bound to PP2A, but not to 14-3-3. Taken together, these results confirm that S37 is the critical site of CHK1-mediated phosphorylation on FAM122A and that the unphosphorylated form of FAM122A is a negative regulator of the PP2A-B55α phosphatase.

Finally, we performed a functional assessment of these two proteins (Figure 5G). A549 cells with a CRISPR knockout of FAM122A, which are resistant to CHK1is, were transiently transfected with either the FAM122A-WT or the FAM122A-S37A mutant. Both proteins were able to partially resensitize the cells to CHK1is. A model describes how CHK1 phosphorylation of FAM122A can upregulate the expression of the WEE1 protein (Figure 5H, panels I and II).

FAM122A Is a Predictive Biomarker for CHK1i Response in Cancer

Since knockout of FAM122A promotes CHK1is as ATRi resistance, we next determined whether mutation or low-level expression of FAM122A is observed in human cancers or cancer cell lines (Figure 6). CBioPortal analysis revealed only a low number of human cancer samples with bona fide FAM122A mutations (data not shown). Cancers did, however, exhibit a low level of FAM122A mRNA expression compared to normal tissue, suggesting a role of FAM122A as a tumor suppressor (Figure 6A).

We next selected ovarian and lung tumor cell lines for CHK1i resistance (Figure S6A). Interestingly, resistant lines exhibited a low level of FAM122A and a high level of WEE1 protein (Figures 6B and S6B–S6D). Consistent with these results, the knockdown of FAM122A in other cell lines, including HEPG2, two pancreatic cancer cell lines, and HeLa cells, also resulted in an increase in WEE1 protein expression levels (Figures S6E–S6H). Correction of one of these cell lines (line ES2), through transfection with WT-FAM122A cDNA, conferred partial restoration of their CHK1i sensitivity (Figures 6C and 6D). We next determined the effect of the WEE1 inhibitor AZD1775 on the CHK1i-sensitive and CHK1i-resistant ES2 ovarian cancer cells (Figures 6E and 6F). For these experiments, we used a low dose of AZD1775 (50 nM), which has little effect on either cell line as a single agent. Indeed, AZD1775 alone did not increase pKAP1 levels in either the ES2 parent cells or the ES2-resistant cells (Figure 6E, lanes 3 and 7). Prexasertib activated replication stress (pKAP1) in ES2 cells, but not in ES2-resistant cells (Figure 6E, lane 2 versus lane 6). Addition of low-dose AZD1775 in combination with a CHK1i (prexasertib) significantly increased replication stress (elevated pKAP1) and DNA damage (increased γH2AX) in ES2-resistant cells (Figure 6E, lane 8), providing a rationale for this drug combination in the clinic. Similarly, siRNA knockdown of PP2A-B55α restored replication stress (increased pKAP1) and DNA damage (elevated γH2AX) in the ES2-resistant cells exposed to prexasertib (Figure S6I, lane 8). We next analyzed the ability of the drug combination of CHK1i plus low-dose (50 nM) AZD1775 on ES2-resistant cells (Figure 6F), and this combination demonstrated significant killing of the resistant cells. Taken together, the cells showing decrease in FAM122A protein expression in response to prexasertib selection have decreased sensitivity to the CHK1i that can be overcome with the addition of a low dose of a WEE1 inhibitor.

FAM122A Is a Novel Tumor Suppressor Gene

Given that the FAM122A level correlates inversely with ATR-CHK1 pathway activity, we reasoned that cancer cells with elevated replication stress and a high dependency on *CHEK1* would exhibit a lower dependency on *FAM122A*. To test this hypothesis, we analyzed more than 500 cancer cell lines covering 20 cancer types (https://www.broadinstitute.org/ccle and https://www.depmap.org). Indeed, cell lines with high *CHEK1* dependency were less dependent on *FAM122A* for survival (Figures 7A, S7A, and S7B). There was no association between *CHEK2* dependency and *FAM122A* dependency, suggesting that this phenomenon is *CHEK1* specific (Figure 7B).

The rapid growth of tumor cells leads to more replication stress. However, a compensatory low expression of FAM122A

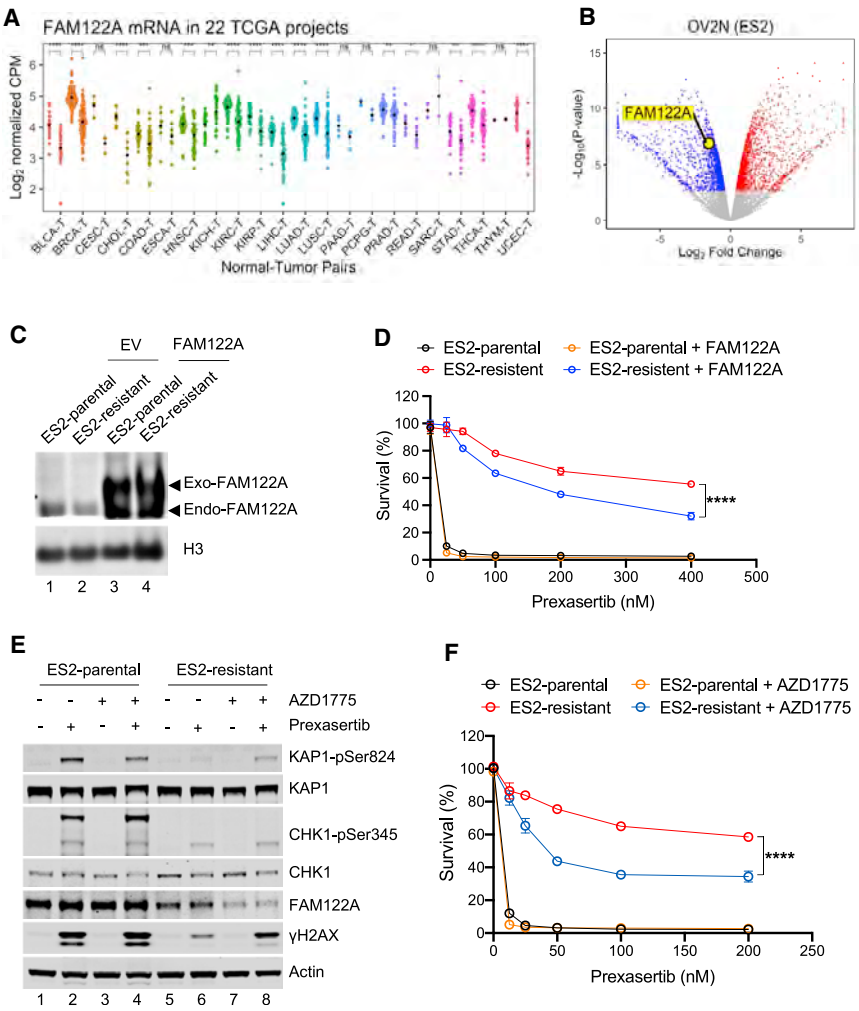


Figure 6. FAM122A Is a Predictive Biomarker for CHK1i Response in Multiple Cancer Types

(A) FAM122A RNA-sequencing (RNA-seq) expression in tumors (right) with matched normal tissues (left). The data were analyzed using 22 The Cancer Genome Atlas (TCGA) projects. Note that FAM122A expression is lower in many cancers compared to matched normal tissues.

(B) Volcano plot showing genes differentially downregulated or upregulated in prexasertib-resistant ES2 ovarian cells (OV2N ES2) compared to parental cells. Note that FAM122A expression is significantly decreased in ES2 cells with acquired resistance to prexasertib.

(C) Western blots of the lysates from stably expressing FAM122A-FLAG or FLAG-control (empty vector [EV]) parental and prexasertib-resistant ES2 cells. Note that anti-H3 was used as a protein loading control.

(D) Survival plots of stably expressing FAM122A-FLAG or FLAG-control parental and prexasertib-resistant ES2 cells after treatment with graded concentrations of prexasertib for 3 days. Data are shown as mean ± SD from three independent experiments. ES2-resistant versus ES2-resistant+FAM122A, ****$p < 0.0001$; statistical analysis was performed using two-way ANOVA.

(E) Western blots of the lysates from parental and prexasertib-resistant ES2 cells after treatment with prexasertib (50 nM) and AZD1775 (50 nM) for 24 h.

(F) Survival plots of parental and prexasertib-resistant ES2 cells after treatment AZD1775 (50 nM) and graded concentrations of prexasertib for 3 days. Data are shown as mean ± SD from three independent experiments. ES2-resistant versus ES2-resistant+AZD1775, ****$p < 0.0001$; statistical analysis was performed using two-way ANOVA.

in these tumor cells would reduce the replication stress and promote tumor growth. Consistent with this concept, in multiple cancer types, patients whose tumors have low *FAM122A* expression have significantly worse overall survival than patients whose tumors have high *FAM122A* expression (Figure 7C).

DISCUSSION

The activation of the ATR-CHK1-WEE1 pathway is a critical upstream event in the G2/M checkpoint. Cancer cells, due to their intrinsic high levels of replication stress, are hyperdependent on this pathway. Accordingly, inhibitors of ATR, CHK1, or WEE1 are under active investigation in anti-cancer clinical trials (Matheson et al., 2016; Qiu et al., 2018; Weber and Ryan, 2015), with early promising results (Fang et al., 2019). Inhibition of the ATR-CHK1-WEE1 pathway disrupts the G2/M checkpoint and promotes rapid exit into mitosis, and cancer cell death.

Role of FAM122A in the Regulation of the G2/M Checkpoint

In the current study, we used a CRISPR screen to uncover a novel mechanism of CHK1i resistance. We identified FAM122A as a novel physiologic regulator of the ATR-CHK1-WEE1 pathway. In the normal cellular setting, FAM122A binds and inhibits the heterotrimeric PP2A complex containing the B55α subunit, resulting in limited dephosphorylation of WEE1. In this setting, WEE1 levels are maintained at a low level. When CHK1 is activated by replication stress, FAM122A is phosphorylated and released from PP2A. In cancer cells, replication stress can result from any one of several events, including p53 loss, MYC amplification, cyclin E amplification, or loss of a homologous recombination (HR) gene (Zeman and Cimprich, 2014). The activated PP2A now dephosphorylates WEE1, resulting in increased WEE1 expression and contributing to the G2/M checkpoint (Figure 5H, panel I). An inhibitor of CHK1, such as prexasertib, blocks the phosphorylation of FAM122A, resulting in reduced PP2A activity and increased WEE1 degradation (Figure 5H, panel II). Thus, the CHK1i further increases replication stress in the cancer cells, thereby promoting cancer cell death.

Knockout of FAM122A results in increased PP2A-mediated dephosphorylation of WEE1, stabilization and increase levels of WEE1 protein, and restoration of the G2/M checkpoint. Indeed, cancer cells often reduce the expression of FAM122A by genetic or epigenetic means, thereby reducing their replication stress and

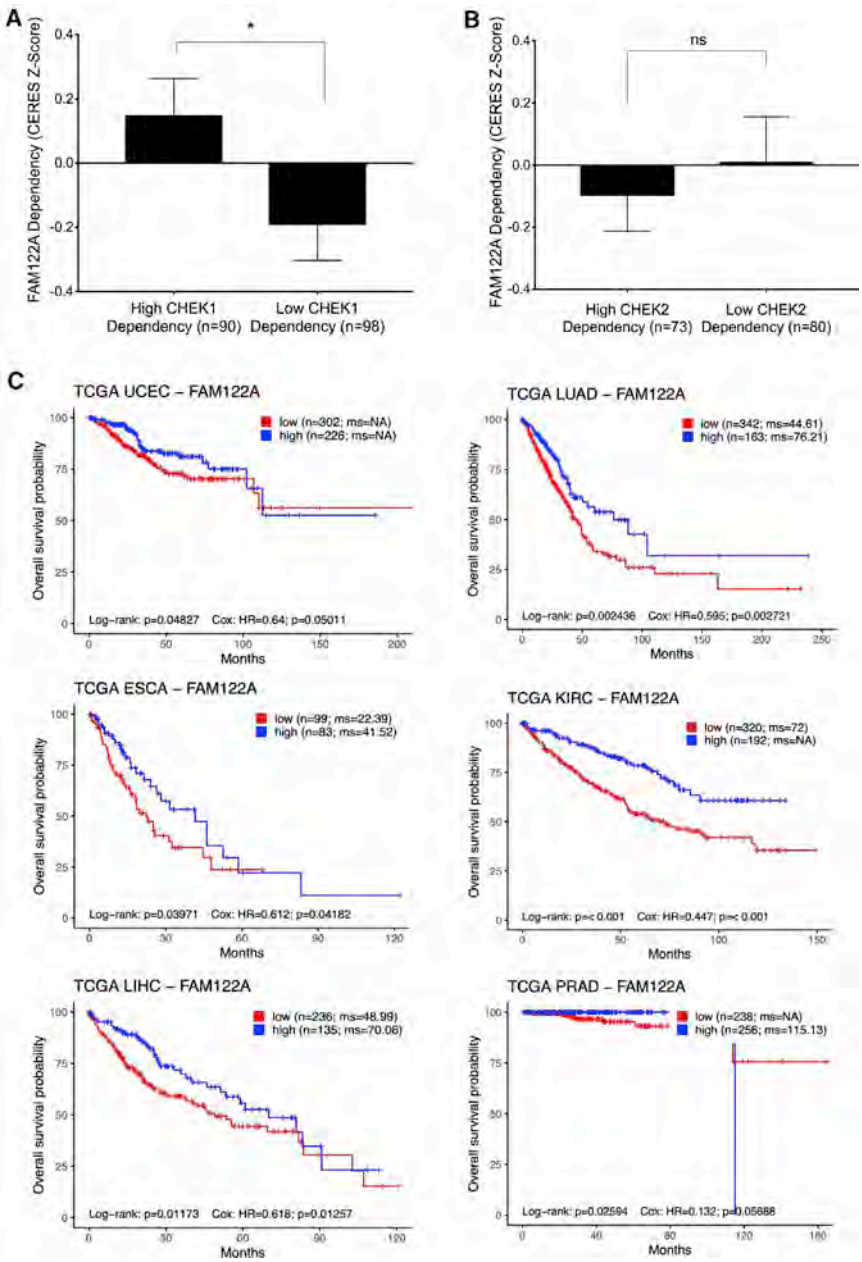


Figure 7. FAM122A Is a Novel Tumor Suppressor Gene

(A) *CHEK1* dependency inversely correlates with *FAM122A* dependency. Cell lines that were 1 SD below or above the mean in terms of *CHEK1* CERES score were grouped into cell lines that have high *CHEK1* dependency and low *CHEK1* dependency, respectively. A lower CERES score indicates a cell line is more dependent on that gene.

(B) Absence of an association between *CHEK2* dependency and *FAM122A* dependency. Cell lines that were 1 SD below or above the mean in terms of *CHEK2* CERES score were grouped into cell lines that have high *CHEK2* dependency and low *CHEK2* dependency, respectively.

(C) Kaplan-Meier survival curves of patients with indicated cancers stratified by FAM122A gene expression level using TCGA data. The log-rank test was used to assess statistical significance.

allowing them to survive their high replication rate. Adding a PP2A inhibitor, such as OA, in this setting could block WEE1 dephosphorylation and reduce WEE1 protein levels, thereby resensitizing the tumor cells to prexasertib. Of note, the CRISPR screen identified an endogenous inhibitor of PP2A-B55α as opposed to inhibitors of other heterotrimeric PP2A complexes containing different B subunits such as ENSA. This implies that targets of PP2A-B55α, including WEE1, are involved in CHK1i resistance conferred by FAM122A loss, as opposed to protein targets of alternate PP2A heterotrimeric complexes. Development of pharmacologic inhibitors specifically of PP2A-B55α as opposed to PP2A more broadly given its myriad roles would be anticipated to reverse CHK1i resistance with lower toxicity.

Role of WEE1 in the Regulation of the G2/M Checkpoint

Our results indicate that CHK1 enhances the WEE1-mediated checkpoint via at least two distinct mechanisms. First, previous studies have shown that CHK1 directly phosphorylates WEE1 (Lee et al., 2001; O'Connell et al., 1997). Activated WEE1 in turn phosphorylates the CDK1/cyclin B complex and promotes the G2/M checkpoint. Second, through its phosphorylation of FAM122A and release of the FAM122A/PP2A-B55α complex, CHK1 enhances PP2A activity and reduces WEE1 degradation by ubiquitin-mediated degradation. Enhanced WEE1 protein levels further reduce replication stress and augment the G2/M checkpoint. Paradoxically, before mitotic initiation, PP2A-B55α can also dephosphorylate WEE1 and reduce its kinase activity (Harvey et al., 2011). PP2A-B55α can also dephosphorylate and subsequently inactivate CDC25 as cells exit mitosis (Forester et al., 2007; Johnson and Kornbluth, 2012). Hence, PP2A has multiple substrates and may function as both a positive and negative regulator of the G2/M checkpoint.

Effective Anti-cancer Therapy Increases the Replication Stress of Cancer Cells

Our results have important implications for the evolution of cancer and cancer drug development. Cancer cells tend to reduce their expression of FAM122A by either genetic or epigenetic mechanisms. The reduction in FAM122A allows the cells to increase their expression level of WEE1 protein, to enhance their G2/M arrest, and to tolerate the replication stress intrinsic to their oncogene-driven high replication rate. The purpose of inhibiting the ATR-CHK1-WEE1 pathway in cancer therapy is to further increase

replication stress of the cancer cells, to put the cells over the threshold of replication stress, and to kill the cells. A combination of a kinase inhibitor of ATR, CHK1, or WEE1 with a drug that inhibits the PP2A enzyme will further increase replication stress but in many cases also be toxic to normal cells. Since cancer cells have elevated replication stress to begin with, this strategy exploits the underlying therapeutic index of normal versus transformed cells.

Biomarker-Driven Development of New CHK1is

Our study suggests predictive and pharmacodynamic biomarkers for these ongoing trials. Tumors with high baseline levels of replication stress, as measured by high pKAP1 and pRPA, are more likely to respond to these inhibitors. Tumors with low levels of FAM122A are less likely to respond. FAM122A is a critical regulator of PP2A-B55α phosphorylation activity. Based on its cellular functions, we propose changing the name of FAM122A to PABIR1, standing for PP2A alpha (PPP2R1A) and B55A (PPP2R2A) interacting phosphatase regulator 1. Finally, FAM122A plays a normal function in non-transformed cells. FAM122A likely functions to suppress PP2A-B55α and promote the phosphorylation and degradation of WEE1 in these cells as well. Under conditions of elevated replication stress, in cancer cells exposed to DNA-damaging agents, CHK1 phosphorylates FAM122A. Phosphorylated FAM122A is sequestered by 14-3-3 proteins in the cytoplasm, thereby freeing active PP2A-B55α to rescue WEE1 protein levels from degradation. Identification of tumors that are the most likely to respond to CHK1is as monotherapy, or the development of combination therapies that overcome potential mechanisms of CHK1i resistance, will hopefully increase the likelihood that CHK1is will have a durable role in the treatment of cancers.

Limitations of Study

Finally, our study has limitations. First, CHK1 kinase has multiple substrates in addition to FAM122A. Indeed, cellular knockout of FAM122A results in a reduction of replication stress and an increase in the G2/M checkpoint. However, the relative contribution of FAM122A compared to other CHK1 substrates with respect to these cellular activities remains largely unknown. Second, our studies indicate that knockout of FAM122A results in resistance to either a CHK1i or an ATR inhibitor. How FAM122A knockout confers resistance to an ATR inhibitor will require additional experimentation. Third, our results suggest that a combination of a CHK1i plus a WEE1 inhibitor may override the cellular resistance to CHK1i monotherapy. However, this drug combination may prove to be too toxic for use in a clinical setting. Finally, our study focused mostly on cancer cell lines, especially lung cancer cell lines. It will be important to determine whether FAM122A knockdown is also an *in vivo* mechanism of acquired CHK1i resistance of tumors isolated from treated cancer patients.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell Lines
- METHOD DETAILS
  - Generation of Stable Cell Lines
  - Drug Selection CRISPR Screens
  - Genomic DNA Preparation, Sequencing, and Identification of Top Hits
  - Western Blotting
  - CRISPR cas9 Knockout
  - siRNA Transfection
  - Clonogenic Assay
  - CellTiter-Glo Assay and Synergy Study
  - Immunofluorescence
  - Measuring Cell Cycle and Rate of DNA Replication
  - Cytogenetics
  - Subcellular Fractionation
  - Alkaline Single-Cell Electrophoresis (Comet) Assays
  - DNA Fiber Assay
  - Kinase Reactions
  - RNA Sequencing and Analysis
  - Computational Analysis
- QUANTITATION AND STATISTICAL ANALYSIS
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.molcel.2020.10.008.

ACKNOWLEDGMENTS

We thank all members of the D'Andrea laboratory for their helpful suggestions and comments. This research was supported by grants from the National Institutes of Health (R37HL052725 and P01HL048546), the Department of Defense (BM110181 and BC151331P1), the Breast Cancer Research Foundation, and the Fanconi Anemia Research Fund (to A.D.D.). This work was supported in part by Public Health Service grants R35CA232128 and P01CA203655 to J.A.D., K08CA172354 to D.K., and F31CA189328 to T.B. This work was also supported by the Joint Center for Radiation Therapy (to D.K.) and by a sponsored research agreement from Eli Lilly (to A.D.D. and G.I.S.).

AUTHOR CONTRIBUTIONS

F.L., D.K., J.D., G.I.S, D.C., and A.D.D. conceived the study, analyzed the data, and wrote the manuscript. P.D., T.B., and X.-F.Z. performed the experiments and computational analysis, analyzed the data, and helped write the manuscript. P.D. and K.P. performed the experiments and analyzed the data. D.K., H.N., and C.D. performed the computational analysis, analyzed the data, and helped write the manuscript. J.D. and C.D. contributed to writing the manuscript. D.K. helped with the design of the CRISPR screen and performed bioinformatics analysis.

DECLARATION OF INTERESTS

D.K. has served as a consultant to Vertex. J.D. has served as a consultant to Merck & Co. and EMD Serono and has received research funding from Constellation Pharmaceuticals. G.I.S. is a consultant/advisory board member for Lilly, Sierra Oncology, Merck-EMD Serono, Pfizer, Astex, Almac, Roche,

Bicycle Therapeutics, Fusion Pharmaceuticals, G1 Therapeutics, Bayer, Ipsen, Cybrexa Therapeutics, Angiex, Daiichi Sankyo, and Seattle Genetics and reports receiving commercial research grants from Lilly, Sierra Oncology, Merck-EMD Serono, and Merck & Co. A.D.D. is a consultant and/or advisory board member for AstraZeneca, Bayer AG, Cedilla Therapeutics, Celgene, Cyteir Therapeutics, Epizyme, GalaxoSmithKline, Ideaya, Impact Therapeutics, LAV Global Management Company Limited, L.E.K. Consulting, Lilly Oncology, Merck KGAa, Novartis Bioventures, Novo Ventures, Pfizer, Scholar Rock, and Tango Therapeutics; is a stockholder in Ideaya, Cedilla Therapeutics, and Cyteir; and reports receiving commercial research grants from Lilly Oncology and Merck KGAa.

Received: April 17, 2020
Revised: July 14, 2020
Accepted: October 4, 2020
Published: October 26, 2020

REFERENCES

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.

Cai, M.Y., Dunn, C.E., Chen, W., Kochupurakkal, B.S., Nguyen, H., Moreau, L.A., Shapiro, G.I., Parmar, K., Kozono, D., and D'Andrea, A.D. (2020). Cooperation of the ATM and Fanconi anemia/BRCA pathways in double-strand break end resection. Cell Rep. *30*, 2402–2415.e2405.

Ceccaldi, R., Liu, J.C., Amunugama, R., Hajdu, I., Primack, B., Petalcorin, M.I., O'Connor, K.W., Konstantinopoulos, P.A., Elledge, S.J., Boulton, S.J., et al. (2015). Homologous-recombination-deficient tumours are dependent on Polθ-mediated repair. Nature *518*, 258–262.

Cornwell, M., Vangala, M., Taing, L., Herbert, Z., Köster, J., Li, B., Sun, H., Li, T., Zhang, J., Qiu, X., et al. (2018). VIPER: Visualization Pipeline for RNA-seq, a Snakemake workflow for efficient and complete RNA-seq analysis. BMC Bioinformatics *19*, 135.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat. Biotechnol. *34*, 184–191.

Duong-Ly, K.C., and Peterson, J.R. (2013). The human kinome and kinase inhibition. Curr. Protoc. Pharmacol. *Chapter 2*, Unit2.9.

Fan, L., Liu, M.H., Guo, M., Hu, C.X., Yan, Z.W., Chen, J., Chen, G.Q., and Huang, Y. (2016). FAM122A, a new endogenous inhibitor of protein phosphatase 2A. Oncotarget *7*, 63887–63900.

Fang, Y., McGrail, D.J., Sun, C., Labrie, M., Chen, X., Zhang, D., Ju, Z., Vellano, C.P., Lu, Y., Li, Y., et al. (2019). Sequential therapy with PARP and WEE1 inhibitors minimizes toxicity while maintaining efficacy. Cancer Cell *35*, 851–867.e857.

Forester, C.M., Maddox, J., Louis, J.V., Goris, J., and Virshup, D.M. (2007). Control of mitotic exit by PP2A regulation of Cdc25C and Cdk1. Proc. Natl. Acad. Sci. USA *104*, 19867–19872.

Harvey, S.L., Enciso, G., Dephoure, N., Gygi, S.P., Gunawardena, J., and Kellogg, D.R. (2011). A phosphatase threshold sets the level of Cdk1 activity in early mitosis in budding yeast. Mol. Biol. Cell *22*, 3595–3608.

Hong, D., Infante, J., Janku, F., Jones, S., Nguyen, L.M., Burris, H., Naing, A., Bauer, T.M., Piha-Paul, S., Johnson, F.M., et al. (2016). Phase I study of LY2606368, a checkpoint kinase 1 inhibitor, in patients with advanced cancer. J. Clin. Oncol. *34*, 1764–1771.

Hong, D.S., Moore, K., Patel, M., Grant, S.C., Burris, H.A., 3rd, William, W.N., Jr., Jones, S., Meric-Bernstam, F., Infante, J., Golden, L., et al. (2018). Evaluation of prexasertib, a checkpoint kinase 1 inhibitor, in a phase Ib study of patients with squamous cell carcinoma. Clin. Cancer Res. *24*, 3263–3272.

Janghorban, M., Farrell, A.S., Allen-Petersen, B.L., Pelz, C., Daniel, C.J., Oddo, J., Langer, E.M., Christensen, D.J., and Sears, R.C. (2014). Targeting c-MYC by antagonizing PP2A inhibitors in breast cancer. Proc. Natl. Acad. Sci. USA *111*, 9157–9162.

Janssens, V., and Goris, J. (2001). Protein phosphatase 2A: a highly regulated family of serine/threonine phosphatases implicated in cell growth and signalling. Biochem. J. *353*, 417–439.

Janssens, V., Longin, S., and Goris, J. (2008). PP2A holoenzyme assembly: in cauda venenum (the sting is in the tail). Trends Biochem. Sci. *33*, 113–121.

Johnson, E.S., and Kornbluth, S. (2012). Phosphatases driving mitosis: pushing the gas and lifting the brakes. Prog. Mol. Biol. Transl. Sci. *106*, 327–341.

Junttila, M.R., Puustinen, P., Niemelä, M., Ahola, R., Arnold, H., Böttzauw, T., Ala-aho, R., Nielsen, C., Ivaska, J., Taya, Y., et al. (2007). CIP2A inhibits PP2A in human malignancies. Cell *130*, 51–62.

Khanna, A., Kauko, O., Böckelman, C., Laine, A., Schreck, I., Partanen, J.I., Szwajda, A., Bormann, S., Bilgen, T., Helenius, M., et al. (2013). Chk1 targeting reactivates PP2A tumor suppressor activity in cancer cells. Cancer Res. *73*, 6757–6769.

Kim, J.W., Berrios, C., Kim, M., Schade, A.E., Adelmant, G., Yeerna, H., Damato, E., Iniguez, A.B., Florens, L., Washburn, M.P., et al. (2020). STRIPAK directs PP2A activity toward MAP4K4 to promote oncogenic transformation of human cells. eLife *9*, 9.

King, C., Diaz, H.B., McNeely, S., Barnard, D., Dempsey, J., Blosser, W., Beckmann, R., Barda, D., and Marshall, M.S. (2015). LY2606368 causes replication catastrophe and antitumor effects through CHK1-dependent mechanisms. Mol. Cancer Ther. *14*, 2004–2013.

Kinoshita, E., Kinoshita-Kikuta, E., and Koike, T. (2015). Advances in Phos-tag-based methodologies for separation and detection of the phosphoproteome. Biochim. Biophys. Acta *1854*, 601–608.

Lee, J., Kumagai, A., and Dunphy, W.G. (2001). Positive regulation of Wee1 by Chk1 and 14-3-3 proteins. Mol. Biol. Cell *12*, 551–563.

Leung-Pineda, V., Ryan, C.E., and Piwnica-Worms, H. (2006). Phosphorylation of Chk1 by ATR is antagonized by a Chk1-regulated protein phosphatase 2A circuit. Mol. Cell. Biol. *26*, 7529–7538.

Lim, K.S., Li, H., Roberts, E.A., Gaudiano, E.F., Clairmont, C., Sambel, L.A., Ponnienselvan, K., Liu, J.C., Yang, C., Kozono, D., et al. (2018). USP1 is required for replication fork protection in BRCA1-deficient tumors. Mol. Cell *72*, 925–941.e924.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Madeira, F., Tinti, M., Murugesan, G., Berrett, E., Stafford, M., Toth, R., Cole, C., MacKintosh, C., and Barton, G.J. (2015). 14-3-3-Pred: improved methods to predict 14-3-3-binding phosphopeptides. Bioinformatics *31*, 2276–2283.

Margolis, S.S., Perry, J.A., Forester, C.M., Nutt, L.K., Guo, Y., Jardim, M.J., Thomenius, M.J., Freel, C.D., Darbandi, R., Ahn, J.H., et al. (2006). Role for the PP2A/B56delta phosphatase in regulating 14-3-3 release from Cdc25 to control mitosis. Cell *127*, 759–773.

Matheson, C.J., Backos, D.S., and Reigan, P. (2016). Targeting WEE1 kinase in cancer. Trends Pharmacol. Sci. *37*, 872–881.

Meyers, R.M., Bryan, J.G., McFarland, J.M., Weir, B.A., Sizemore, A.E., Xu, H., Dharia, N.V., Montgomery, P.G., Cowley, G.S., Pantel, S., et al. (2017). Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. Nat. Genet. *49*, 1779–1784.

Mochida, S., Maslen, S.L., Skehel, M., and Hunt, T. (2010). Greatwall phosphorylates an inhibitor of protein phosphatase 2A that is essential for mitosis. Science *330*, 1670–1673.

O'Connell, M.J., Raleigh, J.M., Verkade, H.M., and Nurse, P. (1997). Chk1 is a wee1 kinase in the G2 DNA damage checkpoint inhibiting cdc2 by Y15 phosphorylation. EMBO J. *16*, 545–554.

Parmar, K., Kochupurakkal, B.S., Lazaro, J.B., Wang, Z.C., Palakurthi, S., Kirschmeier, P.T., Yang, C., Sambel, L.A., Färkkilä, A., Reznichenko, E., et al. (2019). The CHK1 inhibitor prexasertib exhibits monotherapy activity in

high-grade serous ovarian cancer models and sensitizes to PARP inhibition. Clin. Cancer Res. *25*, 6127–6140.

Patil, M., Pabla, N., and Dong, Z. (2013). Checkpoint kinase 1 in DNA damage response and cell cycle regulation. Cell. Mol. Life Sci. *70*, 4009–4021.

Paulsen, R.D., and Cimprich, K.A. (2007). The ATR pathway: fine-tuning the fork. DNA Repair (Amst.) *6*, 953–966.

Qiu, Z., Oleinick, N.L., and Zhang, J. (2018). ATR/CHK1 inhibitors and cancer therapy. Radiother. Oncol. *126*, 450–464.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Sangodkar, J., Farrington, C.C., McClinch, K., Galsky, M.D., Kastrinsky, D.B., and Narla, G. (2016). All roads lead to PP2A: exploiting the therapeutic potential of this phosphatase. FEBS J. *283*, 1004–1024.

Sanjiv, K., Hagenkort, A., Calderón-Montaño, J.M., Koolmeister, T., Reaper, P.M., Mortusewicz, O., Jacques, S.A., Kuiper, R.V., Schultz, N., Scobie, M., et al. (2016). Cancer-specific synthetic lethality between ATR and CHK1 kinase activities. Cell Rep. *14*, 298–309.

Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelson, T., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G., and Zhang, F. (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. Science *343*, 84–87.

Watanabe, N., Arai, H., Nishihara, Y., Taniguchi, M., Watanabe, N., Hunter, T., and Osada, H. (2004). M-phase kinases induce phospho-dependent ubiquitination of somatic Wee1 by SCFbeta-TrCP. Proc. Natl. Acad. Sci. USA *101*, 4419–4424.

Watanabe, N., Arai, H., Iwasaki, J., Shiina, M., Ogata, K., Hunter, T., and Osada, H. (2005). Cyclin-dependent kinase (CDK) phosphorylation destabilizes somatic Wee1 via multiple pathways. Proc. Natl. Acad. Sci. USA *102*, 11663–11668.

Weber, A.M., and Ryan, A.J. (2015). ATM and ATR as therapeutic targets in cancer. Pharmacol. Ther. *149*, 124–138.

Wlodarchak, N., and Xing, Y. (2016). PP2A as a master regulator of the cell cycle. Crit. Rev. Biochem. Mol. Biol. *51*, 162–184.

Yan, Y., Cao, P.T., Greer, P.M., Nagengast, E.S., Kolb, R.H., Mumby, M.C., and Cowan, K.H. (2010). Protein phosphatase 2A has an essential role in the activation of gamma-irradiation-induced G2/M checkpoint response. Oncogene *29*, 4317–4329.

Zeman, M.K., and Cimprich, K.A. (2014). Causes and consequences of replication stress. Nat. Cell Biol. *16*, 2–9.

Zhao, H., and Piwnica-Worms, H. (2001). ATR-mediated checkpoint pathways regulate phosphorylation and activation of human Chk1. Mol. Cell. Biol. *21*, 4129–4139.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
FAM122A	Invitrogen	Cat# MA5-24510; RRID: AB_2633482
β-actin	Cell Signaling Technology	Cat# 4967S; RRID: AB_330288
Vinculin	Cell Signaling Technology	Cat# 13901S; RRID: AB_2728768
KAP1	Abcam	Cat# ab22553; RRID: AB_447151
KAP1-phospho-S824	Abcam	Cat# ab84077; RRID: AB_1861686
CHK1	Cell Signaling Technology	Cat# 2360S; RRID: AB_2080320
CHK1-phospho-S345	Cell Signaling Technology	Cat# 2348S; RRID: AB_2080326
WEE1	Cell Signaling Technology	Cat# 4936S; RRID: AB_2288509
WEE1-phospho-S53	Bioss	Cat# bs-5589R; RRID: AB_11075853
WEE1-phospho-S123	Bioss	Cat# BS-5218R; RRID: AB_11111792
CDK1	Abcam	Cat# ab18; RRID: AB_2074906
CDK1-phospho-Y15	Cell Signaling Technology	Cat# 9111S; RRID: AB_331460
PP2A-B55-alpha	Sigma	Cat# SAB4300573; RRID: AB_10622182
PP2A-B55-alpha	Santa Cruz Biotechnology	Cat# SC-81606; RRID: AB_2252935
HA-Tag	Cell Signaling Technology	Cat# 3724S; RRID: AB_1549585
Flag	Sigma	Cat# F3165; RRID: AB_259529
Flag	Sigma	Cat# F7425; RRID: AB_439687
GST	Cell Signaling Technology	Cat# 2625S; RRID: AB_490796
14-3-3	Cell Signaling Technology	Cat# 8312S; RRID: AB_10860606
GAPDH	Cell Signaling Technology	Cat# 5174S; RRID: AB_10622025
γ-H2AX	Cell Signaling Technology	Cat# 2577S; RRID: AB_2118010
BrdU	Cell Signaling Technology	Cat# 5292S; RRID: AB_10548898
BrdU	Abcam	Cat# ab6326; RRID: AB_305426
Histone H3	Cell Signaling Technology	Cat# 9715S; RRID: AB_331563
Goat anti-Rat IgG (Alexa Fluor 488)	Invitrogen	Cat# A-11006; RRID: AB_2534074
Goat anti-Mouse IgG (Alexa Fluor 488)	Invitrogen	Cat# A32723; RRID: AB_2633275
Chicken anti-mouse IgG (Alexa Fluor 594)	Invitrogen	Cat# A-21201; RRID: AB_141630
Chemicals, Peptides, and Recombinant Proteins		
Prexasertib	Eli Lilly & Co.	Cat# S7178
SRA737	Sierra Oncology	Cat# S8253
VE822	Selleckchem	Cat# S7102
AZD1775	Selleckchem	Cat# S1525
LB-100	Selleckchem	Cat# S7537
5-Chloro-2′-deoxyuridine (CldU)	Sigma Aldrich	Cat# C6891
5-Iodo-2′-deoxyuridine (IdU)	Sigma Aldrich	Cat# I7125
EdU (5-ethynyl-2′-deoxyuridine)	ThermoFisher	Cat# E10415
Cycloheximide	Sigma Aldrich	Cat# 01810
Critical Commercial Assays		
CellTiter-Glo Luminescent Viability Assay	Promega	Cat# G7570
DNeasy Blood & Tissue Kit	QIAGEN	Cat# 69504
Click-iT EdU Imaging Kit with Alexa Fluor 488	ThermoFisher	Cat# C10086
PowerUp SYBR Green Master Mix	Life Technologies	Cat# A25741

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Subcellular Fractionation Kit	Cell Signaling Technology	Cat# 9038
Deposited Data		
Raw imaging data	This paper; Mendeley data	https://data.mendeley.com/datasets/ndvs5mf8hx/1
RNA-seq data	This paper	GEO: GSE158338
Experimental Models: Cell Lines		
NCI-H520	ATCC	N/A
HCC827	ATCC	N/A
NCI-H1563	ATCC	N/A
NCI-H2087	ATCC	N/A
SKLU-1	ATCC	N/A
NCI-H1915	ATCC	N/A
CALU-6	ATCC	N/A
NCI-H1299	ATCC	N/A
NCI-H522	ATCC	N/A
NCI-H23	ATCC	N/A
A549	ATCC	N/A
NCI-H460	ATCC	N/A
U2OS	ATCC	N/A
HeLa	ATCC	N/A
HepG2	ATCC	N/A
MIAPaCa	Gift from Dr. Andrew Aguirre	N/A
PaTu8902	Gift from Dr. Andrew Aguirre	N/A
293T	ATCC	N/A
ES2	ATCC	N/A
A2780	Ceccaldi et al., 2015	N/A
Oligonucleotides		
siFAM122A	Dharmacon	E-015499-00
siPPP2CA	QIAGEN	SI00041853
siPPP2R2A	QIAGEN	SI00041909
siPPP2R5A	QIAGEN	SI02225846
siPPP2R5D	QIAGEN	SI00086240
siCHEK1	QIAGEN	SI02660007
siENSA	QIAGEN	SI00063581
AllStars Neg. Control siRNA	QIAGEN	0001027281
siBRCA1	QIAGEN	SI02654575
siTP53BP1:	QIAGEN	SI02653168
FAM122A-qPCR-Forward Primer: AAGATGGAGCTAGACCTGGAG	eurofins	N/A
FAM122A-qPCR-Reverse Primer: CCGGCGAAGTGTCACTGAG	eurofins	N/A
gRNA targeting FAM122A #1: GCAGATAAGCCACTCCTGGG	Invitrogen	N/A
gRNA targeting FAM122A #2: CAGTGACACTTCGCCGGTGT	Invitrogen	N/A
Software and Algorithms		
Prism8 GraphPad software	N/A	https://www.graphpad.com/
imageJ software	N/A	https://imagej.nih.gov/ij/

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Alan D. D'Andrea (Alan_Dandrea@dfci.harvard.edu).

Materials Availability

The knockout cell lines generated in this study are available from the Lead Contact without restriction.

Data and Code Availability

All the original data for immunofluorescence and western blots have been deposited with Mendeley and can be accessed with https://data.mendeley.com/datasets/ndvs5mf8hx/1. The STARS analysis for the CHKi (Prexasertib) genome-wide CRISPR screens in A549 and H460 cells, positive selection hits, is available in Table S1. The STARS analysis for the CHKi (Prexasertib) genome-wide CRISPR screens in A549 and H460 cells, negative selection hits, is available in Table S2. edgeR Analysis for the CHKi (Prexasertib) Genome-wide CRISPR Screens in A549 and H460 cells is available in Table S3. The RNA-seq datasets discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (PMID: 11752295) and are accessible through GEO Series accession number GSE158338.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell Lines

A549 and NCI-H460 cells stably expressing the Cas9 endonuclease were obtained from the Genetic Perturbation Platform at the Broad Institute and were cultured at 37°C/5% CO_2 incubator in RPMI 1640 (GIBCO) media supplemented with 10% fetal bovine serum (FBS, Sigma), 1% penicillin-streptomycin (GIBCO). U2OS and HEK293T cells were cultured at 37°C/5% CO_2 in DMEM (high glucose, Hyclone) media supplemented with 10% fetal bovine serum (Hyclone) and 1% penicillin-streptomycin (GIBCO). Other human NSCLC cell lines A549, Calu-6, HCC827, NCI-H23, NCI-H460, NCI-H520, NCI-H522, NCI-H1299, NCI-H1563, NCI-H1915, NCI-H2087, and SKLU-1 were grown at 37°C in a humidified 5% CO_2 incubator in GIBCO RPMI 1640 (GIBCO) supplemented with 10% fetal bovine serum (FBS, Sigma) and 1 μg/mL Normocin (GIBCO). Each of these cell lines were obtained from the American Type Culture Collection (ATCC, Manassas, VA) and validated by short tandem repeat (STR) testing.

METHOD DETAILS

Generation of Stable Cell Lines

For generation of cells with stable expression of exogenous genes, the constructs were transfected into 293T cells for retroviral packaging and subsequent transduction. For generation of prexasertib-resistant cell lines, cells were cultured with graded concentrations of prexasertib for one week and surviving cells in the drug concentrations of IC_{90} or IC_{80} were harvested. These cells were then cultured in the IC_{50} drug concentrations for 1-2 weeks and were gradually adapted to higher doses of drug for almost 2-3 months. The resistant cells were then maintained in a medium without the drug.

Drug Selection CRISPR Screens

NSCLC cell lines A549 and NCI-H460 stably expressing the Cas9 endonuclease (pXPR_311) were obtained from the Genetic Perturbation Platform at the Broad Institute. The lines were grown at 37°C in a humidified 5% CO_2 incubator in GIBCO RPMI 1640 (Life Technologies, Grand Island, NY) supplemented with 10% fetal bovine serum (FBS, Sigma-Aldrich, St. Louis, MO) and 100 μg/ml Normocin (InvivoGen, San Diego, CA). The pXPR_311 plasmid contains a blasticidin resistance tag to allow for selection of infected cells. For each line, the selected Cas9 positive cells were infected in four 12-well tissue culture plates with the Brunello whole genome sgRNA pooled lentiviral library consisting of 76,441 sgRNAs targeting 19,114 different genes (Doench et al., 2016). For adequate representation of each sgRNA, e.g., an average of 500 cells/sgRNA, a total of 135 million cells divided into wells each containing 3 million cells were infected at a multiplicity of infection (MOI) of 0.3 in media containing 8 μg/ml polybrene. Plates were centrifuged at 2000 rpm for 2 h at 37°C, and then 2 h after centrifugation, media was replaced with fresh media without polybrene and cells were incubated overnight. Infected cells were transferred 15 cm tissue culture dishes and selected with 2 μg/mL puromycin for 4 days prior to drug treatment. After selection, 40 million cells (8 plates each containing 5 million cells) were plated for each biological replicate, and an additional 40 million cells per replicate were collected and frozen for subsequent genomic DNA (gDNA) extraction. A549 and NCI-H460 cells were then treated with 100nM (A549) or 50nM (NCI-H460) prexasertib versus DMSO control, for 19 days. Cells were re-plated every 3-4 days and maintained in drug or DMSO-containing media, each time plating a minimum of 40 million cells per replicate. At the end of the screen, the cells were harvested and frozen for gDNA extraction.

Cells were collected pre- and post-drug selection, genomic DNA was extracted, and sgRNA sequences were PCR amplified and Illumina sequenced. Top hits among the gene knockouts that conferred drug resistance were identified by the STARS gene-ranking

algorithm (Doench et al., 2016). Validation of hits was performed with sgRNAs (Broad Institute) and pooled siRNAs (QIAGEN). LY2606368 IC_{50}s were assessed using colony formation and CellTiter-Glo cell viability assays.

Genomic DNA Preparation, Sequencing, and Identification of Top Hits

gDNA was isolated using QIAGEN QIAamp DNA blood midi and maxi kits, according to manufacturer's protocol. PCR of gDNA was performed to attach sequencing adaptors and barcode samples for sgRNA sequencing, as previously described (Doench et al., 2016). Samples were sequenced on an Illumina HiSeq2000. Reads per million were $\log_2$-transformed by first adding one to all values, which was necessary to take the log of sgRNAs with zero reads. For analysis, these normalized data were used to determine the $\log_2$-fold change of each sgRNA for the drug treatment group relative to DMSO control group (Shalem et al., 2014). The STARS gene-ranking algorithm was used to determine the top hits (Tables S1 and S2) (https://portals.broadinstitute.org/gpp/public/software/stars); at least two sgRNAs targeting a gene had to rank within the top 10% of sgRNAs by $\log_2$-fold change in one direction or another (Doench et al., 2016). The hit genes were scored using the STARS score and the false discovery rate. Additionally, edgeR package was used to perform guide-level analysis (Tables S1, S2, and S3). The volcano plots of these results were created using R (https://www.r-project.org).

Western Blotting

Cells were lysed with lysis buffer (300 mM NaCl, 50 mM Tris-Cl, 1 mM EDTA, 0.5% NP-40), lysates were resolved on denaturing Nu-page (Invitrogen) polyacrylamide gels and transferred onto nitrocellulose membranes. Membranes were blocked with 5% milk in TBST, and probed with primary and secondary antibodies respectively, then detected with chemiluminescence (Western Lightning, Perkin Elmer).

CRISPR cas9 Knockout

To generate A549, NCI-H460, HeLa and U2OS cells expressing the indicated sgRNAs, the listed oligonucleotides were cloned into the pLentiCRISPR V2 viral vector.

siRNA Transfection

siRNA negative control and siRNA targeting FAM122A, PP2ACa, ENSA, BRCA1 and 53BP1 were ON-TARGET plus siRNA SMART-pools purchased from Dharmacon. For siRNA transfection, cells were seeded at about 50% confluence into 6-well plates and then transfected with a specific siRNA duplex using Lipofectamine RNAiMAX Reagent Agent (Life Technologies, #13778150) according to the manufacturer's instructions for 48 hr.

Clonogenic Assay

Cells were seeded at a concentration of 500-1000 cells per well into 6-well plates. After 24 hr, cells were continuously treated with indicated compounds for about 2 weeks. Colonies were fixed with fixation solution (methanol:acetic acid = 5:1) at room temperature for 20 min and then stained with a solution of 0.5% crystal violet in methanol for 2 h.

CellTiter-Glo Assay and Synergy Study

Cells were seeded into 96-well plates at a cell density of 1,000-2,000 cells per well overnight and treated with indicated compounds. The ATP-based, cell viability assay was carried out after incubation for 4 days (CellTiter-Glo, Promega). Survival curves were calculated by best-fit analysis of the log of the drug concentration to fold change of treated cells over vehicle-treated cells. All survival assays included technical triplicates.

For the prexasertib/PP2Ai synergy assay, cells were plated in the same fashion as above, and treated with 6 increasing doses of both drugs as a matrix. 72 h after drug treatment, CellTiter-Glo was applied and absorbances were recorded. Each replicate was compared and analyzed using ComBenefit synergy software for BLISS analysis.

Immunofluorescence

Cells plated on coverslips were washed once with PBS and fixed with 4% paraformaldehyde for 10 min on ice. 0.5% Triton X-100 was added to cells for 30 min for extraction. Coverslips were then blocked with 3% BSA for 30 min and incubated with primary antibody (diluted in 3% BSA) overnight. Following overnight incubation, coverslips were washed three times with PBST, then incubated with anti-rabbit secondary antibody (Alexa Fluor, Life Technologies) for 1 h. Coverslips were then mounted with DAPI medium and visualized under the fluorescence microscope.

Measuring Cell Cycle and Rate of DNA Replication

For Cell cycle-FACS analysis, cells were collected and fixed with 70% ethanol. Fixed cells were washed once with PBS and then resuspended with PI/RNase Staining Solution (Invitrogen). After incubation for 20 min at 37°C, the cells were analyzed using a flow cytometer (BD Biosciences). For analysis of DNA replication, cells were pulse-labeled with 10 μM EdU for 30 min, and then detected using the Click-iT EdU Alexa Fluor 488 kit (Invitrogen).

Cytogenetics

Cells were treated with/without 100nM prexasertib before cytogenetic analysis. Cells were exposed for 2 h to colcemid (0.1mg/ml) and harvested using a 0.075M KCl hypotonic solution and fixed with 3:1 methanol: acetic acid. Slides were stained with Wright's stain and when possible, at least 50 metaphase spreads were scored for aberrations. Metaphase spreads were observed using a Zeiss Axio Imager microscope and captured using CytoVision software from Applied Imaging.

Subcellular Fractionation

Subcellular fractionations were isolated using a subcellular fractionation kit (Cell signaling Technology) according to the manufacturer's instruction. The fractionation protein lysates were analyzed with western blotting using indicated antibodies.

Alkaline Single-Cell Electrophoresis (Comet) Assays

Alkaline comet assays were performed as previous described (Cai et al., 2020). Cells were trypsinized and suspended in cold 1 × PBS, then mixed with low melting agarose (Trevigen) at a ratio of 1:10 (v/v) and immediately plated onto Comet slide (Trevigen). Alkaline electrophoresis was run at 25 V for 30 min in the electrophoresis system. Cell comets were imaged with a fluorescence microscope (Zeiss).

DNA Fiber Assay

DNA fiber assays were done as described in Lim et al. (2018). In brief, cells were treated with prexasertib or DMSO for 24 h, then labeled with 25 μM CIdU for 30 min, washed three times with PBS, and labeled with 200 μM IdU for 30 min. Cells were collected and embedded in low melting point (LMP) agarose plugs, then treated with proteinase K at 42°C overnight. Agarose plugs were then washed and digested with agarase. Agarase-treated samples were then combed onto silanized coverslips using FiberComb (Genomic Vision). Coverslips were blocked with blocking solution (PBS containing 1% BSA) for 30 min, and incubated with Rat anti-BrdU antibody (Abcam) and mouse anti-BrdU antibody (clone B44, BD Biosciences) at 4°C overnight, followed by incubation with Alexa Fluor 594 chicken anti-mouse and Alexa Fluor 488 goat anti-rat antibodies (Invitrogen). DNA fibers were visualized by fluorescence microscopy (Zeiss). Pictures were taken of at least 100 fibers per condition. DNA fibers were measured with ImageJ and graphed.

Kinase Reactions

293T cells were transfected with pHAGE-FAM122A-FLAG-HA and cellular lysates were generated 2 days post transfection with lysis buffer (150 mM NaCl, 50 mM Tris-HCl, 1 mM EDTA, 0.5% NP-40, 10% glycerol, and protease and phosphatase inhibitor cocktail sets (Millipore)). Negative control lysates were generated by mock transfecting 293T cells. FLAG-HA-FAM122A was immunoprecipitated from cellular lysates using M2 Magnetic Beads (Sigma-Millipore) at room temperature for 2 h. Immunoprecipitated FLAG-HA-FAM122A protein was added to CHK1 kinase reactions with 100 ng purified CHK1 (Millipore) and 20 uM ATP-g-S (Abcam). CHK1 was omitted from denoted reactions as a negative control. Kinase Reactions were conducted as previously described (Kim et al., 2020). Briefly, reactions were incubated at 30°C for 30 min and thiophosphate species were alkylated with p-nitrobenzyl mesylate (Abcam) for 2 h at room temperature. 6x SDS buffer was added to the reactions, samples were boiled, and FAM122A phosphorylation was assessed by SDS-PAGE followed by immunoblot.

RNA Sequencing and Analysis

Total RNA was prepared using a QIAGEN kit according to the manufacturer's instructions. Libraries were prepared using Roche Kapa mRNA HyperPrep strand specific sample preparation kits from 200ng of purified total RNA according to the manufacturer's protocol on a Beckman Coulter Biomek i7. The finished dsDNA libraries were quantified by Qubit fluorometer, Agilent TapeStation 2200, and RT-qPCR using the Kapa Biosystems library quantification kit according to the manufacturer's protocols. Uniquely dual-indexed libraries were pooled in equimolar ratios and sequenced on an Illumina NovaSeq 6000 with paired-end 50bp reads by the Dana-Farber Cancer Institute Molecular Biology Core Facilities. Sequenced reads were aligned to the UCSC hg19 reference genome assembly and gene counts were quantified using STAR (v2.7.3a) (Dobin et al., 2013). Differential gene expression testing was performed by DESeq2 (v1.22.1) (Love et al., 2014), and limma (Ritchie et al., 2015). RNaseq analysis was performed using the VIPER snakemake pipeline (Cornwell et al., 2018). The volcano plot of the results was generated using the ggplot2 package in R.

Computational Analysis

The analyses of FAM122A mRNA expression in the Cancer Genome Atlas (TCGA) tumors and matched normal tissues were performed as follows. The RNASeq expression raw HTSeq counts (Anders et al., 2015) of TCGA were downloaded from NCI Genomic Data Commons Data Portal (https://portal.gdc.cancer.gov/) using a custom Python script. For each cancer type, lowly expressed genes were filtered out if they did not have counts per million (CPM) of greater than 1 for at least 5% of the samples. The CPM values were normalized by the TMM (weighted trimmed mean of M-values) method using the edgeR package (Robinson et al., 2010). Tumor and matched normal tissue pairs were selected by identifying the sample pairs that were derived from the same patient (same patient barcode), and the tumor sample was a primary solid tumor (sample type code = 1) and the normal sample was a solid tissue normal (sample type code = 11). The ggpubr R package was used to perform the Wilcoxon rank sum test to show the difference in FAM122A

mRNA expression levels (log_2-transformed normalized CPMs) between the tumor and the matched normal tissue groups for each cancer type. The test results and the CPM data were plotted using the ggplot2 package in R.

The survival analyses of the TCGA patients were performed using the clinical and RNASeq expression data of TCGA Pan-Cancer study for 32 cancer types downloaded from the cBioPortal for Cancer Genomics (cbioportal.org). For each cancer type, samples were grouped into two groups, those with low mRNA expression of FAM122A and those with high mRNA expression of FAM122A. The two groups were determined based on the expression z-scores relative to the diploid samples. Samples with z-scores less than zero belonged to the low group; the remaining samples belonged to the high group. Survival analysis was then performed in R for each cancer type to determine whether there was a difference in the overall survival between the two groups. Kaplan-Meier curves were created, and the log-rank test was used to test for a difference in overall survival using the survival package in R. The p values were calculated from the chi-square distribution. The survminer R package was used to estimate median survivals, and to plot the Kaplan-Meier curves. Additionally, Cox proportional hazards regression was performed to estimate the hazard ratio between the low- and the high-FAM122A groups for each cancer type.

The CRISPR dependency analyses with cancer cell lines were performed using the data from the Broad Institute Cancer Dependency Map (DepMap) (Meyers et al., 2017) (https://depmap.org/portal/). The focused dependency analysis with FAM122A and CHEK1 was performed using the Achilles CRISPR Avana 17Q2 dataset. In this analysis, cell lines that were one standard deviation below or above the mean in terms of CHEK1 or CHEK2 CERES score were grouped into cell lines that had high CHEK1 dependency and low CHEK1 dependency, respectively. The CRIPSR co-dependency analysis of FAM122A was performed using the Achilles CRISPR DepMap Public 19Q4 dataset. This analysis involved computing association between the dependency of FAM122A and that of every other gene. Only cell lineages with at least 5 cell lines were included in the analysis. The limma R package (Ritchie et al., 2015) was used to compute the associations using the CERES scores from all cell lines, with robust eBayes option and correcting for cell lineage. Volcano plot of the results and the scatterplots of the CERES scores of FAM122A and CHEK1 for the cell lineages were created using the ggplot2 package in R.

QUANTITATION AND STATISTICAL ANALYSIS

All quantitative data were analyzed and graphed using Prism 8 GraphPad software. All data are represented as mean ± SD calculated using the Prism software, unless indicated otherwise. Significance was tested using 2-tailed Student's t test and two-way ANOVA as indicated. Clinical data were analyzed using the Kaplan-Meier method and the log-rank (Mantel-Cox) regression test for significance.

ADDITIONAL RESOURCES

This study did not generate any additional resources.

3014–3028 *Nucleic Acids Research, 2020, Vol. 48, No. 6* *Published online 25 January 2020*
doi: 10.1093/nar/gkaa038

C a E a Fa a a/B CA a a a a

a a a a a C a a A a

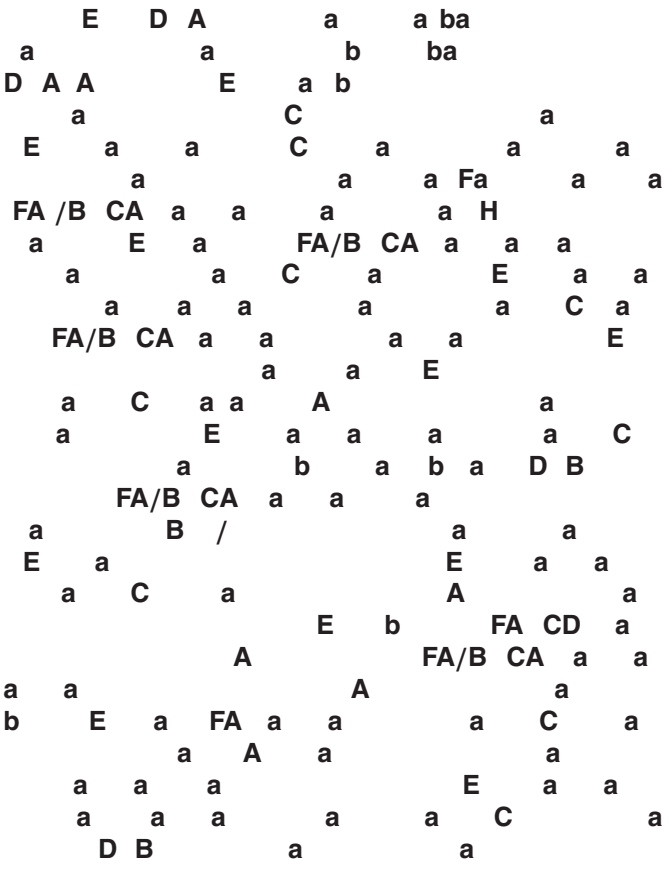
D D A a

[1]Department of Medical Genetics and Molecular Diagnostic Laboratory, Shanghai Children's Medical Center, Shanghai Jiaotong University School of Medicine, Shanghai, China, [2]Key Laboratory of Pediatric Hematology and Oncology Ministry of Health, Department of Hematology and Oncology, Shanghai Children's Medical Center, Shanghai Jiao Tong University School of Medicine, Shanghai, China, [3]Department of Radiation Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA, [4]Department of Microbiology and Molecular Genetics, University of Vermont, Burlington, VT, USA, [5]Center for DNA Damage and Repair, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA and [6]Susan F. Smith Center for Women's Cancers, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA

Received September 15, 2019; Revised December 12, 2019; Editorial Decision January 11, 2020; Accepted January 22, 2020

Downloaded from https://academic.oup.com/nar/article/48/6/3014/5715802 by NIH Library user on 29 December 2020

AB AC



D C

DNA interstrand cross-links (ICLs) are toxic lesions that prevent DNA replication and transcription by blocking DNA strand separation, and unrepaired ICLs lead to apoptosis and cell death (1). The Fanconi anemia (FA) pathway is essential for the repair of DNA-ICLs, and defects in the FA pathway result in Fanconi anemia, a chromosomal instability disorder characterized by congenital abnormalities, progressive bone marrow failure, and cancer predisposition (2). The FA proteins function in a multistep pathway required for the repair of endogenous and exogenous ICLs, such as ICLs induced by the therapeutic agent Mitomycin C (MMC). To date, 23 FA genes have been identified, which are grouped into three categories: the FA core complex (an E3 ligase complex), the FANCI/FANCD2 (ID) complex, and the downstream effector proteins, such as structure-specific nuclease and double-strand break (DSB) repair proteins (3,4). When DNA replication is blocked by an ICL, the FA core complex (containing FANCA, B, C, E, F, G, L and M) monoubiquitinates the FANCI–FANCD2 complex (ID2), a pivotal step in the FA pathway (5). The FA core complex is recruited to a stalled replication fork by an ICL via the anchoring complex containing FANCM subunit, along with Fanconi-associated proteins (FAAPs). Nucleolytic processing of the ICLs, which involves nucleases recruited by the SLX4/FANCP scaffold protein (6,7), generates DSBs that can be repaired by multiple downstream repair pathways (8).

Base excision repair (BER) is the major pathway for repairing DNA base damage and single strand breaks. If left unrepaired, these lesions can be mutagenic, blocking repli-

*To whom correspondence should be addressed. Tel: +1 617 632 2080; Fax: +1 617 632 6069; Email: alan_dandrea@dfci.harvard.edu
Correspondence may also be addressed to Jia Zhou. Tel: +1 617 582 8523; Email: jia_zhou@dfci.harvard.edu

© The Author(s) 2020. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

cation fork (9), or even perturbing epigenetic marks (10,11). The first and most critical step of BER is the searching and excision of damaged bases, a step that is carried out by DNA glycosylases. NEIL3 (DNA Endonuclease VIII-like 3) is a member of the Fpg/Nei glycosylase family (12,13), also including NEIL1 (14) and NEIL2 (15). Like other members of the Fpg/Nei family, NEIL3 contains a DNA glycosylase activity that excises damaged bases and an AP (apurinic/apyrimidinic site) lyase activity that cleaves the DNA backbone at an AP site, thus generating a single-strand break (SSB) (13,16). NEIL3 is distinguished from the other NEILs by its long C-terminal domain (CTD) (13). The glycosylase domain of NEIL3 prefers base lesions in single-stranded DNA (ssDNA) or ssDNA-containing structures (i.e. fork DNA) (13,16). NEIL3 also possesses the unique activity of removing damaged bases from G-quadruplex DNA (17–19).

The biochemical features of NEIL3 have been well characterized in the past decade, although the cellular function of NEIL3 has just begun to be understood. NEIL3 repairs telomere damage and protects telomeres during S phase to ensure accurate segregation of chromosome during mitosis (20). NEIL3 also plays a critical role in preventing autoimmunity, and its glycosylase activity is required for this function (21). NEIL3 appears to be important for cell proliferation, as evidenced by its role in regulating proliferation of cardiac fibroblasts in the heart and neural progenitor cells in the human brain (22,23). Its expression is increased in highly replicative tissues such as bone marrow (12,24) and in cancerous tissues (24). These studies demonstrate that NEIL3 is a versatile DNA glycosylase that functions outside of BER.

A new role for NEIL3 in ICL repair has recently been discovered. NEIL3 directly unhooks psoralen- and AP-ICLs during DNA replication in Xenopus egg extracts (25). This function of NEIL3 appears to be the first choice for the repair of a psoralen-ICL or AP-ICL (formed by AP site and an adenosine on opposite strands), and its failure activates the Fanconi anemia pathway. Purified NEIL3 and NEIL1 excises psoralen-induced DNA–DNA cross-links *in vitro* (26,27). Moreover, NEIL3 unhooks ICLs between glycosidic bonds and does not nick the DNA backbone, and accordingly does not generate DSBs (25,26). The ability of NEIL3 to accommodate those bulky lesions is supported by its crystal structure, which features a large lesion binding pocket of its glycosylase domain (28). In addition, mouse embryonic fibroblasts (MEFs) derived from *Neil3*$^{-/-}$ mice are sensitive to ICL-inducing agents such as cisplatin (29). These studies show that NEIL3 plays an important role in ICL repair.

Additional players in ICL repair have been recently discovered. The RING-Type E3 ubiquitin ligase TRAIP (TRAF Interacting Protein) for example is essential for cell proliferation and replication fork progression through DNA damage, and TRAIP mutations cause microcephalic primordial dwarfism (30,31). TRAIP is required for optimal phosphorylation of H2AX and RPA2 during S-phase after UV-C irradiation (31), and knockdown of TRAIP sensitizes cells to MMC (32). More recently, TRAIP was found to operate a switch between the NEIL3 and FA pathways in ICL repair by controlling the length of ubiquitin chain on CMG complex (33). In addition, the AAA^{+} ATPase family proteins RuvB Like 1 (RUVBL1) and its paralog RUVBL2 are DNA-dependent ATPases/helicases that are involved in diverse cellular activities. The RUVBL1/2 complex facilitates homologous recombination (34) and ICL repair (35,36). The RUVBL1/2 complex also regulates the abundance of FA core complex proteins, and depletion of RUVBL1/2 causes FA phenotypes in human cells, including defective in FA pathway activation, chromosomal instability and sensitivity to MMC (36).

Although NEIL3 was shown to act on ICLs biochemically in the Xenopus egg extracts, little is known about how NEIL3 contributes to ICL repair in human cells, and how the NEIL3 and FA pathways interact. Here we show that the NEIL3 pathway is the major pathway that repairs psoralen-induced ICLs (psoralen-ICLs), while the FA pathway is mostly responsible for repairing MMC- and cisplatin- induced ICLs. NEIL3 provides a strong response to psoralen-ICLs, resulting in a rapid, PARP-dependent recruitment to psoralen-ICL sites and protein accumulation/stabilization. Unlike the canonical FA pathway, NEIL3 unhooks psoralen-ICLs without creating DSBs in human cells. TRAIP appears to function upstream of the NEIL3 and the FA pathways, while RUVBL1/2 binds directly to NEIL3 and functions within the NEIL3 pathway. Collectively, we show the NEIL3 pathway is prioritized in psoralen-ICL repair through a unique mechanism in human cells.

A E A A D E H D

Cell lines and culture conditions

Human FA fibroblasts including PD20 (*FANCD2*$^{-/-}$), and PD326 (*FANCG*$^{-/-}$) were derived from FA patients as previously described (37). All human FA fibroblasts, U2OS and HeLa cells were grown in Dulbecco's modified Eagle's medium supplemented with 10% (v/v) fetal bovine serum (Thermo Fisher Scientific) and 1% penicillin/streptomycin (Thermo Fisher Scientific) in a 5% CO_2 incubator at 37°C. The primary wild-type and *Nelil3*$^{-/-}$ mouse embryonic fibroblasts (MEFs) were gifts from Magnar Bjørås. The MEFs were immortalized with SV40 antigen according to the protocol described previously (38). The immortalized MEFs were grown in minimum essential medium supplemented with 10% (v/v) fetal bovine serum (Thermo Fisher Scientific), 2 mM L-glutamine and 1% penicillin/streptomycin (Thermo Fisher Scientific) in a 5% CO_2 incubator at 37°C.

Plasmids, siRNA, and transfection

The EGFP-FANCD2 and NEIL3-HA plasmids were described previously (20,39). The RUVBL1-Flag (#OHu26261) and RUVBL2-Flag (#OHu25367) plasmids were purchased from GenScript. Plasmid was transfected using Lipofectamine LTX Reagent (Thermo Fisher Scientific). For siRNA treatment, cells were transfected with specific siRNA (Dharmacon and Qiagen; see Supplementary Table S1) using Lipofectamine RNAiMAX Reagent (Thermo Fisher Scientific).

Downloaded from https://academic.oup.com/nar/article/48/6/3014/5715802 by NIH Library user on 29 December 2020

Generation of *NEIL3*$^{-/-}$, *FANCA*$^{-/-}$ and *FANCD2*$^{-/-}$ cells

HeLa-*FANCA*$^{-/-}$ and U2OS-*FANCD2*$^{-/-}$ cell lines were generated using the CRISPR-Cas9 system.To generate HeLa-*NEIL3*$^{-/-}$ and U2OS-*NEIL3*$^{-/-}$ cell lines, HeLa- and U2OS-CAS9 stable cell lines were firstly generated by transfection with the Cas9 plasmid and selected by blasticidin (Thermo Fisher Scientific). Oligonucleotides encoding guide RNAs targeting NEIL3 (5′-ATTCGCGCGCGGGTGCTCCC-3′) was cloned into the lentiGuide-Puro vector, a gift from Feng Zhang (Addgene plasmid # 52963). To produce lentivirus, HEK293T cells were seeded at 60% confluence in 60-mm dishes 24 h before transfection. The gRNA-lentiGuide-Puro construct was transfected together with the packaging plasmids (psPAX2 and PMD2.G) using LipoLTX reagent (Thermo Fisher Scientific). Virus was harvested 48 h after transfection, filtered through a 0.45 mm low-protein-binding membrane (Millipore), and then placed on the HeLa- or U2OS-CAS9 cells with a density of 10%. Mono-clone was collected after 2–3 weeks of selection by puromycin, and knockout clones were identified by western blotting.

PUVA (psoralen plus ultraviolet A) treatment

To induce psoralen-ICLs, the cells were washed using 1× PBS and were treated with the indicated concentrations of psoralen (Sigma-Aldrich #P8399) in 1× Hank's Balanced Salt Solution (HBSS) containing 2% FBS. After incubation for 30 min in the dark, cells were exposed to ultraviolet A (UVA) light (365 nm) at a dose of 0.8 J/cm^2 (40,41). Following treatment, cells were allowed to recover at 37°C in complete medium.

CellTiter-Glo cell viability assay

The CellTiter-Glo luminescent cell viability assay kit (Promega) was used to measure cell viability after drug treatment. The CellTiter-Glo kit determines the number of viable cells in culture based on quantitation of the ATP present, an indicator of metabolically active cells Generally, to determine the drug sensitivity of *NEIL3*$^{-/-}$ cells, 2000 cells (HeLa, U2OS or MEF) per well were plated in a 96-well plate. The sources and doses of DNA damaging agents were as following: camptothecin (Selleckchem #S1288: 0, 0.625, 1.25, 2.5, 5, 10 nM), mitomycin C (MMC, Sigma Aldrich #M0503: 0, 1.25, 2.5, 5, 10, 20 ng/ml), cisplatin (Sigma Aldrich #P4394: 0, 0.0625, 0.125, 0.25, 0.5, 1 μM), Hydroxyurea (HU, Sigma Aldrich #H8627: 0, 0.03125, 0.0625, 0.125, 0.25, 0.5 mM), psoralen (Sigma-Aldrich #P8399, 0, 0.03125, 0.0625, 0.125, 0.25, 0.5 μM), X-ray irradiation (IR: 0, 1, 2, 5, 7.5, 10 Gy), PUVA (0, 0.015625, 0.03125, 0.0625, 0.125 or 0.25 μM psoralen incubation for 30 min followed by exposure to 0.8 J/cm^2 UVA), UVA (0, 0.0625, 0.125, 0.25, 0.5, 1.0 J/cm^2), Ultraviolet B (UVB: 0, 50, 100, 200, 400, 800 J/m^2), or Ultraviolet C (UVC: 0, 2, 5, 10, 20, 40 J/m^2). Cells were cultured in fresh medium for 4–5 days after treatment, by then cell viability was measured. To test knockdown of NEIL3, FANCA, FANCD2 and TRAIP (Supplementary Table S1), indicated cells were transfected with control or specific siRNA (Final concentration of 20 nM siRNA) in 6-well plate and the cells were seeded into 96-well plates next day prior to treatment with DNA damage agents, using the condition stated above. To test knockdown of RUVBL1 and RUVBL2 (Supplementary Table S1), cells were transfected with control or specific siRNA (Final concentration of 10 nM siRNA) in 96-well plate (5000 cells/well). 24 h after transfection, the cells were treated with PUVA (0, 0.03125, 0.0625, 0.125, 0.25 or 0.5 μM psoralen incubation for 30 min followed by exposure to 0.8 J/cm^2 UVA) or MMC (0, 1.25, 2.5, 5, 10, 20 ng/ml) and were subsequently incubated for a further 48 h to measure the cell viability. To test PARP inhibitor, cells were pretreated with 10 μM olaparib (Selleckchem #AZD8821) for 2 h and then treated with indicated concentrations of psoralen (0, 0.015625, 0.03125, 0.0625, 0.125 or 0.25 μM) in HBSS containing 2% FBS and 10 μM olaparib for 30 min. After exposed with 0.8 J/cm^2 UVA, cells were cultured in fresh medium with olaparib (0.5 μM) for 4–5 days before measuring cell viability.

Immunofluorescence

Cells were seeded on cover slide in 12-well plate and then were treated with PUVA (50 nM psoralen for 30 min and then irradiation with 0.8 J/cm^2 UVA) or MMC (5 ng/ml). The γH2AX foci was determined at 0, 1, 6 and 20 h, respectively, after releasing into fresh medium. Briefly, cells were washed by 1× PBS and fixed using 4% paraformaldehyde at room temperature for 20 min. After washing with 1× PBS three times and blocking (blocking buffer, 1× PBS/5% goat serum/0.3% Triton X-100) for 1h, cells were incubated with anti-γH2AX (Ser139) (Cell Signaling Technology #2577) at 4°C overnight. γH2AX foci were generated after incubation with goat anti-rabbit IgG (H+L) secondary antibody (Life Technologies #A32731) for 1 h and detected using fluorescence microscope. For siRNA, 48 h after transfection, cells were then treated with PUVA and followed with above protocols to detect the γH2AX foci formation.

Neutral comet assay

Cells were seeded in six-well plate and exposed to PUVA (10 μM psoralen for 1 h followed by 1.0 J/cm^2 UVA). After 4 h, the neutral comet assay was performed to detect DSBs, using a commercial kit (Trevigen, # 4250-050-K) according the manufacturer's protocol. In siRNA transfected groups, cells were treated with PUVA (10 μM psoralen for 1 h followed by 1.0 J/cm^2 UVA) 72 h after siRNA transfection and were subjected to the neutral comet assay 4 h after PUVA. Quantification of the comet length was performed using Image J v.1.8.0_112 software.

Psoralen/UVA laser striping

To test the recruitment of NEIL3, the U2OS cells were seeded in glass bottomed dishes and transfected indicated siRNAs. Forty eight hours after transfection, NEIL3-EGFP plasmid was transfection into the cells. Next day the cells were incubated with psoralen (10 μM) for 1 h. A Zeiss PALM microdissection microscope equipped with a 360 nm UV laser at the 34% energy dose was then used to observe the recruitment.

Downloaded from https://academic.oup.com/nar/article/48/6/3014/5715802 by NIH Library user on 29 December 2020

To test the recruitment of FANCD2, we generated the U2OS stable cell line using the EGFP-FANCD2 plasmid by selection with puromycin. Cells were seeded in glass bottomed dishes and treated with indicated siRNAs. Seventy two hours after transfection, cells were incubated with psoralen (10 μM) in culture medium for 1 h and then used to perform the laser irradiation.

To test the PARPi, NEIL3-EGFP transfection or EFFP-FANCD2 stable U2OS cells were pre-treated with DMSO or 10 μM olaparib for 3 h, and then were used to perform the laser irradiation described as above.

Cell synchronization and Cell cycle analysis

HeLa cells were double blocked with 2 mM thymidine (arrested for 16 h, released for 8 h, and arrested for another 16 h), and then the cells were harvested for cell cycle and NEIL3 expression analysis at indicated time point. Cells were collected by digestion using trypsin and were washed twice with 1× cold-PBS. Cells were then fixed in 70% ethanol at −20°C overnight and stained with propidium iodide for 20 min. Cell cycle distribution was analyzed using a FACS Canto machine. The data was analyzed by FlowJo software v10.07 (BD Biosciences) or Modfit LT software v5.0 (Verity Software House).

Mass spectrometry

HEK293T cells were seeded in 15 cm dishes and transfected with NEIL3-EGFP or empty vector. Twenty four hours after transfection, cells were treated with 10 μM psoralen plus UVA irradiation (0.8 J/cm^2) and released to culture for indicated time. Cells were then washed with pre-chilled PBS three times and lysed into the lysis buffer (20 mM HEPES-K^+, pH 7.6, 0.1 mM KCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 10% glycerol, 0.01% NP-40, 1 mM DTT, 0.2 mM PMSF, 0.5 mM benzamidine, 1 ug/ml each of leupeptin, aprotinin and pepstatin A) for 30 min at 4°C. Cell lysates were incubated with GFP-Trap resin (ChromoTek) at 4°C overnight and the resin was washed with cold lysis buffer four times. The immunoprecipitated NEIL3-eGFP and its binding partners were eluted with SDS elution buffer (2% SDS, 50 mM Tris–Cl, pH 6.8). The samples were prepared by trichloroacetic acid (TCA) precipitation as previously described (42), and sent to for mass spec analysis.

Immunoprecipitation (IP)

At 48 h after plasmid transfection in 10 cm dish, HEK293T cells were washed twice with ice-cold phosphate-buffered saline. Cells were scraped from dish using 1 ml cold PBS and then were spun at 500 g for 3 min at 4°C. After removal of the supernatant, the cell pellet were lysed in 600 μl IP lysis buffer (250 mM NaCl, 50 mM pH 7.5 Tris–HCl, 1 mM EDTA, 10% glycerol, 0.5% NP40) containing EDTA-free protease inhibitor (Roche) at 4°C for 30 min. Centrifuge cell lysate at 20 000 g for 10 min at 4°C and transfer supernatant to a pre-cooled tube. The FLAG® M2 magnetic beads (Sigma-Aldrich) or GFP-Trap resin (ChromoTek) were washed three times using the IP lysis buffer and 25 μl beads were added to each supernatant tube. After incubating for 4 h at 4°C, the beads were washed three times with the IP lysis buffer and then were boiled with 100 μl 2× SDS-sample buffer for 10 min at 95°C to dissociate immunocomplexes.

Immunoblotting and antibodies

To prepare whole-cell extracts, cells were washed once with ice-cold PBS and lysed with SDS lysis buffer (100 Mm pH 6.8 Tris–HCl, 20% glycerol, 1% SDS). Equal amounts of samples were subjected to 3–8% Tris-acetate gel (for FANCD2-ubiquitination detection) or 4–12% Bis–tris gel (Thermo Fisher Scientific). The protein was transferred to a PVDF membrane. Membranes were blocked in 2% BSA in TBS-T for 1 h and then incubated overnight at 4°C with the following antibodies: anti-FLAG M2 antibody (Sigma-Aldrich #F1804), anti-HA-tag (Cell Signaling Technology #3724), anti-GFP (Cell Signaling Technology #2555), anti-beta-actin (Cell Signaling Technology #4967), anti-GAPDH (Cell Signaling Technology #2118), anti-NEIL3 (Proteintech, 11621-1-AP), anti-FANCA (Bethyl Laboratories #A301-980), anti-FANCD2 (Novus Biologicals #NB100-182), anti-FANCI (Santa Cruz Biotechnology #sc-271316), anti-γH2A.X (Ser139) (Cell Signaling Technology #2577), anti-PARP1 (Cell Signaling Technology #9542), anti-PARP3 (Novus Biologicals #NBP1-31415), anti-MCM7 (Cell Signaling Technology #3735), anti-RUVBL1 (Abcam #ab51500) and anti-RUVBL2 (BD Biosciences #612482). Proteins were detected using a chemiluminescence system with a horseradish peroxidase conjugated secondary antibody.

Quantitative RT-PCR

Total RNA was extracted applying a RNeasy Mini kit (Qiagen), and cDNA was obtained using the SuperScript™ III First-Strand Synthesis kit (Thermo Fisher Scientific). The following primers were used for qPCR. NEIL3: forward, 5′-GTCCTAATTGTGGTCAGTGCC-3′; reverse, 5′-TCCA GTGCTCTTCCGACTTC-3′; TRAIP: forward, 5′-CTAA AAGAGGCACGGAAGGC-3′; reverse, 5′- TTTCCTGC AGCATCGTTAGC-3′; GAPDH: forward, 5′-CCATCT TCCAGGAGCGAGAT-3′; reverse, 5′-TGCTGATGAT CTTGAGGCTG-3′: NEIL1: forward, 5′-CTCGCCCTAT GTTTCGTGGA -3′; reverse, 5′-TTCAGCCGGTACAG GATCTC-3′.

E

NEIL3 plays a major role in psoralen-ICL repair independent of the FA/BRCA pathway.

The biochemical role of NEIL3 in psoralen-ICL repair has been demonstrated biochemically *in vitro* (26) and in the Xenopus egg extract system (25). Whether NEIL3 plays a role in psoralen-ICL repair or in other types of DNA damage (i.e. double-strand break) repair in human cells remains unclear. We thus generated *NEIL3*$^{-/-}$ human cells and asked whether these cells are sensitive to any DNA damaging agents. *NEIL3*$^{-/-}$ HeLa and U2OS cells were hypersensitive to PUVA (psoralen plus UVA irradiation) (Figure 1A, B

Downloaded from https://academic.oup.com/nar/article/48/6/3014/5715802 by NIH Library user on 29 December 2020

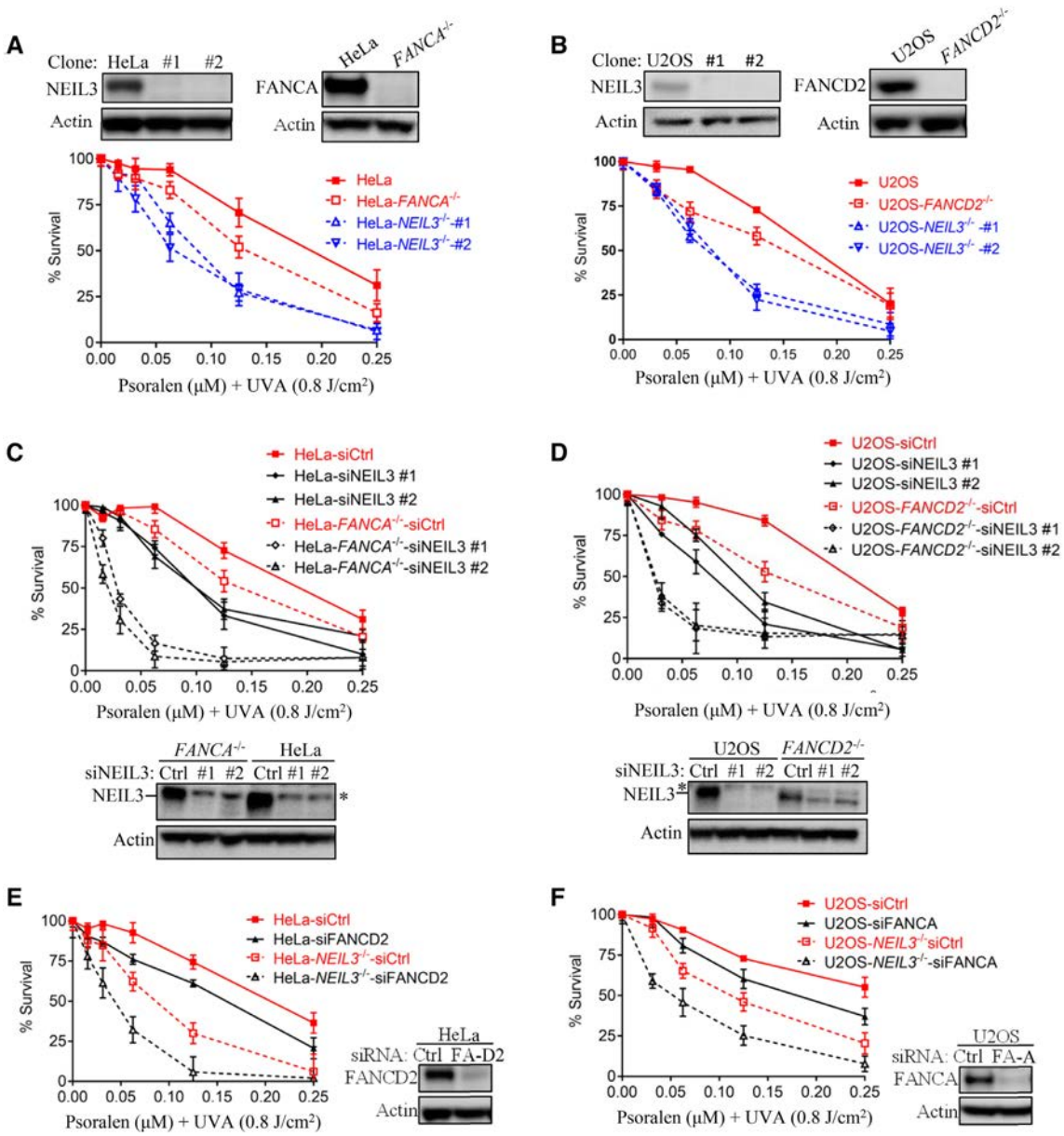


Figure 1. NEIL3 pathway plays a dominant role in psoralen-ICL repair in human cells. (**A, B**) Single cell HeLa-*NEIL3*$^{-/-}$, U2OS-NEIL3$^{-/-}$, HeLa-FANCA$^{-/-}$ and U2OS-FANCD2$^{-/-}$ clones were validated by Western blot. Cell viability was measured by CellTiter-Glo assays (Promega). PUVA, psoralen plus UVA. knockout cell lines were generated by CRISPR-Cas9. Guide RNA sequences and method used are shown in Materials and Methods. (**C, D**) Cell viability after PUVA treatment was measured after NEIL3 knockdown in HeLa WT versus HeLa-FANCA$^{-/-}$ and U2OS WT versus U2OS-FANCD2$^{-/-}$ cells. Western blots show that NEIL3 was knocked down. *Asterisk indicates a non-specific band above the NEIL3 band detected by the NEIL3 antibody. (**E, F**) PUVA sensitivity was determined after knockdown of FANCA or FANCD2 in U2OS-NEIL3$^{-/-}$ and HeLa-NEIL3$^{-/-}$ cells, respectively. Western blots show successful knockdown of FANCD2 and FANCA. Data are mean ± standard error of the mean (s.e.m.) from three independent experiments.

and Supplementary Figure S1A), while treatment with psoralen or UVA alone had no impact on cell growth (Supplementary Figure S1B). *Neil3*$^{-/-}$ MEFs were also sensitive to PUVA (but not UVA or psoralen) when compared to WT control (Supplementary Figure S1C). *NEIL3*$^{-/-}$ HeLa cells were also sensitive to ICL-inducing agents MMC and cisplatin, but to a lesser extent than PUVA. (Supplementary Figure S1A). *NEIL3*$^{-/-}$ HeLa also showed mild sensitivity to HU and no sensitivity to camptothecin, IR, UVB or UVC (Supplementary Figure S1A). Taken together, the sensitivity profile results suggest that NEIL3 plays a major role in psoralen-ICL repair.

The FA/BRCA pathway has been extensively studied in MMC- and cisplatin-ICL repair, and it was recently implicated in psoralen-ICL repair (25,33). To study the relative contribution of NEIL3 and FA/BRCA pathways to the repair of those three types of ICLs, the sensitivities of NEIL3-deficient and FA-deficient were compared. Although both *FANCA*$^{-/-}$ and *FANCD2*$^{-/-}$ cells were sensitive to PUVA, *NEIL3*$^{-/-}$ cells showed much more sensitivity to PUVA, sug-

Downloaded from https://academic.oup.com/nar/article/48/6/3014/5715802 by NIH Library user on 29 December 2020

gesting that NEIL3 is the major player in psoralen-ICL repair in HeLa and U2OS cells (Figure 1A and B). Conversely, NEIL3 pathway plays a less important role than the FA/BRCA pathway in MMC- and cisplatin-ICL repair, as demonstrated by the lesser sensitivity of HeLa-*NEIL3*$^{-/-}$ to those two agents (Supplementary Figure S1D).

To determine the relationship between the NEIL3 and FA pathways in psoralen-ICL repair, we knocked down NEIL3 in FA/BRCA-deficient cells and examined their PUVA sensitivity. siRNA knockdown of NEIL3 in FA/BRCA-deficient HeLa and U2OS cells led to additional sensitivity of those cells to PUVA (Figure 1C and D). Similar results were observed in FANCD2-deficient cells (PD20), which are SV40-transformed fibroblast cells derived from a FANCD2-deficient FA patient. We observed additional PUVA sensitivity when NEIL3 was knocked down in PD20 cells (Supplementary Figure S1E). Consistently, depletion of FANCA or FANCD2 by siRNA in *NEIL3*$^{-/-}$ cells also led to an additional sensitivity to PUVA (Figure 1E and F). These PUVA sensitivity data indicates that the NEIL3 pathway and the FA/BRCA pathway are functionally non-epistatic in psoralen-ICL repair. In contrast, HeLa-*NEIL3*$^{-/-}$ cells showed no additional sensitivity to MMC or cisplatin after FANCD2 knockdown (Supplementary Figure S1F), indicating that NEIL3 participates in MMC- and cisplatin-ICLs repair in a FA/BRCA pathway dependent manner. To investigate whether other NEIL DNA glycosylases is involved in psoralen-ICL repair, we knocked down NEIL1 in HeLa cells and tested their sensitivity to PUVA. NEIL1 knockdown resulted in resistance to PUVA in HeLa cells, where NEIL3 was concurrently up-regulated (Supplementary Figure S1G). These results were consistent with a previous study (43) and supported the idea that NEIL3 is the major enzyme to repair psoralen-ICL. Taken together, NEIL3 plays a major role in psoralen-ICL repair in human cells, which is independent of the FA/BRCA pathway. NEIL3 is also involved in MMC- and cisplatin-ICL repair in human cells, which is governed by the FA/BRCA pathway.

NEIL3 knockdown activates the FA/BRCA pathway in psoralen-ICL repair

In Xenopus egg extracts, psoralen-ICLs are repaired via the FA/BRCA pathway when the NEIL3 pathway is blocked (25). We asked whether this mechanism is also found in human cells. We first monitored FA/BRCA pathway activation after PUVA by detecting FANCD2 mono-ubiquitination. FANCD2 mono-ubiquitination was not detected in U2OS or HeLa parental cells treated with 50 nM psoralen plus 0.8 J/cm^2 UVA, when NEIL3 was present. In contrast, mono-ubiquitination of FANCD2 was clearly observed in *NEIL3*$^{-/-}$ HeLa and in *NEIL3*$^{-/-}$ U2OS cells under the same dose of PUVA treatment (Figure 2A). This is consistent with our observation that NEIL3 plays a dominant role in psoralen-ICL repair in human cells (Figure 1), and with the observation in Xenopus that the FA pathway is activated in the absence of NEIL3. These results suggest that the NEIL3 pathway is the major and prioritized pathway for repairing psoralen-ICLs in human cells. Only when the psoralen-ICLs exceed the repair capability of NEIL3 does the FA/BRCA pathway become activated as a backup repair mechanism.

Expression of NEIL3 and FA genes were interdependent at protein level

Since both NEIL3 and FA pathways contribute to the repair of PUVA-ICLs, we expected a compensatory relationship between the NEIL3 and FA pathway proteins. To our surprise, NEIL3 expression was down-regulated at the protein level in U2OS-*FANCD2*$^{-/-}$ and HeLa-*FANCA*$^{-/-}$ cells, compared to their WT counterparts (Figure 2B and Supplementary Figure S2A). Complementation with EGFP-FANCD2 in the U2OS-*FANCD2*$^{-/-}$ cells rescued the expression of NEIL3 (Figure 2B), and similar rescue results could be achieved in FANCD2 deficient (PD20) or FANCG deficient (PD326) FA patient fibroblasts (Supplementary Figure S2B). We excluded the possibility of robust regulation at NEIL3 mRNA level in FA-deficient cells, since NEIL3 mRNA level was not changed in those cells (Supplementary Figure S2C). We also excluded that the decreased level of NEIL3 was a result of cell cycle alteration. Consistent with our previous report (20), NEIL3 expression increases during S phase and reached maximal levels in G2/M (Supplementary Figure S2D). However, there was no significant change in the cell cycle distribution of NEIL3 in U2OS-*FANCD2*$^{-/-}$ and HeLa-*FANCA*$^{-/-}$ cells (Supplementary Figure S2E and F). Next, we studied how NEIL3 deficiency affected FA genes. FANCA and FANCD2 expression levels were decreased in *NEIL3*$^{-/-}$ HeLa and U2OS cells (Figure 2C). Further investigation showed the proteasome inhibitor MG-132 was able to stabilize NEIL3 in *FANCD2* cells and FANCD2 in *NEIL3*$^{-/-}$ cells, suggesting NEIL3 and FANC protein stabilize each other by reducing proteasome degradation (Supplementary Figure S2G). The mutually dependent expression levels led us to speculate that NEIL3 might form a complex with FANC proteins. However, we were not able to detect an interaction between NEIL3 and the FA proteins after many attempts of co-immunoprecipitation experiments, and the FA proteins were not detected in the NEIL3 immunoprecipitant in our IP-MS experiment (Supplementary Figure S2H).

Response of NEIL3 to psoralen-ICLs was enhanced in FA/BRCA-deficient cells

We next studied how NEIL3 responses to PUVA treatment. Upon PUVA treatment, rather than UVA irradiation or psoralen treatment, the NEIL3 expression was up-regulated in both HeLa and U2OS cells, suggesting that NEIL3 plays an active role in the DNA damage response (Figure 2D and E, Supplementary Figure S2A). Interestingly, NEIL3 protein was strongly induced after PUVA exposure in these FA-deficient cells (more than 4-fold), despite a low starting level of NEIL3 protein in these cells (Figure 2D and E). This increase is not likely an effect of cell cycle, since the S and G/M populations combined showed no difference in WT and *FANCA*$^{-/-}$ HeLa cells after PUVA treatment (Supplementary Figure S2I and J). These results demonstrated an active response of NEIL3 to PUVA, and also explained why FA/BRCA-deficient cells were only mildly sensitive to PUVA.

Downloaded from https://academic.oup.com/nar/article/48/6/3014/5715802 by NIH Library user on 29 December 2020

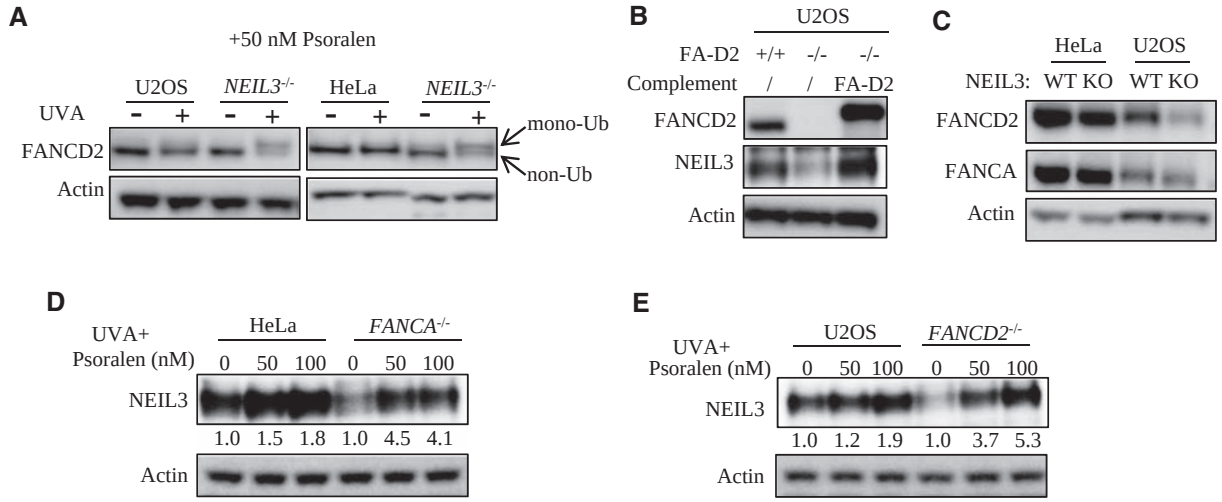


Figure 2. The relationship between the NEIL3 and the FA/BRCA pathways in psoralen-ICL repair. (**A**) WT and *NEIL3*$^{-/-}$ cells were treated with PUVA (pre-incubated with 50 nM psoralen for 30 min and then irradiation with 0.8 J/cm^2 UVA). 24 h after PUVA, the level of FANCD2 mono-ubiquitination was detected by WB to determine activation of the FA/BRCA pathway. (**B**) U2OS-*FANCD2*$^{-/-}$ cells were complemented with EGFP-FANCD2 and NEIL3 expression level was determined by western blot in those cells. (**C**) The FANCA and FANCD2 expression levels were determined in wild-type and NEIL3 deficiency cells. (**D**, **E**) The NEIL3 expression level was determined in wild-type and FANC gene deficiency cells, 24 h after being treated with indicated dose of psoralen plus UVA (0.8 J/cm^2) irradiation. Quantification of the Western blots was performed by ImageJ software and data were normalized to no psoralen.

The NEIL3 pathway repairs psoralen-ICLs without generating double-strand breaks

Previous studies in Xenopus egg extracts revealed that NEIL3 unhooks DNA psoralen-ICLs without generating double-strand breaks (DSBs) (25). We next determined whether a similar mechanism was present in human cells. We investigated DSB generation after PUVA treatment in HeLa cells that were defective in either the NEIL3 pathway or the FA/BRCA pathway, using γH2AX foci as markers. Following PUVA treatment, WT HeLa cells exhibited a slight increase of γH2AX foci compared to the untreated cells (0 h). *FANCA*$^{-/-}$ cells, which only retained the NEIL3 pathway, had a lower level of γH2AX foci compared to WT HeLa cells. Conversely, *NEIL3*$^{-/-}$ cells, which only have the FA/BRCA pathway for repair of the psoralen-ICLs, contained an increased number of γH2AX foci compared to WT HeLa cells (at 6 and 20 h) (Figure 3A and Supplementary Figure S3A). Consistent with the IF results, the γH2AX level was also increased in the *NEIL3*$^{-/-}$ cells after PUVA treatment as determined by western blot (Figure 3B, lanes 5–8). Moreover, depletion of FANCD2 decreased γH2AX foci formation in both WT and *NEIL3* HeLa cells (Figure 3C and Supplementary Figure S3B), indicating the FA pathway is responsible for the DSBs generated after PUVA. Since γH2AX is not a perfect marker for DSBs, we evaluated cellular DSB levels after PUVA treatment using the neutral comet assay, which detects DSBs directly at DNA level. Similar to the γH2AX foci and Western blot data, we observed that there were more DSBs in *NEIL3*$^{-/-}$ HeLa cells compared to WT counterparts after PUVA, as indicated in the increased tail length of the comets. Consistently, much shorter comet tails were observed in *FANCA*$^{-/-}$ cells (Figure 3D-E). In addition, depletion of FANCD2 decreased the tail length of comets in both HeLa WT and *NEIL3* cells (Figure 3F, G), indicating that DSB generation after PUVA is dependent on the FA pathway but not the NEIL3 pathway. To note, the elevated γH2AX foci in *NEIL3*$^{-/-}$ cells were completely rescued by siFANCD2. In addition, we did not observe a difference in γH2AX foci formation in *FANCA*$^{-/-}$ and *NEIL3*$^{-/-}$ cells following MMC treatment (Supplementary Figure S3C-D), suggesting again that NEIL3 shares the same downstream factors with FA/BRCA pathway in MMC-ICLs repair. Taken together, our data demonstrated that, unlike the FA/BRCA pathway, the unique psoralen-ICLs unhooking mechanism of NEIL3 is not associated with DSB formation in human cells.

PARP is required for the recruitment of NEIL3 and FANCD2-Ub to DNA damage

We next examined how NEIL3 is recruited to psoralen-ICL sites. Cells were pre-treated with psoralen and irradiated with a UVA laser (357 nm), and the NEIL3 recruitment was monitored by a GFP tag (Supplementary Figure S4A). In the absence of psoralen, UVA laser alone did not trigger the recruitment of NEIL3-EGFP (Figure S4B). The recruitment of NEIL3-EGFP to the PUVA laser stripe was very efficient; 90% of the cells showed positive NEIL3-EGFP recruitment to the laser track. The recruitment was also rapid, which occurs within one minute (Figure 4A, B). To determine whether NEIL3 recruitment to psoralen-ICL is dependent on any known DNA damage response pathways during psoralen-ICLs repair, we tested inhibitors of ATM, ATR, DNA-PK and PARP. We found only the PARP inhibitor (olaparib) strongly blocked NEIL3 recruitment to psoralen-ICLs (Figure 4A, B), while ATM, ATR or DNA-PKcs inhibitors did not affect NEIL3 recruitment (Supplementary Figure S4C). Note that olaparib almost completely blocked the recruitment of NEIL3, indicating this event is highly PARP dependent. In contrast, the recruit-

Downloaded from https://academic.oup.com/nar/article/48/6/3014/5715802 by NIH Library user on 29 December 2020

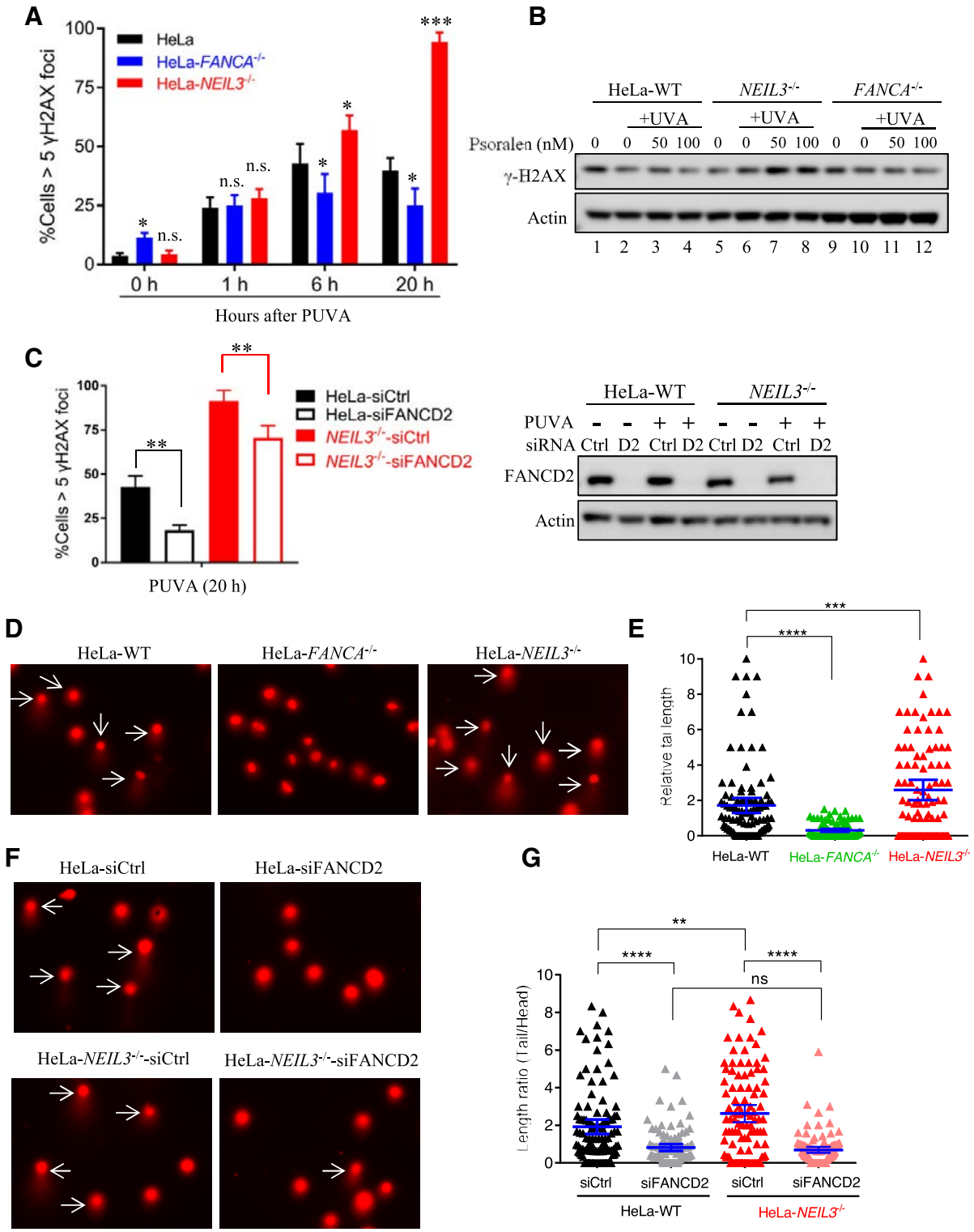


Figure 3. *NEIL3*$^{-/-}$ cells show increased DSBs compared to WT cells after PUVA treatment. (**A**) Quantification of the γH2AX foci at 0, 1, 6 and 20 h after PUVA. Cells with more than five γH2AX foci were counted and quantified, from IF image shown in Figure S3. Data are mean ± s.e.m. from three independent experiments. (**B**) γH2AX level was determined in WT, *FANCA*$^{-/-}$ and N*EIL3*$^{-/-}$ HeLa cells after treatment. (**C**) Quantification and statistical analysis of the γH2AX foci 20 h after PUVA in WT and *NEIL3*$^{-/-}$ HeLa cells after FANCD2 knockdown. Data are mean ± s.e.m. from three independent experiments. Western blot on the right shows the knockdown of FANCD2. PUVA, 50 nM psoralen + 0.8 J/cm^2 UVA. $^{**}P < 0.01$ (paired *t* test). (**D**) Representative images of comets, showing tail formation (arrows) 4 h after PUVA in WT, *FANCA*$^{-/-}$ and *NEIL3*$^{-/-}$ HeLa cells. (**E**) Quantification of the comet length at 4 h after PUVA shown in (D). At least 100 cells were analyzed for each group. (**F**) Representative images of comets showing tail formation in WT and *NEIL3*$^{-/-}$ HeLa cells after FANCD2 knockdown and PUVA. 72 h after siRNA transfection, cells were treated with PUVA and released for 4 h before comet assays. PUVA, 10 μM psoralen plus 1.0 J/cm^2 UVA; White arrows, cells with comet tails. (**G**) Quantification and statistical analysis of comet lengths from experiments shown in (F). ns, not significant, $^{**}P < 0.01$, $^{****}P < 0.0001$ (paired t test).

Downloaded from https://academic.oup.com/nar/article/48/6/3014/5715802 by NIH Library user on 29 December 2020

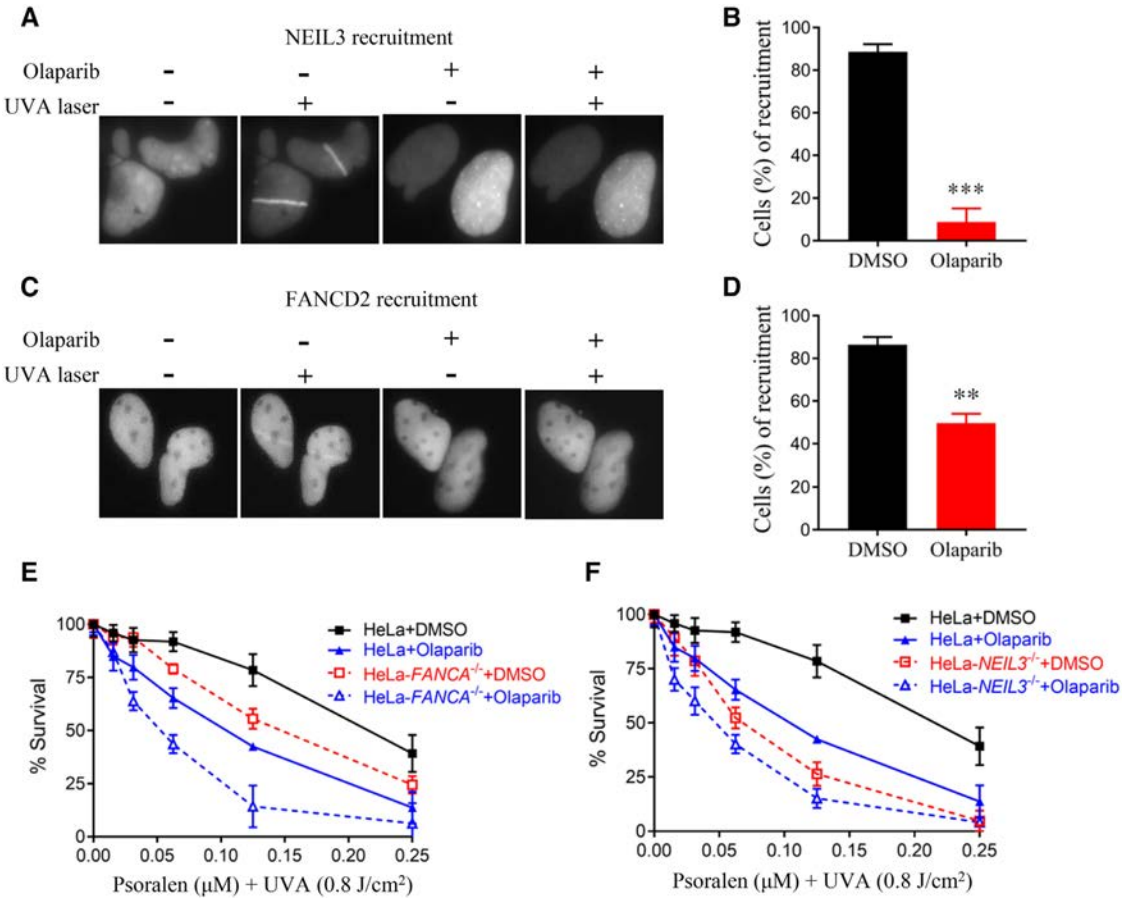


Figure 4. PARP is essential for NEIL3 recruitment to psoralen-ICLs and facilitates FANCD2 recruitment. (**A**) Recruitment of NEIL3 to PUVA laser tracts. U2OS cells expressing NEIL3-GFP were pre-incubated with 10 μM psoralen and the recruitment of NEIL3 was visualized 1 min after exposure to UVA laser. PARP inhibitor (olaparib) was added 1 hour before laser striping. (**B**) Quantification of the NEIL3 recruitment to psoralene-ICL laser tracts after treatment with DMSO or olaparib. A total of 100 cells were quantified for each group. (**C**) U2OS cells stably expressing EGFP-FANCD2 were pre-treated with 10 μM psoralen for 1 h before laser striping. Recruitment efficiency was determined 10 min after exposure to UVA laser. Olaparib or DMSO was added 1 h before laser striping. (**D**) Quantification of the recruitment efficiency of EGFP-FANCD2 after treatment with DMSO or olaparib, as shown in (C). A total of 100 cells were counted for each group. (**E**, **F**) PUVA sensitivity was determined in WT, *FANCA*$^{-/-}$ and *NEIL3*$^{-/-}$ HeLa cells after DMSO or olaparib treatment. CellTiter-Glo cell viability assays were performed. Mean and s.e.m. of three independent experiments are shown. $^{**}P < 0.01$; $^{***}P < 0.001$ (paired *t* test).

ment of EGFP-FANCD2 was much weaker and less efficient, and the recruitment could only be clearly observed five minutes after PUVA laser irradiation. PARPi olaparib only partially suppressed FANCD2 recruitment (Figure 4C, D and Supplementary Figure S4D). The PARP-dependent recruitment of NEIL3 and FANCD2 (to a lesser extent) was confirmed in PUVA laser irradiation experiments with PARP knockdown (Supplementary Figure S4E–G). Interestingly, the recruitment of NEIL3 and FANCD2 was mediated by different PARP family members. Knockdown of PARP1 blocked NEIL3 recruitment, while depletion of PARP3 blocked FANCD2 recruitment (Supplementary Figure S4E–G). In addition, depletion of NEIL3 had no impact on FANCD2 recruitment, and depletion of FANCD2 did not affect NEIL3 recruitment (Supplementary Figure S4H and I). These results are consistent with the idea that the NEIL3 pathway is promptly activated upon PUVA exposure, while the FA pathway is delayed.

So far, we showed that NEIL3 was critical for psoralen-ICL repair and PARP was indispensable for NEIL3 recruitment to psoralen ICL. We asked whether treating cells with a PARP inhibitor would have same effect as depleting NEIL3 in PUVA sensitivity assays. Consistently, olaparib treatment sensitized HeLa cells to PUVA and caused an additional sensitivity of *FANCA*$^{-/-}$ cells to PUVA (Figure 4E), in agreement with the PUVA laser striping data (Figure 4A, B) and PUVA sensitivity data with NEIL3 depletion (Figure 1C, D). On the other hand, olaparib caused a mild increase in PUVA sensitivity in NEIL3$^{-/-}$ cells (Figure 4F), which may be due to a less efficient inhibition of FANCD2 by olaparib. These results show that NEIL3 repairs PUVA-ICLs in a strictly PARP-dependent manner.

NEIL3 and RUVBL1/2 interact and function in the same pathway

To better understand the function of NEIL3 in psoralen-ICL repair, we performed mass spectrometry to identify novel NEIL3 binding partners upon PUVA treatment. NEIL3-GFP was expressed in HEK293T cells and im-

Downloaded from https://academic.oup.com/nar/article/48/6/3014/5715802 by NIH Library user on 29 December 2020

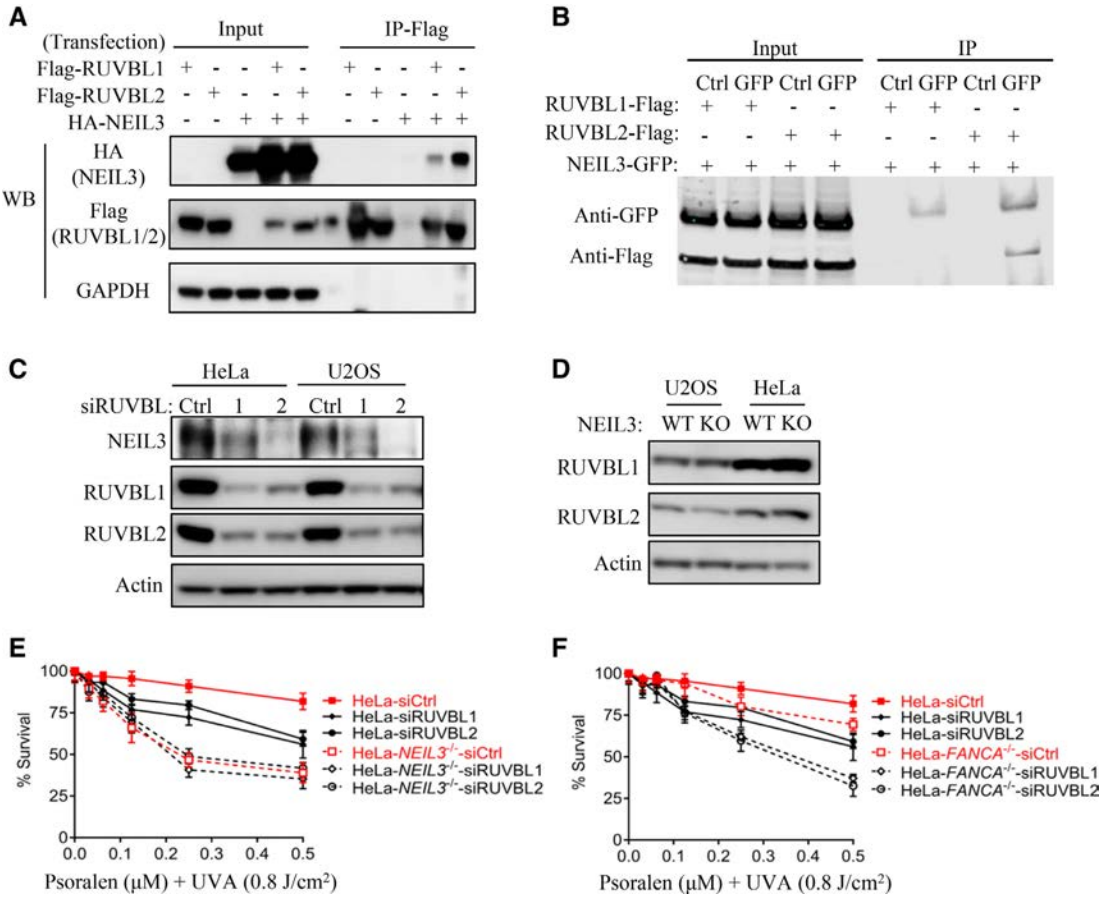


Figure 5. The RUVBL1/RUVBL2 complex is involved in psoralen-ICLs repair via the NEIL3 pathway. (**A**) 293T cells were transiently transfected with indicated plasmids, and co-IP were performed using anti-Flag antibody. WB of the inputs and IPs are shown to detect interactions. (**B**) 293T cells were transiently transfected with indicated plasmids, and co-IP were performed using normal IgG or anti-GFP antibody. WB of the inputs and IPs are shown to detect interactions. (**C**) NEIL3 protein levels were analyzed after knockdown of RUVBL1 or RUVBL2 in HeLa or U2OS cells. **(D)** RUVBL1 and RUVBL2 protein levels were analyzed by WB in WT and *NEIL3*$^{-/-}$ cells. (**E, F**) Cell viability was measured and compared after knockdown of RUVBL1 or RUVBL2 and PUVA treatment in WT, *NEIL3*$^{-/-}$ (E) and *FANCA*$^{-/-}$ (F) HeLa cells. CellTiter-Glo cell viability assays were performed.

munoprecipitated by the GFP tag. The quality of immunoprecipitation was accessed by silver staining, in which both the bait NEIL3-GFP and the binding partners were evident (Supplementary Figure S5A). A list of the top interactors is shown in Supplementary Figure S5B. Consistent with the PARP-dependent recruitment of NEIL3, PARP1 was detected by the mass spectrometry. To our interest, RUVBL1 and RUVBL2, known to form a heterodimeric complex (44), were among top hits. The RUVBL1/2 complex has been reported to associate with FA/BRCA pathway proteins and its knockdown sensitizes cells to mitomycin C (MMC) (36), a phenotype of FA/BRCA pathway deficient cells. We therefore investigated whether NEIL3 functions in concert with the RUVBL1/RUVBL2 complex. To confirm the interaction between NEIL3 and RUVBL2 and/or RUVBL1, we co-transfected the NEIL3-HA and the RUVBL1-Flag or RUVBL2-Flag plasmids Supplementary Figure S5C), and co-IP experiments were carried out. RUVBL2, and RUVBL1 to a lesser extent, pulled down NEIL3 (Figure 5A), while NEIL3 pulled down RUVBL2 but not RUVBL1 (Figure 5B). These data suggest NEIL3 primarily binds to RUVBL2 of the RUVBL1/2 complex. Depletion of RUVBL1 or RUVBL2 had no impact on the recruitment of NEIL3 or FANCD2 to PUVA laser stripes (Supplementary Figure S5D and E). However, a decreased NEIL3 protein level was observed when RUVBL1 or RUVBL2 was knocked down (Figure 5C). A similar reduction of FANCD2 and FANCI levels was observedSupplementary Figure S5F), which was consistent with previous work (36). Since NEIL3 and RUVBL1/2 form a complex in cells, the down-regulation of NEIL3 might be due to its instability resulting from knockdown of RUVBL1/RUVBL2. On the opposite, RUVBL1 and RUVBL2 levels were not changed significantly in *NEIL3*$^{-/-}$ and *FANCA*$^{-/-}$ cells (Figure 5D and Supplementary Figure S5G). In addition, the cell cycle profiles of RUVBL1-depleted HeLa cells were not altered (Supplementary Figure S5H and (36)). These results show that the RUVBL1/2 complex either stabilize NEIL3 in a complex or regulate NEIL3 through transcription.

We next studied the contribution of the RUVBL1/2 complex to PUVA-ICL repair and its relationship with NEIL3 in this process. Depletion of either RUVBL1 or RUVBL2 led to mild PUVA sensitivity in WT HeLa cells, while *NEIL3*$^{-/-}$ cells were very sensitive to PUVA. We did not observe further PUVA sensitivity after RUVBL1/2 knockdown in *NEIL3*$^{-/-}$ cells, indicating NEIL3 and RUVBL1/2

Downloaded from https://academic.oup.com/nar/article/48/6/3014/5715802 by NIH Library user on 29 December 2020

have an epistatic function in PUVA-ICL repair (Figure 5E). On the other hand, FANCA and RUVBL1/2 showed a non-epistatic relationship in PUVA repair (Figure 5F). Given that the role of RUVBL1/RUVBL2 complex in MMC-ICL repair was dependent on the FA/BRCA pathway (36), which was distinct from psoralen-ICL repair, we next examined MMC sensitivity in *FANCA*$^{-/-}$ and *NEIL3*$^{-/-}$ cells after depletion of RUVBL1/RUVBL2. We did not observe any additional MMC sensitivity after knocking down of RUVBL1 or RUVBL2 in either HeLa-*NEIL3*$^{-/-}$ or HeLa-*FANCA*$^{-/-}$ (Supplementary Figure S5I and J). These data suggest that the RUVBL1/RUVBL2 complex are epistatic with both NEIL3 and FA pathways in MMC-ICL repair. Taken together, these data suggest that RUVBL1/2 and NEIL3 function in the same pathway for psoralen-ICL repair, and this pathway is distinct from the FA/BRCA pathway, which is majorly responsible for MMC-ICL repair.

TRAIP is required for both NEIL3 and FA/BRCA pathways in psoralen-ICLs repair

Recently, TRAIP was identified as a master regulator of psoralen-ICL repair in both the NEIL3 and FA/BRCA pathways in Xenopus egg extracts (33). To investigate the role of TRAIP in psoralen-ICLs repair in human cells, we investigated PUVA sensitivity after TRAIP depletion. Lacking a good TRAIP antibody, the efficiency of siRNA-mediated depletion of TRAIP was evaluated by qPCR and Western blotting of exogenously expressed TRAIP-myc (Supplementary Figure S6A and B). Knockdown of TRAIP caused PUVA sensitivity, and more importantly, TRAIP knockdown presented additional PUVA sensitivity in both *FANCA*$^{-/-}$ and *NEIL3*$^{-/-}$ HeLa cells (Figure 6A and B), indicating that TRAIP is non-epistatic with both the FA/BRCA and the NEIL3 pathways in psoralen-ICL repair. In addition, depletion of TRAIP partially impeded NEIL3 recruitment to psoralen-ICLs, but it did not affect FANCD2 recruitment (Figure 6C and Supplementary Figure S6D and E). These results are consistent with a model in which ubiquitylation of the CMG complex by TRAIP is required for NEIL3 recruitment to chromatin in Xenopus egg extracts (33). Although MCM7 ubiquitylation has been identified as a key event that is required for NEIL3 recruitment, we only observed a weak impact of MCM7 knockdown in NEIL3 recruitment (Supplementary Figure S6C). Intriguingly, depletion of TRAIP increased the expression of NEIL3 and FANCD2 proteins, and it also increased the mono-ubiquitylation of FANCD2 (Figure 6D). We speculate that this elevated activation of NEIL3 and FA pathways might be a compensatory response to their nonfunctional recruitment in the absence of TRAIP. In conclusion, these data show TRAIP is non-epistatic with both the FA/BRCA and the NEIL3 pathways in psoralen-ICL repair, and suggest that TRAIP may function upstream of both pathways.

D C

In this study, we investigated the role of NEIL3 in the repair of psoralen-induced ICLs. We show that NEIL3 plays a major role in psoralen-ICL repair while the FA pathway is primarily responsible for the repair of MMC- and cisplatin-induced ICLs. The two pathways are non-epistatic in psoralen-ICL repair. Unlike the FA pathway, NEIL3 repair mechanism is PARP dependent and does not introduce DSBs. On the one hand, RUVBL1/2 forms a complex with NEIL3 and is functionally epistatic with NEIL3. On the other hand, TRAIP is functionally non-epistatic with both NEIL3 and FA/BRCA pathways, indicating that TRAIP may function upstream of the two pathways. A model is provided to summarize our findings (Figure 6E)

The hypersensitivity of NEIL3 depleted cells to psoralen-ICLs in multiple human cell lines demonstrates that NEIL3 is the major pathway to repair psoralen-ICLs, and that the NEIL3 pathway is prioritized over the FA pathway in human cells (Figure 1). Based on in vitro studies, NEIL3 is also important for repair of AP-ICLs in human cells (25,26). We also tested the sensitivity of *NEIL3*$^{-/-}$ cells to other types of ICLs – namely, cisplatin- and MMC-induced ICLs. Although *NEIL3*$^{-/-}$ cells were sensitive to these ICLs, they were more sensitive to psoralen-ICL, and the repair of cisplatin-ICLs and MMC-ICLs was primarily governed by FA pathway (Supplementary Figure S1). These results are consistent with previous studies indicating that NEIL3 does not process cisplatin-ICL (25,33). It will be interesting to determine how cells choose a pathway upon different ICL exposures. The rapid recruitment kinetics of NEIL3 and relatively slow kinetics of FANCD2 suggest that NEIL3 is the first line of defense. Cells will proceed with the mechanistically simpler NEIL3 pathway when the ICL substrate can be directly unhooked by NEIL3 (i.e. psoralen-ICL) and activate the more complex FA pathway when the ICL cannot be unhooked by NEIL3.

Wu *et al.* showed that both the FA and NEIL3 pathways must to be eliminated in order to generate cellular hypersensitivity to psoralen-ICLs in HAP1 cells (33). Synthetic lethality between FA and NEIL3 pathway was also evident in our study. However, we observed that NEIL3 depletion alone, either by siRNA or CRISPR-Cas9 knockout, led to hypersensitivity of HeLa and U2OS cells to psoralen-ICLs (PUVA), emphasizing that the NEIL3 pathway is the more efficient pathway in psoralen-ICL repair. This discrepancy may be a result of different cell lines being used. One question was whether the NEIL3 pathway is prioritized in human cells, similar to the observation in Xenopus extracts. Our data in human cell lines support this hypothesis. First, NEIL3 is recruited to PUVA-induced DNA damages faster than the FA protein FANCD2 (Figure 4). In our hands, strong NEIL3 recruitment could be observed within one minute, while FANCD2 recruitment was only evident after 5 minutes of PUVA treatment. Secondly, FANCD2 is not mono-ubiquitinated when NEIL3 is present in WT HeLa and U2OS cells, while, in response to PUVA, NEIL3 protein levels are elevated regardless of FA pathway availability (Figure 2). Thirdly, PUVA does not trigger significant phosphorylation of H2AX in WT cells, a sign that NEIL3-mediated repair occurs in the normal setting, while γH2AX foci levels are higher after PUVA treatment of *NEIL3*$^{-/-}$ cells, indicating that the FA pathway is activated as a backup pathway (Figure 3). In summary, our

Downloaded from https://academic.oup.com/nar/article/48/6/3014/5715802 by NIH Library user on 29 December 2020

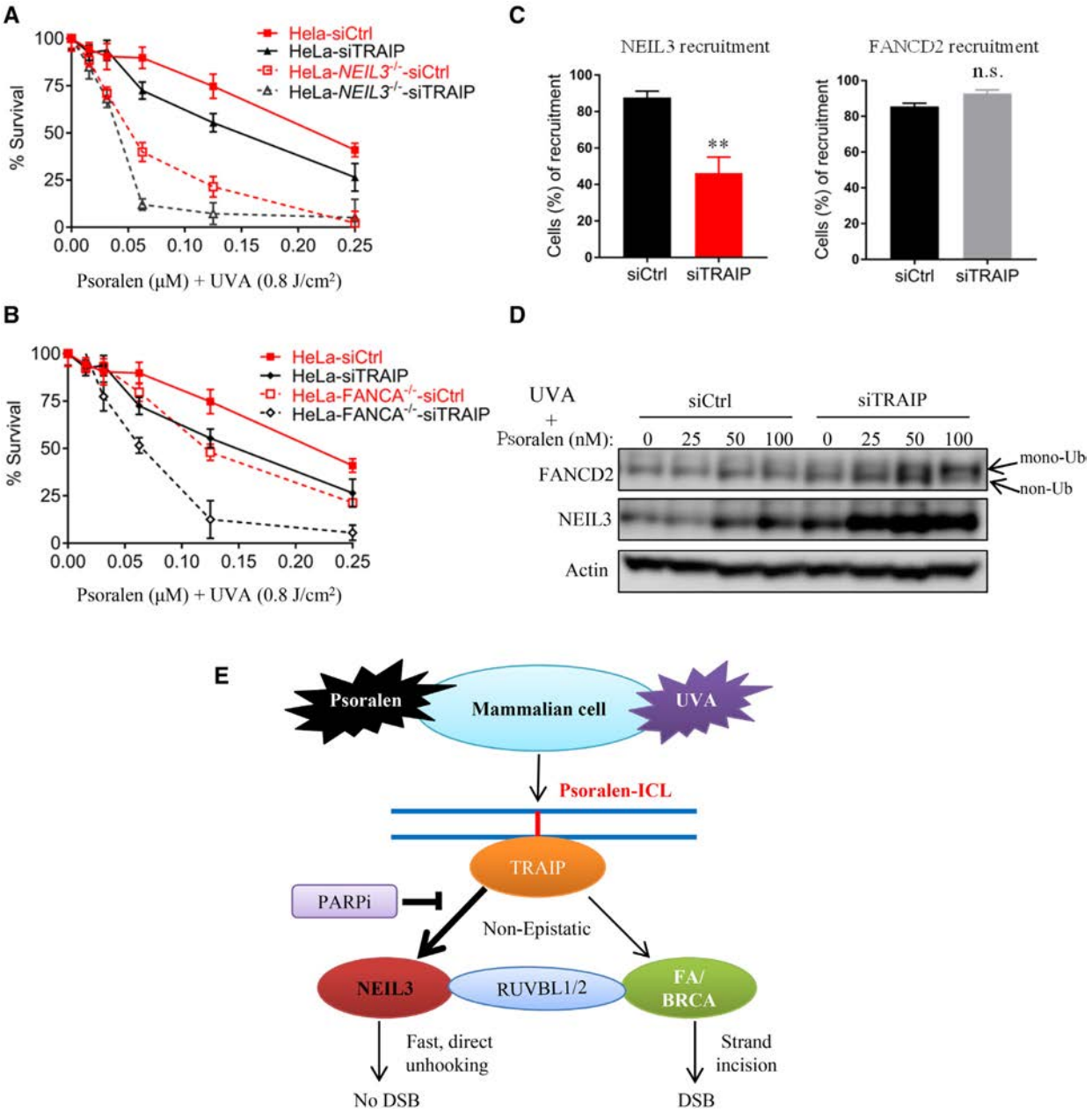


Figure 6. TRAIP show synthetic lethality with both the NEIL3 pathway and the FA pathway upon PUVA treatment. (**A, B**) Cell viability was measured after knockdown of TRAIP and PUVA treatment in WT, *FANCA*$^{-/-}$ and *NEIL3*$^{-/-}$ HeLa cells. (**C**) PUVA laser striping was used to analyze the recruitment efficiency of NEIL3-EGFP and EGFP-FANCD2 in U2OS cells after siRNA knockdown of TRAIP. A total of 100 cells were counted in each group. $^{**}P < 0.01$, n.s., no significance (paired t test). (**D**) FANCD2 and NEIL3 were analyzed by Western blotting in cell lysates from HeLa cells, which were transfected with control or TRAIP siRNA and treated with PUVA. (**E**) A model showing the NEIL3 pathway in psoralen-ICL repair in the context of the FA/BRCA pathway.

data show that NEIL3 pathway is prioritized in PUVA repair in human cells.

The Fanconi Anemia pathway plays an essential role in ICL repair during DNA replication (3). One of the key steps of this pathway is nucleolytic incision of DNA strands flanking the ICL (45), which generates a DSB that need further processing. Both purified NEIL3 and NEIL1 remove psoralen-induced DNA–DNA cross-links *in vitro*, the former does it without nicking DNA backbone (26), which is consistent with results in Xenopus extracts (25). Our data in human cells are consistent with those *in vitro* observations. When NEIL3 was deleted in HeLa cells, a strong accumulation of γH2AX was observed after PUVA treatment, and this γH2AX accumulation was attenuated by FANCD2 depletion (Figure 4). Our data demonstrate that the direct unhooking mechanism utilized by NEIL3, which occurs without the generation of DSBs, is conserved in human cells.

Replication fork stalling at a DNA–protein crosslink leads to TRAIP-dependent ubiquitylation and proteasome degradation (46). Wu *et al.* show that CMG ubiquityla-

Downloaded from https://academic.oup.com/nar/article/48/6/3014/5715802 by NIH Library user on 29 December 2020

tion by TRAIP is required to recruit NEIL3 for ICL unhooking. TRAIP associates with the replisome and non-specifically ubiquitylates lysine residues on bulky protein obstacles. The generation of a short ubiquitin chain can recruit NEIL3 (33). Here, we show that NEIL3 is rapidly recruited to PUVA laser tracts in a PARP-dependent manner. The PARP inhibitor olaparib, or knockdown of PARP1, strongly blocks NEIL3 recruitment (Figure 4). TRAIP, on the other hand, only weakly blocks NEIL3 recruitment (Figure 6). Since we looked at an unsynchronized cell population while the Xenopus extracts represent S phase, it is possible that the recruitment of NEIL3 relies more on TRAIP during S phase as opposed to other phases of the cell cycle. TRAIP itself is recruited to chromatin following ICL exposure during replication (47). BRCA-related FA proteins (BRCA2, FANCJ/BACH1 and FANCN/PALB2) requires replication for crosslink association, unlike other FA core and I/D2 complexes (48). In agreement with this study, we did not see reduction of FANCD2 recruitment after TRAIP knockdown. Our results suggest that psoralen-ICL repair by NEIL3 is not solely governed by TRAIP, and that it may also function outside of S phase.

Another interesting observation is that the protein levels of NEIL3 and FANC proteins are inter-dependent. We observed FANCD2 and FANCA proteins were decreased in *NEIL3*$^{-/-}$ cells and NEIL3 protein was decreased in the FA cells. Transcription regulation did not appear to be involved, since NEIL3 mRNA levels in *FANCD2*$^{-/-}$ or *FANCA*$^{-/-}$ cells were similar to the WT counterparts. Cell cycle alteration cannot explain the inter-dependency, since cell cycle distribution did not change in those FA cells. A previous study demonstrated that the NEIL1 level is dependent on the presence of an intact FA pathway and NEIL1 undergoes proteolytic degradation in FA cells (49). After treatment with the proteasome inhibitor MG-132, NEIL3 in U2OS-*FANCD2*$^{-/-}$ cells was increased to a similar level as in WT cells. Also, FANCD2 was increased after MG-132 treatment in *NEIL3*$^{-/-}$ HeLa cells, although the increase was not as prominent. There data suggest NEIL3 and FANC protein stabilize each other by reducing proteasome-mediated protein degradation.

The interdependency between NEIL3 and FANC proteins suggested that they may share a physical interaction. However, we failed to detect a specific protein-protein interaction between NEIL3 and FANC proteins using co-immunoprecipitation. The interaction may be transient or weak in nature and cannot be captured by co-immunoprecipitation. Alternatively, the stabilization between NEIL3 and FANC proteins may not be mediated by a direct interaction; rather, it may be mediated by a shared interaction with additional factors such as the RUVBL complex. We have shown that NEIL3 interacts with RUVBL2 and the interaction between FANC protein and RUVBL1/2 complex has been previously reported (36).

A potential role of NEIL1 in ICL repair has also been recently described. Macé-Aimé *et al.* demonstrated that the NEIL1 protein level is dependent on an intact FA pathway, suggesting downregulation of NEIL1 may contribute to hypersensitivity of FA cells to ICLs (49). However, McNeill *et al.* later showed that NEIL1 is recruited to psoralen-ICL, but it interferes and delays psoralen-ICL repair (43). Our results were consistent with the later study. NEIL1 knockdown cells were mildly resistant to psoralen-ICL. Interestingly, NEIL3 was concurrently upregulated in the NEIL1-depleted cells, providing a mechanism for the resistance of NEIL1-deficient cells to psoralen-ICL. Taken together, these results further suggest that NEIL3, but not the other NEIL glycosylases, is the major enzyme to repair psoralen-ICL.

In summary, our study shows that NEIL3 plays a major role in psoralen-ICL repair, which is non-epistatic with the FA pathway, a pathway which is primarily responsible for MMC- and cisplatin-induced ICLs. The psoralen-ICL repair mechanism of NEIL3 is PARP-dependent and does not introduce DSBs.

DA A A A AB

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository (50) with the dataset identifier PXD016256.

E E A DA A

Supplementary Data are available at NAR Online.

AC EDGE E

We thank Dr Magnar Bjørås at the Norwegian University of Science and Technology for providing the primary wild-type and *Neil3*$^{-/-}$ MEF cells. We thank Dr Ross Tomaino of the Taplin Biological Mass Spectrometry Facility in Harvard Medical School, for his help in mass spectrometry sample handling and data analysis. The authors also thank members of the D'Andrea lab for technical support and critical discussions.

Author contributions: N.L., J.Z. and A.D.D. designed the experiments. N.L. and J.Z. carried out the experiments and analyzed the data. N.L., J.Z., S.S.W. and A.D.D. wrote and edited the manuscript. A.D.D., J.Z., J.W. and J.C. supervised this work. J.Z. and A.D.D. conceived the original idea of this study. S.S.W. provided critical reagents. All authors were involved in critical discussions of this project.

F D G

U.S. National Institutes of Health [R37HL052725, P01HL048546]; U.S. Department of Defense [BC151331P1]; Breast Cancer Research Foundation; Fanconi Anemia Research Fund (to A.D.D.); National Natural Science Foundation of China [81601869 to N.L.]. Funding for open access charge: U.S. National Institutes of Health [R37HL052725, P01HL048546].

Conflict of interest statement. None declared.

EFE E CE

1. Deans,A.J. and West,S.C. (2011) DNA interstrand crosslink repair and cancer. *Nat. Rev. Cancer*, **11**, 467–480.
2. Rodriguez,A. and D'Andrea,A. (2017) Fanconi anemia pathway. *Curr. Biol.*, **27**, R986–R988.

Downloaded from https://academic.oup.com/nar/article/48/6/3014/5715802 by NIH Library user on 29 December 2020

3. Ceccaldi,R., Sarangi,P. and D'Andrea,A.D. (2016) The Fanconi anaemia pathway: new players and new functions. *Nat. Rev. Mol. Cell Biol.*, **17**, 337–349.
4. Niraj,J., Farkkila,A. and D'Andrea,A.D. (2019) The fanconi anemia pathway in cancer. *Annu Rev Cancer Biol*, **3**, 457–478.
5. Garner,E. and Smogorzewska,A. (2011) Ubiquitylation and the Fanconi anemia pathway. *FEBS Lett.*, **585**, 2853–2860.
6. Crossan,G.P., van der Weyden,L., Rosado,I.V., Langevin,F., Gaillard,P.H., McIntyre,R.E., Sanger Mouse Genetics,P., Gallagher,F., Kettunen,M.I., Lewis,D.Y. *et al.* (2011) Disruption of mouse Slx4, a regulator of structure-specific nucleases, phenocopies Fanconi anemia. *Nat. Genet.*, **43**, 147–152.
7. Kim,Y., Lach,F.P., Desetty,R., Hanenberg,H., Auerbach,A.D. and Smogorzewska,A. (2011) Mutations of the SLX4 gene in Fanconi anemia. *Nat. Genet.*, **43**, 142–146.
8. Moldovan,G.L. and D'Andrea,A.D. (2009) How the fanconi anemia pathway guards the genome. *Annu. Rev. Genet.*, **43**, 223–249.
9. Barnes,D.E. and Lindahl,T. (2004) Repair and genetic consequences of endogenous DNA base damage in mammalian cells. *Annu. Rev. Genet.*, **38**, 445–476.
10. He,Y.F., Li,B.Z., Li,Z., Liu,P., Wang,Y., Tang,Q., Ding,J., Jia,Y., Chen,Z., Li,L. *et al.* (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science*, **333**, 1303–1307.
11. Fleming,A.M., Ding,Y. and Burrows,C.J. (2017) Oxidative DNA damage is epigenetic by regulating gene transcription via base excision repair. *PNAS*, **114**, 2604–2609.
12. Morland,I., Rolseth,V., Luna,L., Rognes,T., Bjoras,M. and Seeberg,E. (2002) Human DNA glycosylases of the bacterial Fpg/MutM superfamily: an alternative pathway for the repair of 8-oxoguanine and other oxidation products in DNA. *Nucleic Acids Res.*, **30**, 4926–4936.
13. Liu,M., Bandaru,V., Bond,J.P., Jaruga,P., Zhao,X., Christov,P.P., Burrows,C.J., Rizzo,C.J., Dizdaroglu,M. and Wallace,S.S. (2010) The mouse ortholog of NEIL3 is a functional DNA glycosylase in vitro and in vivo. *PNAS*, **107**, 4925–4930.
14. Hazra,T.K., Izumi,T., Boldogh,I., Imhoff,B., Kow,Y.W., Jaruga,P., Dizdaroglu,M. and Mitra,S. (2002) Identification and characterization of a human DNA glycosylase for repair of modified bases in oxidatively damaged DNA. *PNAS*, **99**, 3523–3528.
15. Hazra,T.K., Kow,Y.W., Hatahet,Z., Imhoff,B., Boldogh,I., Mokkapati,S.K., Mitra,S. and Izumi,T. (2002) Identification and characterization of a novel human DNA glycosylase for repair of cytosine-derived lesions. *J. Biol. Chem.*, **277**, 30417–30420.
16. Krokeide,S.Z., Laerdahl,J.K., Salah,M., Luna,L., Cederkvist,F.H., Fleming,A.M., Burrows,C.J., Dalhus,B. and Bjoras,M. (2013) Human NEIL3 is mainly a monofunctional DNA glycosylase removing spiroimindiohydantoin and guanidinohydantoin. *DNA Repair (Amst.)*, **12**, 1159–1164.
17. Zhou,J., Liu,M., Fleming,A.M., Burrows,C.J. and Wallace,S.S. (2013) Neil3 and NEIL1 DNA glycosylases remove oxidative damages from quadruplex DNA and exhibit preferences for lesions in the telomeric sequence context. *J. Biol. Chem.*, **288**, 27263–27272.
18. Zhou,J., Fleming,A.M., Averill,A.M., Burrows,C.J. and Wallace,S.S. (2015) The NEIL glycosylases remove oxidized guanine lesions from telomeric and promoter quadruplex DNA structures. *Nucleic Acids Res.*, **43**, 4039–4054.
19. Fleming,A.M., Zhou,J., Wallace,S.S. and Burrows,C.J. (2015) A role for the fifth G-Track in G-Quadruplex forming oncogene promoter sequences during oxidative stress: Do these 'Spare Tires' have an evolved function? *ACS Cent. Sci*, **1**, 226–233.
20. Zhou,J., Chan,J., Lambele,M., Yusufzai,T., Stumpff,J., Opresko,P.L., Thali,M. and Wallace,S.S. (2017) NEIL3 Repairs telomere damage during S Phase to secure chromosome segregation at mitosis. *Cell Rep.*, **20**, 2044–2056.
21. Massaad,M.J., Zhou,J., Tsuchimoto,D., Chou,J., Jabara,H., Janssen,E., Glauzy,S., Olson,B.G., Morbach,H., Ohsumi,T.K. *et al.* (2016) Deficiency of base excision repair enzyme NEIL3 drives increased predisposition to autoimmunity. *J. Clin. Invest.*, **126**, 4219–4236.
22. Jalland,C.M., Scheffler,K., Benestad,S.L., Moldal,T., Ersdal,C., Gunnes,G., Suganthan,R., Bjoras,M. and Tranulis,M.A. (2016) Neil3 induced neurogenesis protects against prion disease during the clinical phase. *Sci. Rep.*, **6**, 37844.
23. Regnell,C.E., Hildrestrand,G.A., Sejersted,Y., Medin,T., Moldestad,O., Rolseth,V., Krokeide,S.Z., Suganthan,R., Luna,L., Bjoras,M. *et al.* (2012) Hippocampal adult neurogenesis is maintained by Neil3-dependent repair of oxidative DNA lesions in neural progenitor cells. *Cell Rep.*, **2**, 503–510.
24. Hildrestrand,G.A., Neurauter,C.G., Diep,D.B., Castellanos,C.G., Krauss,S., Bjoras,M. and Luna,L. (2009) Expression patterns of Neil3 during embryonic brain development and neoplasia. *BMC Neurosci.*, **10**, 45.
25. Semlow,D.R., Zhang,J., Budzowska,M., Drohat,A.C. and Walter,J.C. (2016) Replication-dependent unhooking of DNA interstrand cross-links by the NEIL3 glycosylase. *Cell*, **167**, 498–511.
26. Martin,P.R., Couvé,S., Zutterling,C., Albelazi,M.S., Groisman,R., Matkarimov,B.T., Parsons,J.L., Elder,R.H. and Saparbaev,M.K. (2017) The human DNA glycosylases NEIL1 and NEIL3 excise Psoralen-induced DNA-DNA cross-links in a four-stranded DNA structure. *Sci. Rep.*, **7**, 17438.
27. Yang,Z., Nejad,M.I., Varela,J.G., Price,N.E., Wang,Y. and Gates,K.S. (2017) A role for the base excision repair enzyme NEIL3 in replication-dependent repair of interstrand DNA cross-links derived from psoralen and abasic sites. *DNA Repair (Amst.)*, **52**, 1–11.
28. Liu,M., Imamura,K., Averill,A.M., Wallace,S.S. and Doublie,S. (2013) Structural characterization of a mouse ortholog of human NEIL3 with a marked preference for single-stranded DNA. *Structure*, **21**, 247–256.
29. Rolseth,V., Krokeide,S.Z., Kunke,D., Neurauter,C.G., Suganthan,R., Sejersted,Y., Hildrestrand,G.A., Bjoras,M. and Luna,L. (2013) Loss of Neil3, the major DNA glycosylase activity for removal of hydantoins in single stranded DNA, reduces cellular proliferation and sensitizes cells to genotoxic stress. *Biochim. Biophys. Acta*, **1833**, 1157–1164.
30. Chapard,C., Hohl,D. and Huber,M. (2012) The role of the TRAF-interacting protein in proliferation and differentiation. *Exp. Dermatol.*, **21**, 321–326.
31. Harley,M.E., Murina,O., Leitch,A., Higgs,M.R., Bicknell,L.S., Yigit,G., Blackford,A.N., Zlatanou,A., Mackenzie,K.J., Reddy,K. *et al.* (2016) TRAIP promotes DNA damage response during genome replication and is mutated in primordial dwarfism. *Nat. Genet.*, **48**, 36–43.
32. Hoffmann,S., Smedegaard,S., Nakamura,K., Mortuza,G.B., Raschle,M., Ibanez de Opakua,A., Oka,Y., Feng,Y., Blanco,F.J., Mann,M. *et al.* (2016) TRAIP is a PCNA-binding ubiquitin ligase that protects genome stability after replication stress. *J. Cell Biol.*, **212**, 63–75.
33. Wu,R.A., Semlow,D.R., Kamimae-Lanning,A.N., Kochenova,O.V., Chistol,G., Hodskinson,M.R., Amunugama,R., Sparks,J.L., Wang,M., Deng,L. *et al.* (2019) TRAIP is a master regulator of DNA interstrand crosslink repair. *Nature*, **567**, 267–272.
34. Clarke,T.L., Sanchez-Bailon,M.P., Chiang,K., Reynolds,J.J., Herrero-Ruiz,J., Bandeiras,T.M., Matias,P.M., Maslen,S.L., Skehel,J.M., Stewart,G.S. *et al.* (2017) PRMT5-dependent methylation of the TIP60 coactivator RUVBL1 is a key regulator of homologous recombination. *Mol. Cell*, **65**, 900–916.e907.
35. Runge,J.S., Raab,J.R. and Magnuson,T. (2016) Epigenetic regulation by ATP-dependent chromatin-remodeling enzymes: SNF-ing out crosstalk. *Curr. Top. Dev. Biol.*, **117**, 1–13.
36. Rajendra,E., Garaycoechea,J.I., Patel,K.J. and Passmore,L.A. (2014) Abundance of the Fanconi anaemia core complex is regulated by the RuvBL1 and RuvBL2 AAA+ ATPases. *Nucleic Acids Res.*, **42**, 13736–13748.
37. Li,N., Ding,L., Li,B., Wang,J., D'Andrea,A.D. and Chen,J. (2018) Functional analysis of Fanconi anemia mutations in China. *Exp. Hematol.*, **66**, 32–41.e38.
38. Xu,J. (2005) Preparation, culture, and immortalization of mouse embryonic fibroblasts. *Curr Protoc Mol Biol*, doi:10.1002/0471142727.mb2801s70.
39. Kim,H., Yang,K., Dejsuphong,D. and D'Andrea,A.D. (2012) Regulation of Rev1 by the Fanconi anemia core complex. *Nat. Struct. Mol. Biol.*, **19**, 164–170.
40. Muniandy,P.A., Thapa,D., Thazhathveetil,A.K., Liu,S.T. and Seidman,M.M. (2009) Repair of laser-localized DNA interstrand cross-links in G1 phase mammalian cells. *J. Biol. Chem.*, **284**, 27908–27917.

Downloaded from https://academic.oup.com/nar/article/48/6/3014/5715802 by NIH Library user on 29 December 2020

41. Chowdhari,S. and Saini,N. (2014) hsa-miR-4516 mediated downregulation of STAT3/CDK6/UBE2N plays a role in PUVA induced apoptosis in keratinocytes. *J. Cell Physiol.*, **229**, 1630–1638.
42. Koontz,L. (2014) TCA precipitation. *Methods Enzymol.*, **541**, 3–10.
43. McNeill,D.R., Paramasivam,M., Baldwin,J., Huang,J., Vyjayanti,V.N., Seidman,M.M. and Wilson,D.M. 3rd (2013) NEIL1 responds and binds to psoralen-induced DNA interstrand crosslinks. *J. Biol. Chem.*, **288**, 12426–12436.
44. Gorynia,S., Bandeiras,T.M., Pinho,F.G., McVey,C.E., Vonrhein,C., Round,A., Svergun,D.I., Donner,P., Matias,P.M. and Carrondo,M.A. (2011) Structural and functional insights into a dodecameric molecular machine - the RuvBL1/RuvBL2 complex. *J. Struct. Biol.*, **176**, 279–291.
45. Kim,H. and D'Andrea,A.D. (2012) Regulation of DNA cross-link repair by the Fanconi anemia/BRCA pathway. *Genes Dev.*, **26**, 1393–1408.
46. Larsen,N.B., Gao,A.O., Sparks,J.L., Gallina,I., Wu,R.A., Mann,M., Raschle,M., Walter,J.C. and Duxin,J.P. (2019) Replication-coupled DNA-protein crosslink repair by SPRTN and the proteasome in xenopus egg extracts. *Mol. Cell*, **73**, 574–588.
47. Raschle,M., Smeenk,G., Hansen,R.K., Temu,T., Oka,Y., Hein,M.Y., Nagaraj,N., Long,D.T., Walter,J.C., Hofmann,K. *et al.* (2015) DNA repair. Proteomics reveals dynamic assembly of repair complexes during bypass of DNA cross-links. *Science*, **348**, 1253671.
48. Shen,X., Do,H., Li,Y., Chung,W.H., Tomasz,M., de Winter,J.P., Xia,B., Elledge,S.J., Wang,W. and Li,L. (2009) Recruitment of Fanconi anemia and breast cancer proteins to DNA damage sites is differentially governed by replication. *Mol. Cell*, **35**, 716–723.
49. Mace-Aime,G., Couve,S., Khassenov,B., Rosselli,F. and Saparbaev,M.K. (2010) The Fanconi anemia pathway promotes DNA glycosylase-dependent excision of interstrand DNA crosslinks. *Environ. Mol. Mutagen.*, **51**, 508–519.
50. Vizcaino,J.A., Csordas,A., del-Toro,N., Dianes,J.A., Griss,J., Lavidas,I., Mayer,G., Perez-Riverol,Y., Reisinger,F., Ternent,T. *et al.* (2016) 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.*, **44**, D447–D456.

Downloaded from https://academic.oup.com/nar/article/48/6/3014/5715802 by NIH Library user on 29 December 2020

Published OnlineFirst July 10, 2020; DOI: 10.1158/0008-5472.CAN-20-1672

Exploiting the Microhomology-Mediated End-Joining Pathway in Cancer Therapy

Jeffrey Patterson-Fortin[1] and Alan D. D'Andrea[2,3]

ABSTRACT

Repair of DNA double-strand breaks (DSB) is performed by two major pathways, homology-dependent repair and classical nonhomologous end-joining. Recent studies have identified a third pathway, microhomology-mediated end-joining (MMEJ). MMEJ has similarities to homology-dependent repair, in that repair is initiated with end resection, leading to single-stranded 3′ ends, which require microhomology upstream and downstream of the DSB. Importantly, the MMEJ pathway is commonly upregulated in cancers, especially in homologous recombination-deficient cancers, which display a distinctive mutational signature. Here, we review the molecular process of MMEJ as well as new targets and approaches exploiting the MMEJ pathway in cancer therapy.

Introduction

DNA damage and double-stranded breaks

Maintenance of genome integrity is of upmost importance for cellular survival (1). Loss of genomic integrity results in permanent changes to DNA sequence and is the source of many human diseases, notably cancer (2–7). Genome integrity is achieved by DNA repair pathways, collectively known as the DNA damage response (DDR; ref. 8). Double-stranded breaks (DSB) are the most deleterious form of DNA damage (1). DSBs are generated either exogenously, for example, from exposure to ionizing radiation, or endogenously, for example, from reactive oxygen species secondary to aerobic metabolism or from errors during DNA replication and meiosis (9–11). DSBs were traditionally thought to be repaired by either of two distinct pathways, distinguished by the presence or absence of homology: homologous recombination repair (HR) and classical nonhomologous end-joining (NHEJ), respectively (**Fig. 1A** and **B**).

Microhomology-mediated end-joining as a third pathway for DSB repair

Evidence for a third pathway that could repair DSBs first emerged following the observation that NHEJ-deficient yeast and hamster cells retained some degree of end-joining activity (12, 13). Because this pathway was first reported in NHEJ-deficient cells, it was initially termed alternative end-joining and viewed as a backup mechanism if HR and NHEJ were unavailable. Subsequent work, especially in mammalian cells, has suggested a more substantial role for this pathway, now more commonly referred to as microhomology-mediated end-joining (MMEJ; refs. 14,15).

[1]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts. [2]Department of Radiation Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts. [3]Center for DNA Damage and Repair, Dana-Farber Cancer Institute, Boston, Massachusetts.

Corresponding Author: Alan D. D'Andrea, Dana-Farber Cancer Institute, 450 Brookline Avenue, HIM 243, Boston, MA 02215. Phone: 617-632-2112, Fax: 617-632-6069; E-mail: alan_dandrea@dfci.harvard.edu

Cancer Res 2020;80:4593–600

doi: 10.1158/0008-5472.CAN-20-1672

©2020 American Association for Cancer Research.

Current Understanding of DSB Repair by MMEJ

The MMEJ pathway is a stepwise pathway (**Fig. 1C**). It is initiated by end-resection near the DSB exposing short regions of complementary sequences ranging from 2 to 20 nucleotides (microhomologies). These microhomologies are used to align the DNA ends with the occurrence of end-bridging. Next, the resultant 5′ flaps, created following alignment, are processed. A polymerase then fills in any gaps, followed ultimately by ligation (16).

End-resection

The MMEJ pathway is initiated by end-resection. Following a DSB, PARP-1 competes with Ku70-Ku80 for binding to DSBs ends, facilitating, though not essential for, recruitment of end-resection factors (17, 18). As in HR, the initial end-resection in MMEJ is performed by the MRN complex (MRE11, RAD50, and NBS1) stimulated by C-terminal binding protein interacting protein (CtIP), with the initial creation of a 3′ single-strand region (19–21). If the microhomologies are distant, studies have suggested that further resection is performed by Bloom helicase-DNA2 helicase/nuclease and exonuclease 1 (EXO1); however, their roles have not been genetically confirmed. Furthermore, opposing data suggests that long-range resection by EXO1 decreases MMEJ (14, 19, 20, 22–27).

Normally, ssDNA is bound by replication protein A (RPA) to stabilize the secondary structure. In yeast, RPA functions as a negative regulator of MMEJ. MMEJ is dramatically increased following end resection while HR remains relatively stable when a hypomorphic-mutant allele of one of components of yeast RPA is expressed (28). Recently, Shukla and colleagues demonstrated that the 5-hydroxymethylcytosine binding, ESC-specific protein (HMCES), previously shown to protect stalled replication forks, can bind ssDNA and promote MMEJ repair (29, 30). Furthermore, HMCES binds class switch regions and protects ssDNA overhangs to promote MMEJ, and mice deficient in HMCES have significant defects in class switch recombination, similar to mice deficient in the specialized DNA polymerase of the MMEJ pathway, polymerase theta (POLQ; refs. 30, 31). Taken together, these studies suggest that RPA is the critical ssDNA-binding protein in HR repair, while HMCES is the critical ssDNA-binding protein in MMEJ (**Fig. 1B** and **C**).

End-bridging and alignment

End-bridging and alignment of microhomology sequences are next accomplished through the activities of PARP-1, the MRN complex,

Downloaded from cancerres.aacrjournals.org on December 29, 2020. © 2020 American Association for Cancer Research.

Published OnlineFirst July 10, 2020; DOI: 10.1158/0008-5472.CAN-20-1672

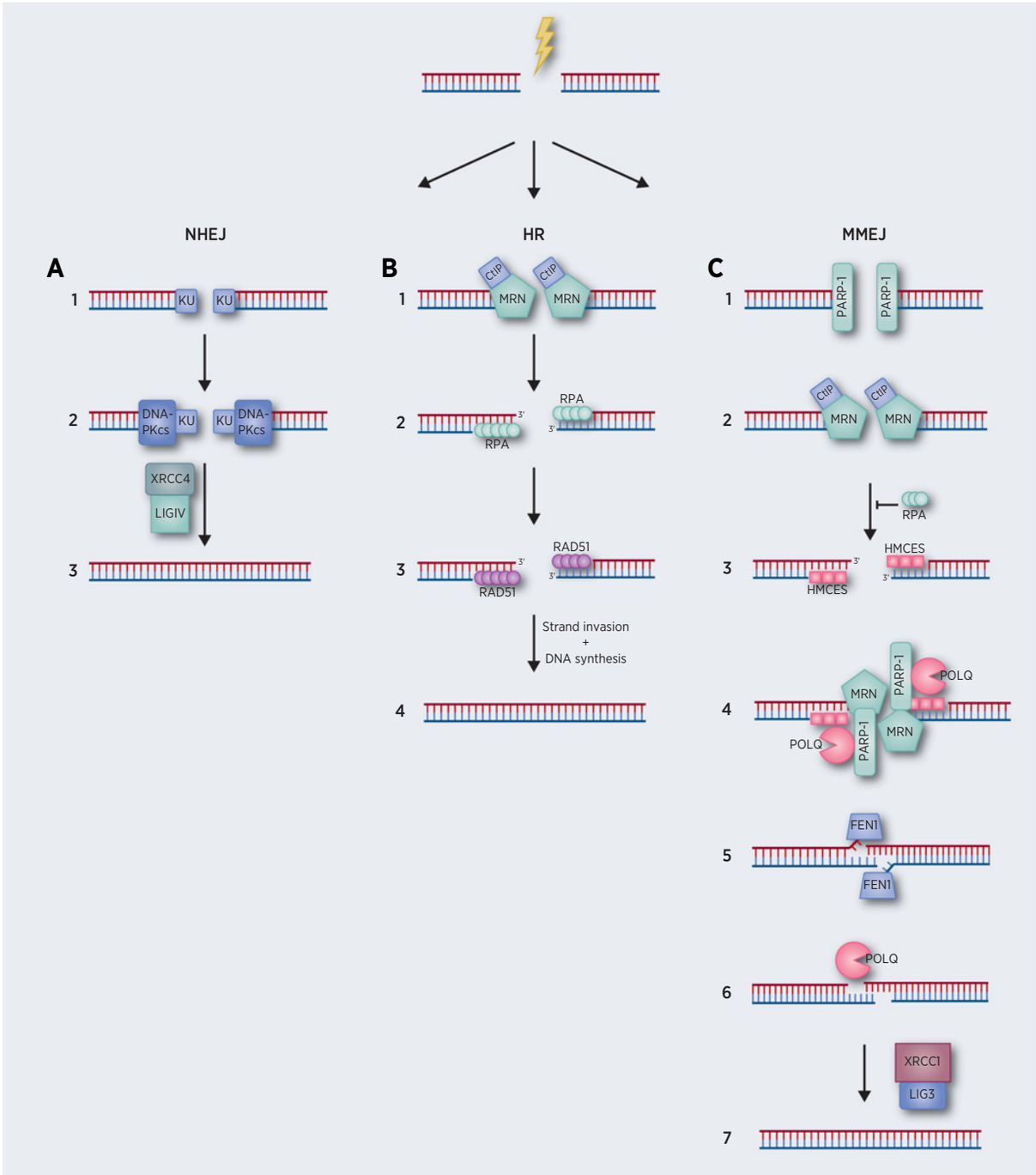


Figure 1.

DSB repair pathways. Repair of DSBs is a stepwise pathway. **A,** NHEJ. **1,** Ku70-Ku80 (Ku) heterodimer recognizes DSB end. **2,** Recruitment of DNA-PKcs to Ku70-Ku80 heterodimer-DSB ends. If incompatible ends, further processing is required. **3,** Ligation of DNA ends by XRCC4-DNA Ligase IV. **B,** HR. **1,** MRN, stimulated by CtIP, performs end-resection. **2,** ssDNA is bound by RPA. **3,** RAD51 replaces RPA, forming a RAD51-ssDNA nucleoprotein filament. **4,** RAD51-ssDNA nucleoprotein filament performs strand invasion of the sister chromatid, leading to template-dependent strand extension, followed by strand annealing. **C,** MMEJ. **1,** PARP-1 competes with Ku70–80 to bind DSBs ends, recruiting MRN and CtIP. **2,** End-resection is initiated by MRN, stimulated by CtIP, in a stepwise manner first, with the creation of a 3′ single-strand region endonucleolytically, subsequently followed by 3′ to 5′ exonucleolytic digestion. **3,** Competition for binding of single-stranded DNA by RPA or HMCES with HMCES-ssDNA coating favored by MMEJ. **4,** End-bridging and microhomology alignment. End-bridging is mediated by a combination of PARP-1, MRN, and POLQ. **5,** Flap processing. Nonhomologous 3′ tails are removed by ERCC1/XPF, FEN1, and likely as-yet identified endonucleases. Nonhomologous 5′ tails are removed by FEN1. **6,** POLQ gap filling. POLQ can uniquely prime DNA synthesis from nonoptimal base-pairing, leading to distinctive templated insertions. **7,** Ligation of ends by LIG3/XRCC1. See text for more details.

Downloaded from cancerres.aacrjournals.org on December 29, 2020. © 2020 American Association for Cancer Research.

Published OnlineFirst July 10, 2020; DOI: 10.1158/0008-5472.CAN-20-1672

and POLQ. Recent *in vitro* studies suggest that PARP-1 may compete with Ku70-80 for binding to DSBs and may exhibit end-bridging activity through an unknown mechanism (32). PARP-1 also appears to have a role in the recruitment of POLQ to sites of DNA damage. Regarding the MRN complex, *in vitro* studies have shown that cohesive ends inhibit MRE11-mediated degradation. If a region of microhomology is revealed during resection, MRE11 stalls, thereby stabilizing the junction site (33, 34). In addition, scanning force microscopy revealed a complex of MRE11 and RAD50 that can tether linear DNA molecules (35). These findings suggest that the MRN complex, while performing nucleolytic degradation, can facilitate DNA end-bridging and microhomology alignment. POLQ contains a N-terminal helicase domain, a long, unstructured central region, and a C-terminal polymerase domain (36). The crystal structure of the POLQ helicase domain revealed a tetrameric organization, suggesting a role of microhomology alignment and strand annealing in preparing a substrate for the polymerase domain (37). Indeed, both domains are required for MMEJ, and mutation of a conserved residue in the helicase domain reduces MMEJ (38, 39).

Tail and flap processing

The processes of DNA end-bridging and alignment result in the generation of nonhomologous 3′ tails. These tails must be removed to complete the repair process. In some DNA repair processes, such as single-strand annealing, nonhomologous 3′ tails are removed by the heterodimeric structure-specific endonuclease ERCC1/XPF (excision repair cross-complementing group 1/Xeroderma pigmentosum, complementation group F; ref. 40). ERCC1/XPF is also employed in nucleotide excision repair, in the Fanconi Anemia interstrand crosslink repair, and in base excision repair (41–43). Consequently, it was originally hypothesized that ERCC1/XPF was the endonuclease complex that removes nonhomologous 3′ flaps in MMEJ. However, murine ERCC1$^{-/-}$ cells demonstrate only a minor MMEJ defect (44). These studies suggest that another functionally distinct flap endonuclease, such as flap endonuclease 1 (FEN1), might be involved in the removal of nonhomologous 3′ tails in the MMEJ pathway.

Gap filling in MMEJ is performed by POLQ, through the process of strand displacement and DNA synthesis, and this process creates displaced 5′ ssDNA flaps requiring subsequent removal (45). FEN1 is a structure-specific endonuclease that recognizes and cleaves these 5′ ssDNA flaps (46). Using a cell-free system created from rat testes and thymus to study proteins involved in MMEJ, Sharma and colleagues demonstrated that siRNA-mediated depletion of FEN1 strongly reduces MMEJ activity (47). More recently, genetic screens to identify genes required for the upregulation of MMEJ and the survival of BRCA2-deficient cells identified FEN1, and specifically the 5′ flap endonuclease activity of FEN1 (48). Furthermore, siRNA-mediated depletion of FEN1 resulted in a defect in MMEJ repair in a cell-based MMEJ reporter assay, confirming that FEN1 is an essential enzyme in MMEJ and suggesting that FEN1 may be the critical endonuclease that cleaves 5′ flaps created by POLQ.

POLQ gap filling

Once the DNA microhomology regions are aligned and the 3′ nonhomologous tails have been removed, there will be gaps between the duplex DNA. The specialized DNA polymerase involved in MMEJ, POLQ, was first identified from studies in *Drosophila* by using DSBs induced by P-element transposition or sequence-specific endonuclease (38, 49). Subsequent studies have shown that POLQ is essential for MMEJ in all mammalian species examined to date (50). POLQ is a 290 kDA A-family DNA polymerase with a unique structure containing an N-terminal helicase domain, a long, unstructured central region, and a C-terminal polymerase domain (36).

The N-terminal domain of POLQ contains a superfamily 2 Hel308-typeS helicase domain. The helicase domain has both ATPase and DNA unwinding activities, which can act independently of each other. The ATPase activity is required for both the suppression of HR by binding RAD51 and inhibiting its assembly along ssDNA, and the survival of HR-deficient cells (36, 51). Expression of POLQ mutated at ATPase catalytic residues did not reduce RAD51 foci formation, nor decrease recombination frequency (51). The ATPase activity also facilitates removal of RPA from ssDNA to allow MMEJ and inhibit HR (52). Conceivably, with removal of RPA from ssDNA, HMCES could bind, protecting ssDNA overhangs and promoting MMEJ (30). *In vitro*, the POLQ helicase domain can unwind DNA, thereby facilitating displacement synthesis by the POLQ polymerase domain. Further studies are required to understand the coordination of these domains *in vivo* (53).

The long, unstructured central region of POLQ contains a RAD51 interaction motif. Protein interaction studies of POLQ demonstrated that this site directly interacts with RAD51 via three distinct motifs, though only one region (amino acids 847–894) was both necessary and sufficient for RAD51 binding. Interaction of the central domain of POLQ with RAD51 inhibited RAD51–ssDNA nucleoprotein filament assembly. In this way, POLQ acts as an antirecombinase capable of inhibiting HR and promoting MMEJ (51). This process appears to mirror that of the POLQ helicase domain, which can displace RPA, inhibit HR, and promote MMEJ (52). These interactions of the POLQ helicase domain with RPA and the POLQ central domain with RAD51, respectively, demonstrate the competitive nature between MMEJ and HR, occurring at the level of ssDNA binding. In addition, the central domain of POLQ regulates POLQ multimerization (54).

The C-terminal domain of POLQ contains an A-family DNA polymerase domain that performs gap filling. *In vitro* and cellular experiments have elucidated POLQ's gap-filling mechanism. POLQ anneals to short regions of ssDNA microhomology of 2 to 6 base-pairs (37, 55, 56). Because POLQ can uniquely prime DNA synthesis from nonoptimal base-pairing, leading to the introduction of insertions at the break sites, with the 3′ flaps acting as a template, it is highly error-prone (45, 57, 58). In addition, the POLQ polymerase is proofreading deficient and quite promiscuous because of its robust terminal transferase activity (59). Biochemically, which domains of POLQ are necessary for DNA synthesis depend on the substrate (54). Purified POLQ polymerase domain alone is active on short ssDNA and short 3′ overhangs. Longer ssDNA substrates require both the POLQ N-terminal helicase domain and the C-terminal polymerase domain, and it is the helicase domain's DNA annealing ability that is required rather than its helicase activity. Furthermore, extension of long ssDNA substrates requires multimers of POLQ, and multimerization requires the central region (54).

End ligation

DSB repair by MMEJ is completed upon ligation of the DNA ends. In mammalian cells, three DNA ligases are encoded: DNA Ligase I, DNA Ligase III, and DNA Ligase IV (60). Ligase IV functions in NHEJ and its deficiency does not impact MMEJ (61). Substantial literature has demonstrated that Ligase III is the major ligase in MMEJ. First, *in vitro* biochemical and plasmid-rejoining studies identified Ligase III (32, 62). Second, these studies were confirmed *in vivo*, where mouse cells deficient in Ligase III demonstrated reduced frequency of

Downloaded from cancerres.aacrjournals.org on December 29, 2020. © 2020 American Association for Cancer Research.

Published OnlineFirst July 10, 2020; DOI: 10.1158/0008-5472.CAN-20-1672

chromosomal translocations. In this case, the few remaining translocations lacked areas of microhomology commonly seen in MMEJ (63). Third, sequential depletion of Ligase III followed by Ligase I depletion was additive, suggesting the latter is a backup ligase (63). Ligase III forms a stable complex with the scaffold protein, X-ray repair cross-complementing protein 1 (XRCC1), and both proteins in the complex interact with PARP-1 (64, 65). This interaction with PARP-1 is not required for recruitment of Ligase III/XRCC1 to DSBs as there is no reduction in recruitment in the presence of PARP inhibition (15). Rather, it appears that the MRN complex might recruit Ligase III/XRCC1 to DSBs during repair by MMEJ (66).

MMEJ's Role in Human Disease

Because of the mutagenic nature of POLQ, it is likely that POLQ and hence MMEJ are only activated under rare circumstances. Indeed, POLQ protein expression is highly limited in normal tissue (51, 67). These rare circumstances might include the following. First, as POLQ expression greatly increases in HR-deficient cancers, the presence of active HR could inhibit expression of POLQ to prevent a more mutagenic DSB repair (51, 68). Perhaps, the BRCA1 protein, which has some transcriptional functions, reduces the transcription of the POLQ gene (69). Second, embryonic stem cells and somatic cells rely on different DSB repair pathways (70). MMEJ might play a greater role during gametogenesis, embryogenesis, or in hematopoietic stem cells as expression of POLQ is greatest in the testis, human placental tissue, and hematopoietic stem cells (36, 71). Third, POLQ appears to have a prominent role in the repair of four-stranded guanine-rich structures (G-quadruplex; refs. 72–74). These structures can have both beneficial and deleterious functions. On the one hand, G-quadruplexes can protect telomeres, regulate transcription, and promote immunoglobulin gene recombination. On the other hand, these structures are prone to genomic instability and can inhibit DNA replication through fork stalling (75). Indeed, mice deficient in POLQ have increased genomic instability (76). Finally, POLQ might play a preferred role in the repair of repetitive DNA, as HR could be a deleterious option given the possibility of recombination between repetitive elements.

MMEJ and cancer

One of the hallmarks of cancer is genomic instability (77). The repair of DSBs by MMEJ is an intrinsically mutagenic pathway. The creation of templated insertions at DSB junction sites is characteristic of MMEJ (38, 51, 78, 79). Templated insertions are small insertions, typically, 3 to 30 bps, synthesized from DNA flanking the DSB junction into sites of microhomology deletions (38). Templated insertions are abundant in the ClinVar database, raising the question of MMEJ's contribution to malignancy (80). Microhomology signatures have been observed at sites of oncogenic chromosomal translocation breakpoints in primary human cancer cells, raising the possibility that MMEJ could be the causative mechanism of these translocation (23, 26). Furthermore, review of The Cancer Genome Atlas demonstrates that POLQ is overexpressed in breast, lung, bladder, colorectal, gastric, glioblastoma, pancreatic, prostate, melanoma, and uterine cancers, correlating with poor prognosis (67, 81). The elevated POLQ expression in these tumors may result, at least in part, from an underlying HR deficiency.

Cancers deficient in HR upregulate the expression of POLQ as a survival strategy. Aberrant use of MMEJ could create more mutations in cancers, promoting both cancer growth and resistance (82). These cancers become addicted to POLQ expression, routing the repair of DSBs to MMEJ, as evidenced by the accumulation of a distinct pattern of MMEJ-mediated templated insertions (51, 68, 83). This mutation signature, initially termed mutation signature 3, was identified by whole-genome sequence analysis of HR-deficient tumor and thought to typify tumors that were potentially HR-deficient (84). In actuality, this signature better represents upregulation and addiction to POLQ rather than a specific signature for HR deficiency (85). Indeed, in a recent revision of mutational signatures in cancers, signature 3 has been further resolved. The MMEJ specific signature, now defined as insertion–deletion signature 6, is characterized predominantly by greater than 5 bp deletions with overlapping microhomology at deletion boundaries and is correlated with single-base substitution signature 3 (86).

Therapeutic Application of MMEJ for Cancer Treatment

Loss of DNA repair pathways is a common feature of cancers and likely provides a growth advantage (87). Concomitantly, there is increased dependence on the remaining DNA repair pathways. This dependence renders the cancer vulnerable to inhibition of the remaining DNA repair pathways and is the basis of the synthetic lethality. This concept of synthetic lethality has been successfully exploited by the use of PARP inhibitors (PARPi) in breast and ovarian cancers harboring inactivating mutations in BRCA1 or BRCA2, and could similarly be employed with MMEJ inhibition (88, 89). Because PARP-1 is also a component of MMEJ, PARPis also inhibit MMEJ (15). Indeed, inhibition of MMEJ appears to be a major mechanism of synthetic lethality of PARPis. Loss of expression of POLQ or other MMEJ genes in HR-deficient cells is synthetically lethal (51, 68). Some cancers also have mutations in NHEJ genes, and these cancers exhibit synthetically lethality from depletion of POLQ (58, 90). Importantly, some HR-deficient tumors acquire PARPi resistance by downregulating NHEJ. This offers the opportunity for using a POLQ inhibitorbecause NHEJ-deficient cells are also dependent on MMEJ. Taken together, these observations provide a strong rationale for targeting MMEJ in the treatment of cancers, including PARPi-resistant cancers (**Fig. 2**; refs. 88, 89).

Synthetic lethality: targeting MMEJ and HR

POLQ has limited protein expression in normal tissue but is overexpressed in wide variety of cancers (51, 67, 81). This overexpression is increased with concomitant p53 loss (91). These cancers may be addicted to POLQ expression because of the need to repair DSBs (51, 68, 83). Furthermore, POLQ, via its ATPase activity and RAD51-binding ability, can prevent the accumulation of toxic RAD51 on resected ssDNA, thereby enhancing cancer cell survival. Loss of POLQ in these tumors results in cell death by impairing MMEJ-mediated insertions of DSBs and promoting the buildup of toxic RAD51 (51, 68). Thus, POLQ is an attractive target for drug development in HR-deficient cancers (**Fig. 2A**; ref. 92).

The POLQ protein has three domains: a N-terminal helicase domain, a long, unstructured central region, and a C-terminal polymerase domain (36). POLQ inhibitors could be designed to target either the N-terminal helicase domain or the C-terminal polymerase domain. Targeting the polymerase domain would inhibit repair of resected DSBs at the gap-filling step in HR-deficient cancers. Targeting the helicase domain, however, might be a better option, as MMEJ repair of DSBs would be inhibited and toxic RAD51 foci would accumulate (51). POLQ inhibitors are currently undergoing active preclinical development. Indeed, Zhou and colleagues have recently identified the antibiotic novobiocin as a specific POLQ helicase

Downloaded from cancerres.aacrjournals.org on December 29, 2020. © 2020 American Association for Cancer Research.

Published OnlineFirst July 10, 2020; DOI: 10.1158/0008-5472.CAN-20-1672

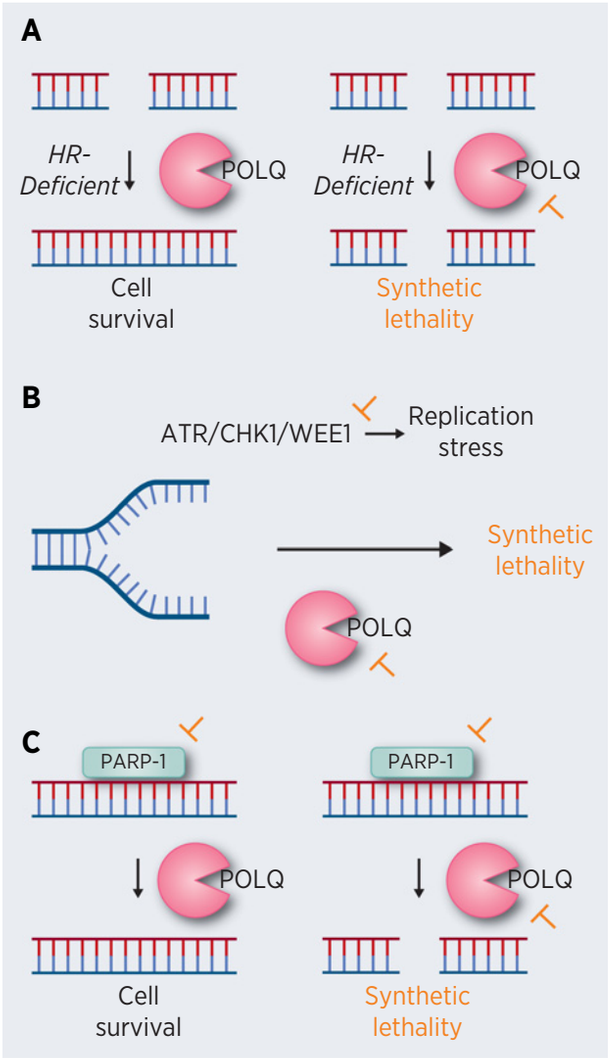


Figure 2.

Therapeutic application of MMEJ for cancer treatment. **A,** HR-deficient cancers require MMEJ to repair DSBs. Inhibition of POLQ results in synthetic lethality in HR-deficient cancers. **B,** POLQ deficiency is synthetically lethal with ATR inhibition. Inhibition of the ATR-CHK1-WEE1 signaling cascade leading to replication stress in combination with POLQ inhibition would result in synthetic lethality. **C,** PARP inhibition results in DSBs that can be repaired by MMEJ. Inhibition of both PARP and POLQ (MMEJ) induces synthetic lethality.

domain inhibitor that selectively kills HR-deficient cancer cells *in vitro* and *in vivo* (93).

In addition to POLQ, FEN1 is an essential enzyme in MMEJ, required at least for removal of the 5′ flap and potentially the 3′ flap (47, 48). Depletion of FEN1 results in a defect in MMEJ repair and is also synthetically lethal with HR deficiency, recapitulating the synthetic lethality observed with POLQ inhibition (48, 51, 68). Hydroxyurea-based FEN1 inhibitors have been developed that induce synthetic lethality and did not induce toxicity as a monotherapy (94). Importantly, FEN1 is not exclusive to the MMEJ pathway. It is also involved in long-patch base excision repair, thereby adding to its value as a target for anticancer drug development.

Synthetic lethality: targeting more than HR deficiency

Synthetic lethality occurs not only in HR-deficient cells but also in cells deficient in NHEJ, where depletion of POLQ decreases cell survival (58). There are only a few cancers with underlying defects in NHEJ (90). However, DNA-dependent protein kinase (DNA-PK) is a key component of the NHEJ pathway, and DNA-PK inhibitors are currently under development because they block NHEJ and enhance the activity of DNA-damaging agents (95, 96). DNA-PK inhibition results in a significant increase in MMEJ-mediated templated insertions and, coupled with genetic downregulation of POLQ, results in significantly reduced viability (91). These results provide evidence for the application of POLQ inhibition beyond HR-deficient cancers and for combination treatment strategies (**Fig. 2B** and **C**).

Cell-cycle DNA damage checkpoints allow time for repair by the DDR. The protein kinases, ataxia-telangiectasia-mutated (ATM) and ataxia-telangiectasia and Rad3-related protein (ATR), play central roles in these DNA damage checkpoints, creating an opportunity for synthetic lethality (97). ATM is activated by DSBs and initiates a DDR via a signaling cascade of phosphorylation with activation of checkpoint kinase 2 (CHK2) and ultimately p53, as well as activation of the HR pathway. ATR, in contrast, is activated by replication stress. Replication stress is common in oncogene-driven cancer cells, but not in normal tissue (98). Through its kinase activity, ATR activates checkpoint kinase 1 (CHK1) and ultimately Wee1-like protein kinase (WEE1), leading to prolongation of the S/G_2 cell-cycle phase (99).

Replication stress activates the ATR pathway to prevent replication fork collapse and allow DNA replication to continue (99). By creating single-strand breaks at replication forks, using either a Cas9 nickase or a topoisomerase inhibitor, Wang and colleagues demonstrated synthetic lethality with POLQ deficiency, demonstrating a role for MMEJ in replication stress tolerance (100). Inhibition of the ATR-CHK1-WEE1 signaling cascade is likely to prove useful in combination with POLQ inhibition (**Fig. 2B**). Inhibitors of ATR, CHK1, or WEE1 are currently under investigation in clinical trials and have activity through their ability to promote replication stress. Gemcitabine is another agent that can promote replication stress. A randomized phase II clinical trial of the ATR inhibitor M6620 plus gemcitabine demonstrated efficacy through the induction of replication stress in cancer cells (101). Thus, combination treatment strategies involving induction of replication stress with gemcitabine coupled with POLQ inhibition provides a strong rationale for future trials. Finally, a targeted CRISPR screen of 309 murine genes known to be involved in DDR uncovered 140 genes synthetically lethal with POLQ deficiency with the majority of these identified genes being previously unknown. It is likely that whole genome-wide CRISPR screening to identify genetic vulnerabilities that interact with POLQ will reveal new combination therapy targets.

Targeting MMEJ to overcome therapy resistance

The success of PARPis in the treatment of breast and ovarian cancers harboring inactivating mutations in BRCA1 or BRCA2 has demonstrated the power of synthetic lethality (88, 89). Unfortunately, these cancers ultimately acquire PARPi resistance, with the primary mechanisms of resistance being either restoration of HR repair or replication fork stabilization (102). Reactivation of HR in PARPi-resistant cancers would be predicted to alleviate dependence on MMEJ for DSB repair. However, POLQ via its RAD51-binding ability, prevents accumulation of toxic RAD51 secondary to abortive HR. Though HR is often reactivated in PARPi-resistant cancers, through somatic reversion of BRCA1/2 mutations or via downregulation of NHEJ, inhibition of POLQ could still result in accumulation of toxic RAD51,

Downloaded from cancerres.aacrjournals.org on December 29, 2020. © 2020 American Association for Cancer Research.

Published OnlineFirst July 10, 2020; DOI: 10.1158/0008-5472.CAN-20-1672

resulting in cellular toxicity (51). Furthermore, POLQ and the MMEJ are required for coping with replication stress (100). Inhibition of POLQ in PARPi-resistant cancers would still result in increased replication stress and cellular toxicity. Indeed, PARPi-resistant tumors are sensitive to the specific POLQ inhibitor, Novobiocin, both *in vitro* and *in vivo* (93). Targeting MMEJ by inhibition of POLQ might help prevent the development of PARPi resistance or resensitize PARPi-resistant cancers (**Fig. 2C**).

Cancer cells often lose expression of p53 or express a mutant version (103). Loss of p53 function in cancer cells increases resistance to ionizing radiation, a common treatment modality for cancer (104, 105). Given POLQ is overexpressed in p53-deficient cancers, this raises the question of whether POLQ inhibition might resensitize p53-deficient cancers to ionizing radiation or if POLQ inhibition could be used as an adjuvant with radiation treatment (91). Indeed, treatment of bone marrow stromal cells derived from POLQ-deficient mice demonstrated increased sensitivity to ionizing radiation (106).

Biomarker-guided development of MMEJ inhibitors

The development of predictive biomarkers to guide the application of MMEJ inhibition in cancers is crucial. There are at least four potential MMEJ/POLQ biomarkers. First, because POLQ is overexpressed in HR-deficient cancers or in p53-deficient cancers, dependence on POLQ activity would be predicted by upregulation of POLQ protein expression (51, 68, 83, 91). Assessment of POLQ expression in cancers, via IHC or RT-PCR, could serve as a predictive biomarker for response to POLQ inhibition. Second, depletion of POLQ in HR-deficient cancers leads to RAD51 accumulation (51). Assessing RAD51 accumulation could serve as a pharmacodynamic biomarker for POLQ inhibition. Third, the MMEJ pathway is required for alleviation of replication stress and is synthetically lethal with ATR inhibition (100). Replication stress activates the ATR pathway (99). Phosphorylation of RPA is downstream of ATR activation and is an IHC marker of replication stress (107). Thus, the assessment of phosphorylated-RPA could be a biomarker of the effectiveness of an MMEJ inhibitor. Finally, templated insertions are a biomarker of ongoing (or previous) MMEJ activity in the tumor cells and are a potential indicator of POLQ inhibitor efficacy (84, 85). The development of biomarkers will be extremely useful in future clinical trials of POLQ inhibitors.

Summary

DNA is under constant threat from both endogenous and exogenous damage, and the maintenance of genomic integrity is critical for cell survival (1). The most deleterious type of DNA damage is DSBs. Mammalian cells have three main pathways to repair DSBs: NHEJ, HR, and MMEJ. Initially, NHEJ and HR were the best studied DSB repair pathways and were thought to be the dominant pathways, with MMEJ as a weaker alternative pathway. However, evidence has since emerged to support MMEJ as a more substantial third pathway given the functional relevance of the pathway to both DSB repair and human disease. Indeed, the prevalence of a distinctive MMEJ mutational signature of templated insertions in up to 20% of all human cancers provides evidence of the widespread relevance of this pathway. Given the success of targeting DDR dependencies as a therapeutic approach to cancers, for example, with PARPis in HR-deficient ovarian and breast cancers, targeting MMEJ is especially enticing. Furthermore, targeting MMEJ could be helpful in the setting of acquired drug resistance to prior therapies.

Disclosure of Potential Conflicts of Interest

A. D. D'Andrea reports grants and personal fees from Eli Lilly and Co. (consulting and sponsored research agreements) and Merck EMD Serono (consulting and sponsored research agreement); and personal fees from Sierra Oncology (consulting/advisory board member; formerly had a sponsored research agreement) outside the submitted work. No potential conflicts of interest were disclosed by the other author.

Acknowledgments

This research was supported by grants from the U.S. NIH (R37HL052725, P01HL048546), the U.S. Department of Defense (BM110181, BC151331P1), the Breast Cancer Research Foundation, and the Fanconi Anemia Research Fund (to A.D. D'Andrea). A.D. D'Andrea is also supported by a Pancreatic Cancer Collective New Therapies Challenge Grant, an initiative of the Lustgarten Foundation and Stand Up To Cancer, Grant number SU2C-AACR-PCC-02-18. Stand Up To Cancer is a division of the Entertainment Industry Foundation. The research grant is administered by the American Association for Cancer Research, the Scientific Partner of SU2C. We thank Connor Clairmont, Wei-Chi Tsai, and Jia Zhou for critical reading of the manuscript.

Received May 18, 2020; revised June 13, 2020; accepted July 7, 2020; published first July 10, 2020.

References

1. Ceccaldi R, Rondinelli B, D'Andrea AD. Repair pathway choices and consequences at the double-strand break. Trends Cell Biol 2016;26:52–64.
2. Sung P, Klein H. Mechanism of homologous recombination: mediators and helicases take on regulatory functions. Nat Rev Mol Cell Biol 2006;7:739–50.
3. Heyer W-D, Ehmsen KT, Liu J. Regulation of homologous recombination in eukaryotes. Annu Rev Genet 2010;44:113–39.
4. Chapman JR, Taylor MRG, Boulton SJ. Playing the end game: DNA double-strand break repair pathway choice. Mol Cell 2012;47:497–510.
5. Bunting SF, Nussenzweig A. End-joining, translocations and cancer. Nat Rev Cancer 2013;13:443–54.
6. Helleday T, Eshtad S, Nik-Zainal S. Mechanisms underlying mutational signatures in human cancers. Nat Rev Genet 2014;15:585–98.
7. Carvalho CMB, Lupski JR. Mechanisms underlying structural variant formation in genomic disorders. Nat Rev Genet 2016;17:224–38.
8. Nilles N, Fahrenkrog B. Taking a bad turn: compromised DNA damage response in leukemia. Cells 2017;6:11.
9. Dudley DD, Chaudhuri J, Bassing CH, Alt FW. Mechanism and control of V(D)J recombination versus class switch recombination: similarities and differences. Adv Immunol 2005;86:43–112.
10. Neale MJ, Keeney S. Clarifying the mechanics of DNA strand exchange in meiotic recombination. Nature 2006;442:153–8.
11. Lieber MR. NHEJ and its backup pathways in chromosomal translocations. Nat Struct Mol Biol 2010;17:393–5.
12. Boulton SJ, Jackson SP. Saccharomyces cerevisiae Ku70 potentiates illegitimate DNA double-strand break repair and serves as a barrier to error-prone DNA repair pathways. EMBO J 1996;15:5093–103.
13. Kabotyanski EB, Gomelsky L, Han JO, Stamato TD, Roth DB. Double-strand break repair in Ku86- and XRCC4-deficient cells. Nucleic Acids Res 1998;26:5333–42.
14. Truong LN, Li Y, Shi LZ, Hwang PY-H, He J, Wang H, et al. Microhomology-mediated end joining and homologous recombination share the initial end resection step to repair DNA double-strand breaks in mammalian cells. Proc Natl Acad Sci U S A 2013;110:7720–5.
15. Dutta A, Eckelmann B, Adhikari S, Ahmed KM, Sengupta S, Pandey A, et al. Microhomology-mediated end joining is activated in irradiated human cells due to phosphorylation-dependent formation of the XRCC1 repair complex. Nucleic Acids Res 2017;45:2585–99.
16. Sallmyr A, Tomkinson AE. Repair of DNA double-strand breaks by mammalian alternative end-joining pathways. J Biol Chem 2018;293:10536–46.
17. Wang M, Wu W, Wu W, Rosidi B, Zhang L, Wang H, et al. PARP-1 and Ku compete for repair of DNA double strand breaks by distinct NHEJ pathways. Nucleic Acids Res 2006;34:6170–82.

Downloaded from cancerres.aacrjournals.org on December 29, 2020. © 2020 American Association for Cancer Research.

Published OnlineFirst July 10, 2020; DOI: 10.1158/0008-5472.CAN-20-1672

18. Haince J-F, McDonald D, Rodrigue A, Déry U, Masson J-Y, Hendzel MJ, et al. PARP1-dependent kinetics of recruitment of MRE11 and NBS1 proteins to multiple DNA damage sites. J Biol Chem 2008;283:1197–208.
19. Garcia V, Phelps SEL, Gray S, Neale MJ. Bidirectional resection of DNA double-strand breaks by Mre11 and Exo1. Nature 2011;479:241–4.
20. Cannavo E, Cejka P. Sae2 promotes dsDNA endonuclease activity within Mre11-Rad50-Xrs2 to resect DNA breaks. Nature 2014;514:122–5.
21. Daley JM, Niu H, Miller AS, Sung P. Biochemical mechanism of DSB end resection and its regulation. DNA Repair 2015;32:66–74.
22. Dinkelmann M, Spehalski E, Stoneham T, Buis J, Wu Y, Sekiguchi JM, et al. Multiple functions of MRN in end-joining pathways during isotype class switching. Nat Struct Mol Biol 2009;16:808–13.
23. Rass E, Grabarz A, Plo I, Gautier J, Bertrand P, Lopez BS. Role of Mre11 in chromosomal nonhomologous end joining in mammalian cells. Nat Struct Mol Biol 2009;16:819–24.
24. Xie A, Kwok A, Scully R. Role of mammalian Mre11 in classical and alternative nonhomologous end joining. Nat Struct Mol Biol 2009;16:814–8.
25. Lee-Theilen M, Matthews AJ, Kelly D, Zheng S, Chaudhuri J. CtIP promotes microhomology-mediated alternative end joining during class-switch recombination. Nat Struct Mol Biol 2011;18:75–9.
26. Zhang Y, Jasin M. An essential role for CtIP in chromosomal translocation formation through an alternative end-joining pathway. Nat Struct Mol Biol 2011;18:80–4.
27. Deng SK, Yin Y, Petes TD, Symington LS. Mre11-Sae2 and RPA collaborate to prevent palindromic gene amplification. Mol Cell 2015;60:500–8.
28. Deng SK, Gibb B, de Almeida MJ, Greene EC, Symington LS. RPA antagonizes microhomology-mediated repair of DNA double-strand breaks. Nat Struct Mol Biol 2014;21:405–12.
29. Mohni KN, Wessel SR, Zhao R, Wojciechowski AC, Luzwick JW, Layden H, et al. HMCES maintains genome integrity by shielding abasic sites in single-strand DNA. Cell 2019;176:144–153.e13.
30. Shukla V, Halabelian L, Balagere S, Samaniego-Castruita D, Feldman DE, Arrowsmith CH, et al. HMCES functions in the alternative end-joining pathway of the DNA DSB repair during class switch recombination in B cells. Mol Cell 2020;77:384–394.e4.
31. Masuda K, Ouchida R, Takeuchi A, Saito T, Koseki H, Kawamura K, et al. DNA polymerase theta contributes to the generation of C/G mutations during somatic hypermutation of Ig genes. Proc Natl Acad Sci U S A 2005;102:13986–91.
32. Audebert M, Salles B, Calsou P. Involvement of poly(ADP-ribose) polymerase-1 and XRCC1/DNA ligase III in an alternative route for DNA double-strand breaks rejoining. J Biol Chem 2004;279:55117–26.
33. Paull TT, Gellert M. The 3′ to 5′ exonuclease activity of Mre 11 facilitates repair of DNA double-strand breaks. Mol Cell 1998;1:969–79.
34. Paull TT, Gellert M. A mechanistic basis for Mre11-directed DNA joining at microhomologies. Proc Natl Acad Sci U S A 2000;97:6409–14.
35. Jäger R, Gisslinger H, Passamonti F, Rumi E, Berg T, Gisslinger B, et al. Deletions of the transcription factor Ikaros in myeloproliferative neoplasms. Leukemia 2010;24:1290–8.
36. Seki M, Marini F, Wood RD. POLQ (Pol theta), a DNA polymerase and DNA-dependent ATPase in human cells. Nucleic Acids Res 2003;31:6117–26.
37. Newman JA, Cooper CDO, Aitkenhead H, Gileadi O. Structure of the helicase domain of DNA polymerase theta reveals a possible role in the microhomology-mediated end-joining pathway. Structure 2015;23:2319–30.
38. Chan SH, Yu AM, McVey M. Dual roles for DNA polymerase theta in alternative end-joining repair of double-strand breaks in Drosophila. PLoS Genet 2010;6:e1001005.
39. Beagan K, Armstrong RL, Witsell A, Roy U, Renedo N, Baker AE, et al. Drosophila DNA polymerase theta utilizes both helicase-like and polymerase domains during microhomology-mediated end joining and interstrand crosslink repair. PLoS Genet 2017;13:e1006813.
40. Motycka TA, Bessho T, Post SM, Sung P, Tomkinson AE. Physical and functional interaction between the XPF/ERCC1 endonuclease and hRad52. J Biol Chem 2004;279:13634–9.
41. Bhagwat N, Olsen AL, Wang AT, Hanada K, Stuckert P, Kanaar R, et al. XPF-ERCC1 participates in the fanconi anemia pathway of cross-link repair. Mol Cell Biol 2009;29:6427–37.
42. Gregg SQ, Robinson AR, Niedernhofer LJ. Physiological consequences of defects in ERCC1-XPF DNA repair endonuclease. DNA Repair 2011;10:781–91.
43. Woodrick J, Gupta S, Camacho S, Parvathaneni S, Choudhury S, Cheema A, et al. A new sub-pathway of long-patch base excision repair involving 5′ gap formation. EMBO J 2017;36:1605–22.
44. Bennardo N, Cheng A, Huang N, Stark JM. Alternative-NHEJ is a mechanistically distinct pathway of mammalian chromosome break repair. PLoS Genet 2008;4:e1000110.
45. Kent T, Chandramouly G, McDevitt SM, Ozdemir AY, Pomerantz RT. Mechanism of microhomology-mediated end-joining promoted by human DNA polymerase θ. Nat Struct Mol Biol 2015;22:230–7.
46. Zou G-M, Maitra A. Small-molecule inhibitor of the AP endonuclease 1/REF-1 E3330 inhibits pancreatic cancer cell growth and migration. Mol Cancer Ther 2008;7:2012–21.
47. Sharma S, Javadekar SM, Pandey M, Srivastava M, Kumari R, Raghavan SC. Homology and enzymatic requirements of microhomology-dependent alternative end joining. Cell Death Dis 2015;6:e1697.
48. Mengwasser KE, Adeyemi RO, Leng Y, Choi MY, Clairmont C, D'Andrea AD, et al. Genetic screens reveal FEN1 and APEX2 as BRCA2 synthetic lethal targets. Mol Cell 2019;73:885–899.e6.
49. Yu AM, McVey M. Synthesis-dependent microhomology-mediated end joining accounts for multiple types of repair junctions. Nucleic Acids Res 2010;38:5706–17.
50. Sfeir A, Symington LS. Microhomology-mediated end joining: a back-up survival mechanism or dedicated pathway? Trends Biochem Sci 2015;40:701–14.
51. Ceccaldi R, Liu JC, Amunugama R, Hajdu I, Primack B, Petalcorin MIR, et al. Homologous-recombination-deficient tumours are dependent on Polθ-mediated repair. Nature 2015;518:258–62.
52. Mateos-Gomez PA, Kent T, Deng SK, McDevitt S, Kashkina E, Hoang TM, et al. The helicase domain of Polθ counteracts RPA to promote alt-NHEJ. Nat Struct Mol Biol 2017;24:1116–23.
53. Ozdemir AY, Rusanov T, Kent T, Siddique LA, Pomerantz RT. Polymerase θ-helicase efficiently unwinds DNA and RNA-DNA hybrids. J Biol Chem 2018;293:5259–69.
54. Black SJ, Ozdemir AY, Kashkina E, Kent T, Rusanov T, Ristic D, et al. Molecular basis of microhomology-mediated end-joining by purified full-length Polθ. Nat Commun 2019;10:4423.
55. Hogg M, Sauer-Eriksson AE, Johansson E. Promiscuous DNA synthesis by human DNA polymerase θ. Nucleic Acids Res 2012;40:2611–22.
56. McVey M, Lee SE. MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. Trends Genet 2008;24:529–38.
57. Yousefzadeh MJ, Wyatt DW, Takata K-I, Mu Y, Hensley SC, Tomida J, et al. Mechanism of suppression of chromosomal instability by DNA polymerase POLQ. PLoS Genet 2014;10:e1004654.
58. Wyatt DW, Feng W, Conlin MP, Yousefzadeh MJ, Roberts SA, Mieczkowski P, et al. Essential roles for polymerase θ-mediated end joining in the repair of chromosome breaks. Mol Cell 2016;63:662–73.
59. Zahn KE, Averill AM, Aller P, Wood RD, Doublié S. Human DNA polymerase θ grasps the primer terminus to mediate DNA repair. Nat Struct Mol Biol 2015;22:304–11.
60. Ellenberger T, Tomkinson AE. Eukaryotic DNA ligases: structural and functional insights. Annu Rev Biochem 2008;77:313–38.
61. Grawunder U, Zimmer D, Fugmann S, Schwarz K, Lieber MR. DNA ligase IV is essential for V(D)J recombination and DNA double-strand break repair in human precursor lymphocytes. Mol Cell 1998;2:477–84.
62. Wang H, Rosidi B, Perrault R, Wang M, Zhang L, Windhofer F, et al. DNA ligase III as a candidate component of backup pathways of nonhomologous end joining. Cancer Res 2005;65:4020–30.
63. Simsek D, Brunet E, Wong SY-W, Katyal S, Gao Y, McKinnon PJ, et al. DNA ligase III promotes alternative nonhomologous end-joining during chromosomal translocation formation. PLoS Genet 2011;7:e1002080.
64. Tomkinson AE, Sallmyr A. Structure and function of the DNA ligases encoded by the mammalian LIG3 gene. Gene 2013;531:150–7.
65. Okano S, Lan L, Tomkinson AE, Yasui A. Translocation of XRCC1 and DNA ligase IIIalpha from centrosomes to chromosomes in response to DNA damage in mitotic human cells. Nucleic Acids Res 2005;33:422–9.
66. Della-Maria J, Zhou Y, Tsai M-S, Kuhnlein J, Carney JP, Paull TT, et al. Human Mre11/human Rad50/Nbs1 and DNA ligase IIIalpha/XRCC1 protein complexes act together in an alternative nonhomologous end joining pathway. J Biol Chem 2011;286:33845–53.
67. Lemée F, Bergoglio V, Fernandez-Vidal A, Machado-Silva A, Pillaire M-J, Bieth A, et al. DNA polymerase theta up-regulation is associated with poor survival in

Downloaded from cancerres.aacrjournals.org on December 29, 2020. © 2020 American Association for Cancer Research.

Published OnlineFirst July 10, 2020; DOI: 10.1158/0008-5472.CAN-20-1672

breast cancer, perturbs DNA replication, and promotes genetic instability. Proc Natl Acad Sci U S A 2010;107:13390–5.
68. Mateos-Gomez PA, Gong F, Nair N, Miller KM, Lazzerini-Denchi E, Sfeir A. Mammalian polymerase θ promotes alternative NHEJ and suppresses recombination. Nature 2015;518:254–7.
69. Kennedy RD, Gorski JJ, Quinn JE, Stewart GE, James CR, Moore S, et al. BRCA1 and c-Myc associate to transcriptionally repress psoriasin, a DNA damage–inducible gene. Cancer Res 2005;65:10265–72.
70. Nagaria P, Robert C, Rassool F V. DNA double-strand break response in stem cells: mechanisms to maintain genomic integrity. Biochim Biophys Acta 2013; 1830:2345–53.
71. Shima N, Munroe RJ, Schimenti JC. The mouse genomic instability mutation chaos1 is an allele of Polq that exhibits genetic interaction with Atm. Mol Cell Biol 2004;24:10381–9.
72. Koole W, van Schendel R, Karambelas AE, van Heteren JT, Okihara KL, Tijsterman M. A polymerase theta-dependent repair pathway suppresses extensive genomic instability at endogenous G4 DNA sites. Nat Commun 2014;5:3216.
73. Lemmens B, van Schendel R, Tijsterman M. Mutagenic consequences of a single G-quadruplex demonstrate mitotic inheritance of DNA replication fork barriers. Nat Commun 2015;6:8909.
74. Zyner KG, Mulhearn DS, Adhikari S, Martínez Cuesta S, Di Antonio M, Erard N, et al. Genetic interactions of G-quadruplexes in humans. Elife 2019;8: e46793.
75. Bryan TM. Mechanisms of DNA replication and repair: insights from the study of G-quadruplexes. Molecules 2019;24:3439.
76. Soulier J, Leblanc T, Larghero J, Dastot H, Shimamura A, Guardiola P, et al. Detection of somatic mosaicism and classification of fanconi anemia patients by analysis of the FA/BRCA pathway. Blood 2005;105:1329–36.
77. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011; 144:646–74.
78. Boboila C, Oksenych V, Gostissa M, Wang JH, Zha S, Zhang Y, et al. Robust chromosomal DNA repair via alternative end-joining in the absence of X-ray repair cross-complementing protein 1 (XRCC1). Proc Natl Acad Sci U S A 2012;109:2473–8.
79. Wood RD, Doublié S. DNA polymerase θ (POLQ), double-strand break repair, and cancer. DNA Repair 2016;44:22–32.
80. Schimmel J, van Schendel R, den Dunnen JT, Tijsterman M. Templated insertions: a smoking gun for polymerase theta-mediated end joining. Trends Genet 2019;35:632–44.
81. Higgins GS, Harris AL, Prevo R, Helleday T, McKenna WG, Buffa FM. Overexpression of POLQ confers a poor prognosis in early breast cancer patients. Oncotarget 2010;1:175–84.
82. Chen X, Zhong S, Zhu X, Dziegielewska B, Ellenberger T, Wilson GM, et al. Rational design of human DNA ligase inhibitors that target cellular DNA replication and repair. Cancer Res 2008;68:3169–77.
83. Carvajal-Garcia J, Cho J-E, Carvajal-Garcia P, Feng W, Wood RD, Sekelsky J, et al. Mechanistic basis for microhomology identification and genome scarring by polymerase theta. Proc Natl Acad Sci U S A 2020;117:8476–85.
84. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature 2013; 500:415–21.
85. Feng W, Simpson DA, Carvajal-Garcia J, Price BA, Kumar RJ, Mose LE, et al. Genetic determinants of cellular addiction to DNA polymerase theta. Nat Commun 2019;10:4286.
86. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire of mutational signatures in human cancer. Nature 2020;578:94–101.
87. Kass EM, Moynahan ME, Jasin M. When genome maintenance goes badly awry. Mol Cell 2016;62:777–87.
88. Bryant HE, Schultz N, Thomas HD, Parker KM, Flower D, Lopez E, et al. Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. Nature 2005;434:913–7.
89. Farmer H, McCabe N, Lord CJ, Tutt ANJ, Johnson DA, Richardson TB, et al. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. Nature 2005;434:917–21.
90. Sishc BJ, Davis AJ. The role of the core non-homologous end joining factors in carcinogenesis and cancer. Cancers 2017;9:81.
91. Kumar RJ, Chao HX, Roberts VR, Sullivan AR, Shah SJ, Simpson DA, et al. Hyperactive end joining repair mediates resistance to DNA damaging therapy in p53-deficient cells. bioRxiv 2020. DOI: https://doi.org/10.1101/2020.04.01.021253.
92. Higgins GS, Boulton SJ. Beyond PARP-POLθ as an anticancer target. Science 2018;359:1217–8.
93. Zhou J, Gelot C, Pantelidou C, Li A, Yücel H, Davis RE, et al. Polymerase theta inhibition kills homologous recombination deficient tumors. bioRxiv 2020. DOI: https://doi.org/10.1101/2020.05.23.111658.
94. Exell JC, Thompson MJ, Finger LD, Shaw SJ, Debreczeni J, Ward TA, et al. Cellularly active N-hydroxyurea FEN1 inhibitors block substrate entry to the active site. Nat Chem Biol 2016;12:815–21.
95. Uematsu N, Weterings E, Yano K, Morotomi-Yano K, Jakob B, Taucher-Scholz G, et al. Autophosphorylation of DNA-PKCS regulates its dynamics at DNA double-strand breaks. J Cell Biol 2007;177:219–29.
96. Blackford AN, Jackson SP. ATM, ATR, and DNA-PK: The trinity at the heart of the DNA damage response. Mol Cell 2017;66:801–17.
97. Curtin NJ. DNA repair dysregulation from cancer driver to therapeutic target. Nat Rev Cancer 2012;12:801–17.
98. Gaillard H, García-Muse T, Aguilera A. Replication stress and cancer. Nat Rev Cancer 2015;15:276–89.
99. Dobbelstein M, Sørensen CS. Exploiting replicative stress to treat cancer. Nat Rev Drug Discov 2015;14:405–23.
100. Wang Z, Song Y, Li S, Kurian S, Xiang R, Chiba T, et al. DNA polymerase θ (POLQ) is important for repair of DNA double-strand breaks caused by fork collapse. J Biol Chem 2019;294:3909–19.
101. Konstantinopoulos PA, Cheng S-C, Hendrickson AEW, Penson RT, Schumer ST, Doyle LA, et al. Randomized phase 2 study of the ATR inhibitor berzosertib in combination with gemcitabine versus gemcitabine alone in platinum resistant high grade serous ovarian cancer. Lancet Oncol 2020;21:957–68.
102. Lord CJ, Ashworth A. Mechanisms of resistance to therapies targeting BRCA-mutant cancers. Nat Med 2013;19:1381–8.
103. Nigro JM, Baker SJ, Preisinger AC, Jessup JM, Hostetter R, Cleary K, et al. Mutations in the p53 gene occur in diverse human tumour types. Nature 1989; 342:705–8.
104. Lee JM, Bernstein A. p53 mutations increase resistance to ionizing radiation. Proc Natl Acad Sci U S A 1993;90:5742–6.
105. McIlwrath AJ, Vasey PA, Ross GM, Brown R. Cell cycle arrests and radiosensitivity of human tumor cell lines: dependence on wild-type p53 for radiosensitivity. Cancer Res 1994;54:3718–22.
106. Goff JP, Shields DS, Seki M, Choi S, Epperly MW, Dixon T, et al. Lack of DNA polymerase theta (POLQ) radiosensitizes bone marrow stromal cells *in vitro* and increases reticulocyte micronuclei after total-body irradiation. Radiat Res 2009;172:165–74.
107. Hill SJ, Decker B, Roberts EA, Horowitz NS, Muto MG, Worley MJ, et al. Prediction of DNA repair inhibitor response in short-term patient-derived ovarian cancer organoids. Cancer Discov 2018;8:1404–21.

Downloaded from cancerres.aacrjournals.org on December 29, 2020. © 2020 American Association for Cancer Research.

Published OnlineFirst July 10, 2020; DOI: 10.1158/0008-5472.CAN-20-1672

Cancer Research
The Journal of Cancer Research (1916–1930) | The American Journal of Cancer (1931–1940)

AACR American Association for Cancer Research

Exploiting the Microhomology-Mediated End-Joining Pathway in Cancer Therapy

Jeffrey Patterson-Fortin and Alan D. D'Andrea

Cancer Res 2020;80:4593-4600. Published OnlineFirst July 10, 2020.

Updated version Access the most recent version of this article at:
doi:10.1158/0008-5472.CAN-20-1672

Cited articles This article cites 105 articles, 26 of which you can access for free at:
http://cancerres.aacrjournals.org/content/80/21/4593.full#ref-list-1

E-mail alerts Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions To request permission to re-use all or part of this article, use this link
http://cancerres.aacrjournals.org/content/80/21/4593.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.

Downloaded from cancerres.aacrjournals.org on December 29, 2020. © 2020 American Association for Cancer Research.



Cell Cycle

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/kccy20

Disassembly of the Shieldin Complex by TRIP13

Prabha Sarangi , Connor S. Clairmont & Alan D. D'Andrea

To cite this article: Prabha Sarangi , Connor S. Clairmont & Alan D. D'Andrea (2020) Disassembly of the Shieldin Complex by TRIP13, Cell Cycle, 19:13, 1565-1575, DOI: 10.1080/15384101.2020.1758435

To link to this article: https://doi.org/10.1080/15384101.2020.1758435

© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

Published online: 18 May 2020.

Submit your article to this journal

Article views: 1034

View related articles

View Crossmark data

Citing articles: 1 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=kccy20

CELL CYCLE
2020, VOL. 19, NO. 13, 1565–1575
https://doi.org/10.1080/15384101.2020.1758435

Taylor & Francis
Taylor & Francis Group

REVIEW

OPEN ACCESS Check for updates

Disassembly of the Shieldin Complex by TRIP13

Prabha Sarangi[a,b], Connor S. Clairmont[a,b], and Alan D. D'Andrea[a,b]

[a]Department of Radiation Oncology, Dana-Farber Cancer Institute, Boston, MA, USA; [b]Center for DNA Damage and Repair, Dana-Farber Cancer Institute, Boston, MA, USA

ABSTRACT

In the past decade, the study of the major DNA double strand break (DSB) repair pathways, homologous recombination (HR) and classical non-homologous end joining (C-NHEJ), has revealed a vast and intricate network of regulation. The choice between HR and C-NHEJ is largely controlled at the step of DNA end-resection. A pro-C-NHEJ cascade commencing with 53BP1 and culminating in the newly discovered REV7-Shieldin complex impedes end resection and therefore HR. Importantly, loss of any component of this pathway confers PARP inhibitor resistance in *BRCA1*-deficient cells; hence, their study is of great clinical importance. The newest entrant on the scene of end resection regulation is the ATPase TRIP13 that disables the pro-C-NHEJ cascade by promoting a novel conformational change of the HORMA protein REV7. Here, we tie these new findings and factors with previous research on the regulation of DSB repair and HORMA proteins, and suggest testable hypotheses for how TRIP13 could specifically inactivate REV7-Shieldin to promote HR. We also discuss these biological questions in the context of clinical therapeutics.

ARTICLE HISTORY
Received 25 March 2020
Revised Vxx xxx xxxx
Accepted 8 April 2020

KEYWORDS
Keywords REV7; Shieldin; TRIP13DNA repair; homologous recombination; PARP inhibitor; HORMA protein; fanconi anemia

Many human cancers, such as ovarian, breast, prostate, and pancreatic cancers, have mutations in homologous recombination (HR) pathway genes. The most commonly altered HR genes are *BRCA1* and *BRCA2*, followed by other genes in the Fanconi Anemia (FA) pathway. Loss of HR causes genomic instability, hyper-dependence on alternative DNA repair mechanisms, and enhanced sensitivity to PARP inhibitors (PARPi) [1,2]. This synthetic lethal interaction with PARPi has been exploited therapeutically, most notably in ovarian cancer, where several PARPis have recently received FDA approval. PARPis are oral drugs and relatively well-tolerated, and they can be combined with other anti-cancer therapies. Unfortunately, PARPi resistance has emerged as a vexing clinical problem for the treatment of *BRCA1/2*-deficient carcinomas.

The most prevalent mechanism of PARPi resistance is secondary events, such as somatic reversion of a germline *BRCA1/2* mutation, leading to restoration of HR repair [3,4]. Another important mechanism of HR restoration in *BRCA1*-deficient cells involves the loss of factors which normally inhibit 5′-3′ DNA end resection at a double strand break (DSB), an important first step in HR (Figure 1a) [5]. The first gene identified in this pathway was 53BP1, whose loss was found to restore viability in *BRCA1*$^{-/-}$ mice and render *BRCA1*-deficient cancers resistant to DNA damaging treatment, including PARPi [6,7]. Subsequent work showed that upon DNA damage, 53BP1 is phosphorylated by the DNA damage responsive kinase ATM at numerous sites which recruit the downstream effector proteins, RIF1 and PTIP, to promote classical non-homologous end joining (C-NHEJ), a DSB repair pathway that competes with and is more error-prone than HR [8–11]. Loss of RIF1 and PTIP also induces PARPi resistance in *BRCA1/2*-deficient backgrounds, similarly to loss of 53BP1. In a surprising discovery, the small DNA polymerase ζ subunit, REV7, was uncovered as an essential downstream effector of RIF1, independently of any known Pol ζ-related binding partners [12,13].

The Polymerase ζ complex, consisting of REV7 and REV3, has long been known to catalyze mutagenic translesion DNA synthesis (TLS), a pathway that allows cells to replicate DNA across damaged bases [14]. All of the catalytic activity of the complex

CONTACT Alan D. D'Andrea Alan_Dandrea@dfci.harvard.edu

© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

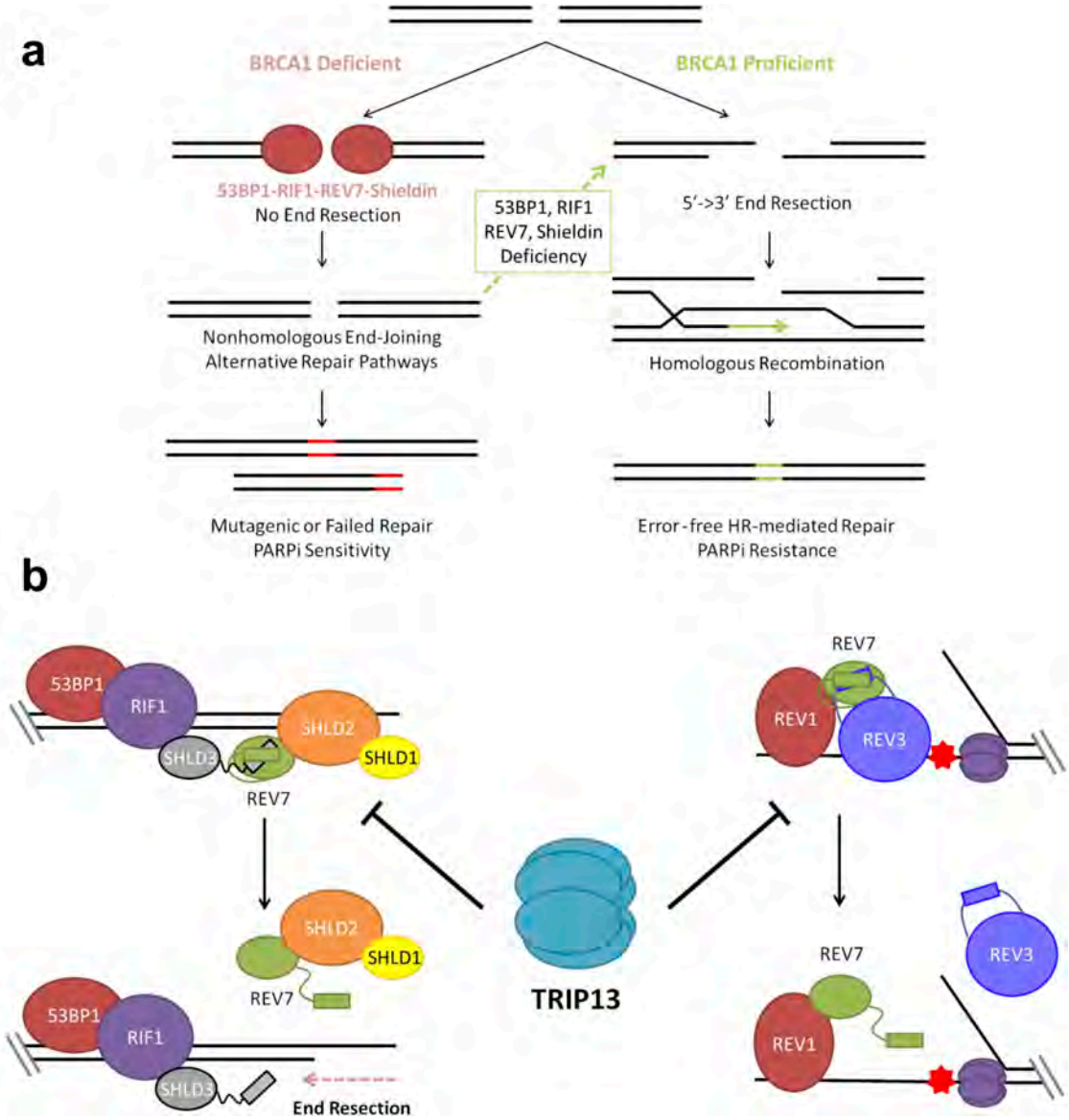


Figure 1. TRIP13 promotes PARP inhibitor sensitivity by disassembling REV7-Shieldin. (a) BRCA1 promotes 5′-3′ DNA end resection, the critical first step in double strand break (DSB) repair through homologous recombination. In the absence of BRCA1; 53BP1, RIF1, REV7 and Shieldin inhibit DNA end resection, preventing repair by HR. End resection and HR can be restored in *BRCA1*-deficient cells by concurrent deficiency of any component of this anti-resection axis. Without HR repair, cells are unable to properly repair all DSBs, particularly those that arise during S phase leading to hypersensitivity to PARP inhibition and other DNA damaging therapies. (b) The Shieldin complex assembles at the break site through sequential recruitment of 53BP1, RIF1, SHLD3, REV7 and SHLD1-SHLD2 (left). REV7 is also recruited to sites of replication blockages, where it interacts with REV1 and REV3 as part of the DNA Polymerase ζ complex (right). TRIP13 catalyzes an inactivating conformational change of REV7, thereby inhibiting both the Shieldin complex and the Polymerase ζ complex.

is contained within the large REV3 subunit, and REV7 has primarily been considered an accessory subunit that may function only to bridge the interaction between Pol ζ and the TLS factor, REV1 [15].

Given the known function of REV7 as an accessory subunit, these studies strongly implied the existence of novel REV7 interaction partners mediating its newly found role in HR regulation. Indeed, a flurry of studies using complementary techniques of CRISPR screening and physical interactome analysis identified three novel interaction partners – SHLD1, SHLD2 and SHLD3 [16–23]. Together with REV7, these four proteins form a complex known as Shieldin. The entire cascade,

starting with 53BP1 and culminating in the Shieldin complex, inhibits end resection-dependent HR and promotes C-NHEJ, such that loss of any of them in the absence of BRCA1 relieves the inhibition and allows HR to occur. This is clinically relevant because these doubly-deficient cells are resistant to PARPis.

Our laboratory has recently demonstrated that tumor cells can also achieve a state of PARPi resistance by upregulating the expression of the ATPase, TRIP13, and that this mode is linked to REV7 [24]. Importantly, *BRCA1*-deficient tumor cells often have high TRIP13 expression, leading to PARPi resistance. We showed that REV7 exists in active "closed" and inactive "open" conformations (Figure 1b). TRIP13, through ATP hydrolysis, catalyzes an inactivating conformational change in REV7. In doing so, it dissociates the REV7-Shieldin complex, promoting end resection and thereby HR. TRIP13 similarly disassembles the REV7-REV3 TLS complex. Thus, TRIP13 overexpression simultaneously inactivates mutagenic TLS and C-NHEJ and promotes high fidelity HR. Importantly, since TRIP13 is an enzyme, an inhibitor of TRIP13 may be useful in re-sensitizing *BRCA1*-deficient tumor cells to PARPis. Despite some of the more obvious biological and clinical implications of TRIP13 overexpression, such as increased HR and the acquisition of PARPi resistance, there are several unanswered questions regarding its regulation and mechanism of action. This review addresses emerging questions in the field of regulation of end resection, with a brief background on the better known apical players and a detailed look at the newest and most downstream components, REV7 and TRIP13.

Regulation of the 53BP1-RIF1-REV7-Shieldin anti-resection pathway

As end resection is a critical determinant of DSB repair pathway and outcome, it is not surprising that nearly every participating factor is regulated. 53BP1, the apical component of the anti-resection pathway, has ubiquitin binding [25] and methyl-lysine binding domains [26,27], both of which aid in its recruitment. The DNA damage specificity of 53BP1 recruitment is thought to be mediated by its binding to H2AK15 ubiquitylation, catalyzed by the DNA damage responsive ubiquitin ligase, RNF168 [28]. RNF168 recruitment is in turn regulated by a DNA damage cascade comprising ATM, MDC1 and RNF8 [29–33]. The ATM kinase further phosphorylates 53BP1 itself, which is required for the recruitment of RIF1, REV7 and Shieldin.

The methyl-lysine binding tudor domains bind specifically to methylated histone H4K20, a ubiquitous histone mark [34]. This interaction is normally inhibited by the protein TIRR, which binds directly to the tudor domain, but is released upon DNA damage [35,36]. It has been suggested that H4K20 methylation accumulates slowly following histone deposition on newly replicated DNA, thus favoring 53BP1 binding to unreplicated DNA [37,38]. This is a sensible model, as HR requires the presence of a sister chromatid, and should therefore be repressed in unreplicated DNA. 53BP1 is also inhibited by phosphorylation during M-phase [39–41], when it can cause aberrant DNA repair. It is unclear whether this aberrancy is related to unscheduled resection, which can occur to some degree during M-phase [42].

Only recently have we learned of the downstream REV7-Shieldin components, named so due to their functional and sequence level homology to the telomeric Shelterin complex. Like Shelterin, the Shieldin complex localizes to DNA ends and binds to single stranded DNA (ssDNA), preventing resection, accumulation of the ssDNA binding protein RPA and subsequent ATR signaling. However, unlike Shelterin, which exhibits sequence-specific DNA binding, Shieldin is recruited to DNA damage through 53BP1-RIF1 interaction, with SHLD3 acting as a bridge between RIF1 and REV7, which in turn recruits SHLD1 and SHLD2. Nothing is known about the regulation of these downstream proteins other than the role of TRIP13 and REV7 conformation, the function and regulation of which still present many mysteries.

The HORMA family of proteins

REV7 is a founding member of the HORMA protein family, so-named for **HO**p1, **REV7** and **MAD2** [43]. In addition to the founding members, recent studies have identified the autophagy proteins Atg13 and Atg101 as HORMA family

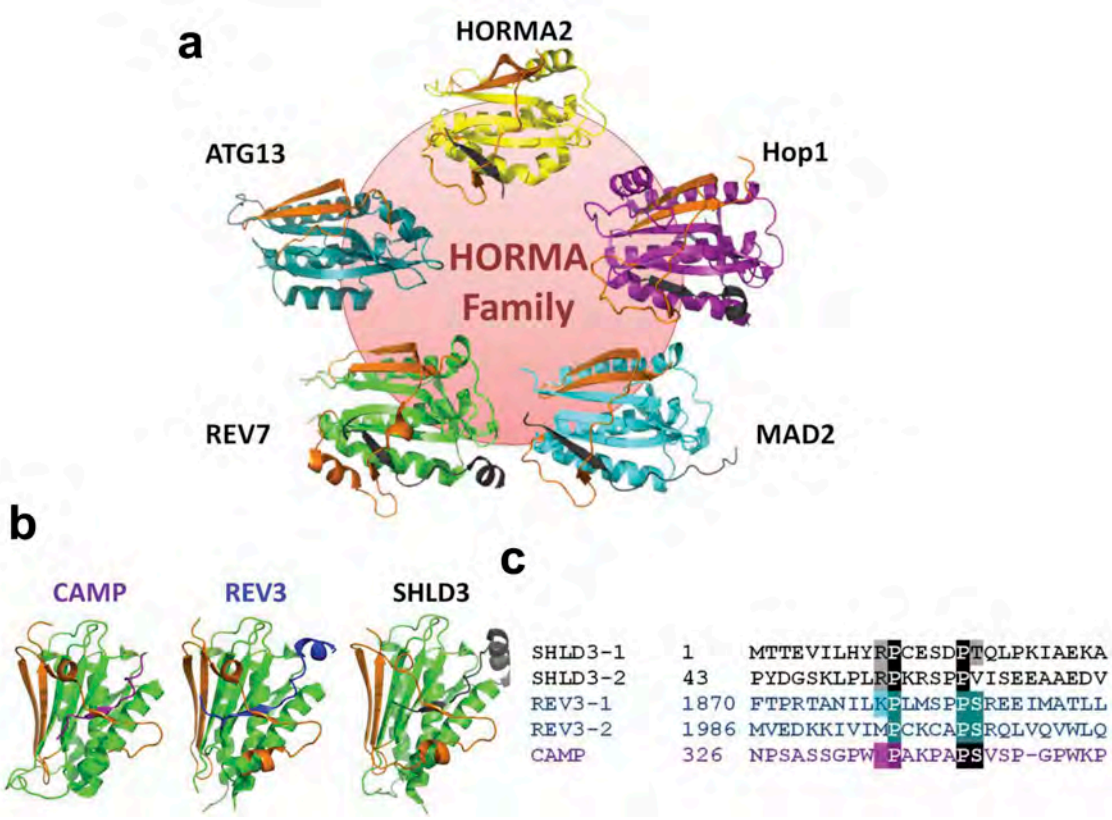


Figure 2. REV7 is a member of the functionally diverse HORMA family and interacts with multiple seatbelt binding partners. (a) The HORMA family is highly structurally conserved across domains of life. Shown are examples of HORMA protein structures in the closed conformation with the seatbelt subdomain highlighted in orange; clockwise from top: HORMA2 – *P. aeruginosa*, PDB: 6P8 U [47]; HIM-3 (Hop1 orthologue) – *C. elegans*, PDB: 4TZJ [60]; MAD2 – *S. pombe*, PDB: 4AEZ [86]; REV7 – *H. sapiens*, PDB: 3VU7 [87], ATG13 – *H. sapiens*, PDB: 5C50 [55]. (b) Structures of closed REV7 bound to seatbelt binding partners: CAMP, REV3 and SHLD3. PDB: 5XPT [62], 3VU7 [87], 6K07 [88] respectively. (c) Primary sequence alignment of REV7 seatbelt binding regions of SHLD3, REV3 and CAMP highlighting conserved amino acids constituting the REV7 seatbelt binding motif (SBM).

members [44–46]. Intriguingly, HORMA proteins and a TRIP13 orthologue have recently been identified in bacteria, suggesting that this family has ancient origins [47]. Proteins of the HORMA family are unified by modest primary sequence homology, but striking homology in tertiary structure (Figure 2a).

Like REV7, Hop1, MAD2 and the bacterial HORMA proteins activity are also regulated by stable conformational changes between open and closed conformations [47–49]. The major structural transition between the open and closed conformation occurs in the C-terminal "seatbelt" subdomain. In the closed conformation, the seatbelt latches tightly to binding partners at so-called seatbelt binding motifs (SBMs), whereas this interaction is not stable in the open conformation. It is this critical binding to SBM-containing partners that is controlled by HORMA conformational transitions. Intriguingly, Atg13 and Atg101 seem to have non-canonical HORMA function as they appear to be locked in the closed and open states respectively and form a heterodimer [46].

While the HORMA proteins share structural and regulatory features, they function in very distinct capacities. MAD2 plays a clearly defined role in the spindle assembly checkpoint (SAC). During mitosis, pools of inactive, open MAD2 are activated at unattached kinetochores through seatbelt-SBM interactions with MAD1 [50–52]. MAD1 promotes the closing of MAD2 which is propagated in a chain reaction by dimerization of open and closed MAD2 conformers [53,54]. It is noteworthy that the Atg13-Atg101 conformationally-locked heterodimer, resembles the assymetric closed-open MAD2 dimer [55], suggesting that asymmetric HORMA protein dimers may be a common feature of regulation. The target of

MAD2 is the anaphase promoting complex (APC) activator, Cdc20, another SBM-containing protein [56,57]. Closed MAD2 rapidly binds and sequesters Cdc20, preventing APC activation until each chromatid is attached to the spindle.

Hop1 is a conserved member of the meiosis-specific chromosome axis complex, which organizes homologous chromosomes and promotes crossover formation [58,59]. Unlike MAD2, which forms discrete soluble MAD2:Cdc20 complexes, Hop1 and the related mammalian proteins, HORMAD1 and HORMAD2, appear to form large chromosome-bound chains. Hop1-type proteins have both a HORMA domain and SBMs, allowing for the formation of complex higher-order structures with one Hop1 monomer's seatbelt bound over the adjacent Hop1 monomer's SBM and so on [60].

REV7 is unique among the HORMA proteins in having multiple distinct SBM-containing partners (Figure 2b,c). SHLD3 has two tandem SBMs in its N-terminus, similar to REV3 which has two functionally important SBMs [61]. It is not known if REV7 can bind to both SBMs simultaneously or if this is functionally relevant. In addition to the well-characterized interaction with REV3 and to a lesser extent SHLD3, the zinc-finger protein CAMP is also a confirmed seatbelt interactor [62]. High-throughput interaction profiling of REV7 has revealed numerous additional binding partners, but whether any of these are also seatbelt-mediated remains to be elucidated [17,24]. Why REV7 is involved in these distinct processes is not understood, but is suggestive of some sort of coordinated regulation of REV7-dependent pathways. MAD2 also has two seatbelt-mediated interaction partners, namely MAD1 and Cdc20; however there is a clear distinction, as MAD1 stimulates MAD2 closing, while Cdc20 can only be bound following MAD1 activity, and both these partners are related to the same (and only) function of MAD2. Whether the known REV7 binding partners can play activator functions, analogous to MAD1, or are mainly targets of REV7 activity, like Cdc20, is not yet clear. It is also not known whether REV7 activation is similarly propagated through asymmetric dimerization, like with MAD2. REV7 homodimerization through a similar interface has recently been shown to be functionally important [63], lending support to this hypothesis; however, more extensive research is still required.

One unifying feature of the HORMA family of proteins is their regulation by the AAA+ ATPase TRIP13. Like REV7, Hop1 and MAD2 are both known to be inactivated by TRIP13 which is thought to remodel closed conformers into the inactive open conformation [64–66]. The involvement of TRIP13 in at least three very distinct cellular processes raises several interesting questions. Is TRIP13 activity somehow directed to specific substrates in the appropriate context, such as through substrate-specific adaptor proteins? Or is the regulation of these seemingly distinct proteins actually coordinated, perhaps at specific cell cycle phases? For instance, TRIP13 may regulate the choice between HR and NHEJ during S phase, and the activity of the SAC in mitosis.

Regulation of TRIP13 activity on HORMA proteins

TRIP13 is a highly conserved AAA+ ATPase that forms single homohexameric rings. It was first linked to the HORMA protein family when studies in yeast revealed a role in meiotic checkpoint and recombination [67–69]. This role is evolutionarily conserved as both yeast Pch2 and mouse Trip13 regulate their respective HORMA substrates, Hop1 and Hormad1 [64,68]. In both organisms, Hop1/Hormad1 monomers bind to DNA in a cooperative manner, and addition of Pch2/Trip13 alters their DNA binding or chromatin association [64,70]. Thus, Pch2/Trip13 has been proposed to remodel Hop1/Hormad1 oligomers already on chromatin, promoting their displacement from DNA and turnover. How Pch2 is targeted to Hop1 is not entirely clear, but it has been shown to interact with the BRCT domains in Xrs2, part of the MRX (Mre11-Rad50-Xrs2) damage sensor in yeast (MRN or MRE11-RAD50-NBS1 in humans), and this interaction is functionally important [68]. Mouse *Hormad1* mutants lacking the N terminal region are refractory to TRIP13 action, indicating that this region is required for the disassembly of HORMAD1 from chromatin.

How TRIP13 is targeted to and inactivates MAD2 is better understood. Like in the case of

Hormad1, the highly conserved N-terminal region of MAD2 is necessary for TRIP13-mediated inactivation [71]. Elegant structural studies have provided insight into why this is so. TRIP13 grasps the N terminus of MAD2, which is fed through a central pore. ATP hydrolysis by TRIP13 causes a limited unfolding of this region of MAD2, which destabilizes the overall closed MAD2 structure by disrupting a conserved hydrogen bond network between residues at the N-terminus and C-terminal seatbelt domain [71]. Interestingly, while TRIP13 can elicit this effect alone, its activity on MAD2 is greatly enhanced by another HORMA-like protein, p31, which directly interacts with both TRIP13 and MAD2, forming a bridge between them [65,72] (Figure 3). p31 binds to MAD2 at the latter's homodimerization interface, and this structural mimicry further attenuates MAD2 signaling by preventing activation through the asymmetric closed-open dimerization pathway [73]. Whether p31 similarly coordinates the interaction of TRIP13 and REV7 is unknown.

Models for TRIP13 activity toward REV7

Given the very recent discovery of TRIP13 action on REV7, we know little about the mechanism and regulation of this interaction; however, using evidence from other HORMA proteins, we can suggest several testable hypotheses. One critical question is whether there is any mechanism for TRIP13 substrate specificity or if all substrates present in the cell are regulated concurrently. A possible mechanism of substrate specificity is through adaptor proteins like p31, which has only been definitively shown to interact with MAD2. If there are alternative p31-like adapters for REV7 and HORMAD1, this could dictate substrate specificity (Figure 4, Model 1). There is some evidence suggesting that REV7 does in fact interact with p31, contradicting this model; however, more research is required on this topic [63]. The potential universality of the p31 adaptor is also supported by reports that the rice orthologue, CRC1, localizes to the chromosome axis and is implicated in meiotic HR [74],

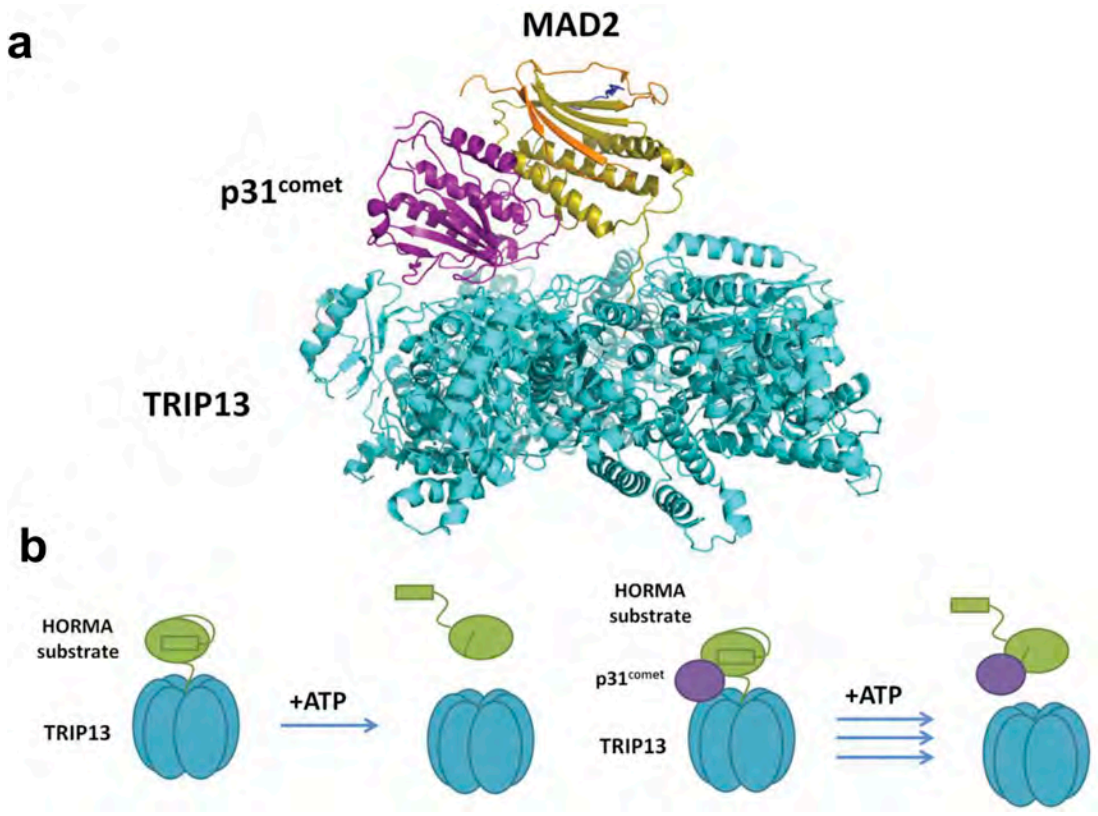


Figure 3. TRIP13 remodels HORMA substrates through N-terminal engagement and is stimulated by an adaptor protein. (a) Cryo-EM structure of TRIP13 acting on C-MAD2, showing bridging by the adaptor protein, $p31^{comet}$, PDB: 6F0X [89]. (b) Schematic of N-terminal engagement of HORMA substrate by TRIP13 and stimulation by an adaptor, $p31^{comet}$, which bridges the interaction. Following engagement of the HORMA protein, TRIP13 catalyzes HORMA remodeling coupled to ATP hydrolysis.

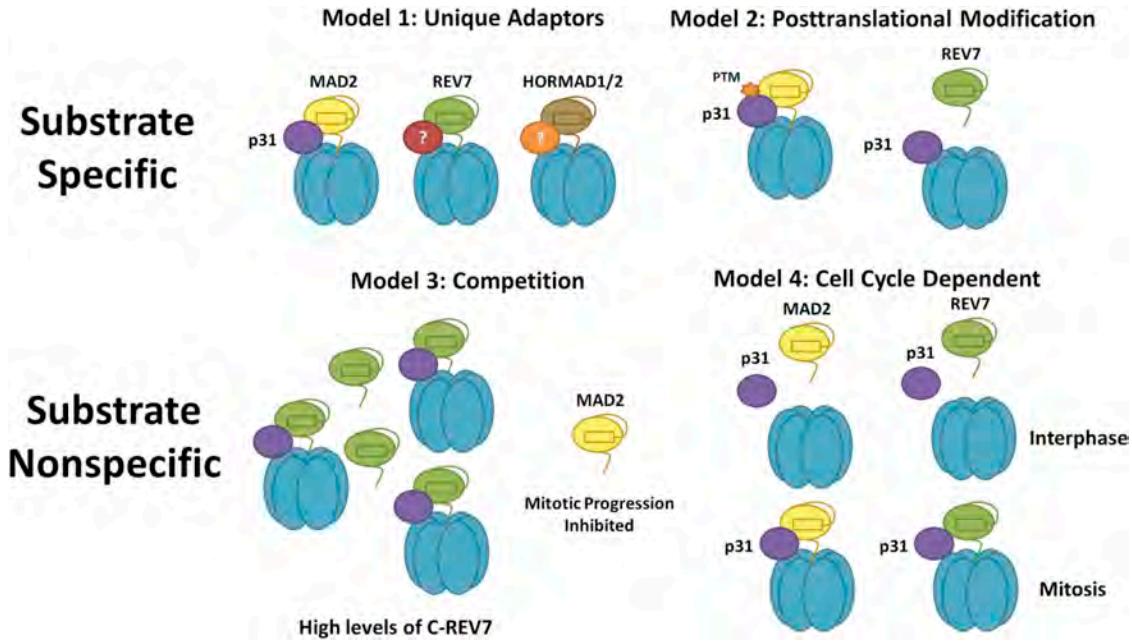


Figure 4. Models for TRIP13 regulation of diverse HORMA proteins. Substrate Specific Models (1): Substrate specificity is enforced by unique adaptors for each HORMA protein; (2): Substrate specificity is enforced by post-translational modification of the appropriate substrate. Substrate Nonspecific Models (3): Each HORMA protein competes for a limited supply of TRIP13-p31. High levels of DNA damage would thereby inhibit anaphase progression through TRIP13-p31 sequestration. C-REV7 indicates closed REV7; (4): TRIP13 acts on all substrates during a specific cell cycle phase or phases. TRIP13 is known to be active in mitosis and may represent another mechanism to suppress DNA repair in mitosis.

suggestive of a HORMAD1 regulatory role. On the other hand, there are some organisms, such as budding yeast, that possess TRIP13 but not p31, and in these cases, TRIP13 appears to have only a meiotic function through Hop1 [75]. If p31 is in fact a universal adaptor, substrate specificity may instead be dictated through posttranslational modifications of TRIP13, p31 or the HORMA substrate (Figure 4, Model 2). Indeed, Plk1-dependent phosphorylation of p31 decreases its affinity for MAD2, thereby dialing down TRIP13-mediated inactivation of MAD2 [76]. Perhaps DNA damage-dependent phosphorylation of p31 or REV7-Shieldin by ATM similarly modulates TRIP13 activity to promote certain types of repair.

It is also possible that TRIP13 does not show substrate specificity; however, this does not mean that TRIP13 activity is not regulated. TRIP13 activity could be controlled by the relative abundance of the closed conformations of each substrate, such that they compete for limited TRIP13-p31 (Figure 4, Model 3). This would have interesting consequences as high levels of DNA damage, and therefore closed REV7, would effectively sequester TRIP13, preventing inactivation of the SAC. This would represent a novel mechanism of cell cycle arrest in response to DNA damage.

Yet another possibility is that TRIP13 and p31 are sufficiently abundant to prevent any competition. In this case, they would regulate all of their HORMA substrates concurrently. It is noteworthy that p31 protein expression is up-regulated specifically in S and G2/M phases, when REV7 and MAD2 activities need to be regulated. Thus, cells could specifically promote the efficient inactivation of REV7 and MAD2 by limiting p31 expression to these cell cycle phases (Figure 4, Model 4). Apart from these possibilities but perhaps related to them, TRIP13 cellular localization could dictate its activity. Support for this idea comes from studies showing that TRIP13 is normally cytoplasmic, but forms foci at kinetochores during mitosis [77] and on chromosome axes during meiosis [78]. Whether TRIP13 similarly forms foci after DSB induction that co-localize with DNA damage factors would be useful to pinpoint when and where TRIP13 inactivates REV7.

Clinical Implications of the TRIP13-REV7 Interaction

TRIP13 activity is not only biologically interesting, it is also clinically relevant. Previous studies have noted that TRIP13 is overexpressed in various cancers and is a negative prognostic marker [79–81]. One report examining TRIP13 in the context of head and neck cancer similarly found that TRIP13 promotes DSB repair and chemoresistance, but concluded that TRIP13 promotes C-NHEJ, not HR [82]. That study identified a physical interaction between TRIP13 and DNA-PK subunits; however, these are common contaminants that appear in nearly 50% of AP-MS experiments [83]. Additionally, there are no HORMA domains in DNA-PK and no mechanism was proposed for TRIP13 regulation of DNA-PK. It remains possible that TRIP13 affects C-NHEJ in the cancer cell lines used in that study; however, our results demonstrate that the major function of TRIP13 in DSB repair is the promotion of HR through REV7 inactivation. Interestingly, biallelic loss-of-function TRIP13 mutations have been seen in Wilms tumor cells, which not surprisingly have an SAC defect [84]. It would be clinically useful to establish whether they are additionally HR-deficient, and therefore sensitive to PARPis.

Alternative Roles of TRIP13 in Regulating Genomic Stability

TRIP13 overexpression has previously been proposed to engender genomic instability by weakening the SAC, allowing cells to replicate more quickly and become aneuploid, promoting oncogenesis [79]. This idea has been challenged by the finding that TRIP13 expression is highly correlated with that of MAD2, suggesting a homeostatic relationship between the two in maintaining a normal SAC [85]. Given the genomic proximity of TRIP13 to hTERT, a strong oncogene, TRIP13 amplification may often occur as a passenger mutation. Regardless of the mechanism of TRIP13 overexpression in cancer, our studies have shown that the consequences are clear: high TRIP13 levels enhance HR and engender chemoresistance.

While elucidating PARP inhibitor resistance mechanisms through the discovery of 53BP1, RIF1, and the Shieldin complex has been a significant advance, none of those factors are druggable. TRIP13, as the newest entrant to this field and an enzymatic, negative regulator of the Shieldin complex, emerges as a critical player whose inhibition could be used in the clinic to extend the utility of PARP inhibitors.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Foundation for the National Institutes of Health [R37HL052725, P01HL048546]; Leukemia and Lymphoma Society [5440-16]; DOD (Department of Defense) [BC151331P1].

References

[1] Bryant HE, Schultz N, Thomas HD, *et al.* Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. Nature. 2005;434(7035):913–917.
[2] Farmer H, McCabe N, Lord CJ, *et al.* Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. Nature. 2005;434(7035):917–921.
[3] Edwards SL, Brough R, Lord CJ, *et al.* Resistance to therapy caused by intragenic deletion in BRCA2. Nature. 2008;451(7182):1111–1115.
[4] Sakai W, Swisher EM, Karlan BY, *et al.* Secondary mutations as a mechanism of cisplatin resistance in BRCA2-mutated cancers. Nature. 2008;451(7182):1116–1120.
[5] Heyer WD, Ehmsen KT, Liu J. Regulation of homologous recombination in eukaryotes. Annu Rev Genet. 2010;44(1):113–139.
[6] Bunting SF, Callén E, Wong N, et al. 53BP1 inhibits homologous recombination in Brca1-deficient cells by blocking resection of DNA breaks. Cell. 2010;141(2):243–254.
[7] Bouwman P, Aly A, Escandell JM, *et al.* 53BP1 loss rescues BRCA1 deficiency and is associated with triple-negative and BRCA-mutated breast cancers. Nat Struct Mol Biol. 2010;17(6):688–695.
[8] Zimmermann M, Lottersberger F, Buonomo SB, et al. 53BP1 regulates DSB repair using Rif1 to control 5′ end resection. Science. 2013;339(6120):700–704.

[9] Escribano-Diaz C, Orthwein A, Fradet-Turcotte A, *et al.* A cell cycle-dependent regulatory circuit composed of 53BP1-RIF1 and BRCA1-CtIP controls DNA repair pathway choice. Mol Cell. 2013;49(5):872–883.
[10] Chapman JR, Barral P, Vannier J-B, *et al.* RIF1 is essential for 53BP1-dependent nonhomologous end joining and suppression of DNA double-strand break resection. Mol Cell. 2013;49(5):858–871.
[11] Callen E, Di Virgilio M, Kruhlak M, *et al.* 53BP1 mediates productive and mutagenic DNA repair through distinct phosphoprotein interactions. Cell. 2013;153(6):1266–1280.
[12] Xu G, Chapman JR, Brandsma I, *et al.* REV7 counteracts DNA double-strand break resection and affects PARP inhibition. Nature. 2015;521(7553):541–544.
[13] Boersma V, Moatti N, Segura-Bayona S, *et al.* MAD2L2 controls DNA repair at telomeres and DNA breaks by inhibiting 5 end resection. Nature. 2015;521 (7553):537–540.
[14] Goodman MF, Woodgate R. Translesion DNA polymerases. Cold Spring Harb Perspect Biol. 2013;5 (10):a010363.
[15] Hara K, Hashimoto H, Murakumo Y, *et al.* Crystal structure of human REV7 in complex with a human REV3 fragment and structural implication of the interaction between DNA polymerase zeta and REV1. J Biol Chem. 2010;285(16):12299–12307.
[16] Gupta R, Somyajit K, Narita T, *et al.* DNA repair network analysis reveals shieldin as a key regulator of NHEJ and PARP inhibitor sensitivity. Cell. 2018;173 (4):972–988 e923.
[17] Noordermeer SM, Adam S, Setiaputra D, *et al.* The shieldin complex mediates 53BP1-dependent DNA repair. Nature. 2018;560(7716):117–121.
[18] Mirman Z, Lottersberger F, Takai H, *et al.* 53BP1-RIF1 -shieldin counteracts DSB resection through CST- and Polalpha-dependent fill-in. Nature. 2018;560 (7716):112–116.
[19] Ghezraoui H, Oliveira C, Becker JR, *et al.* 53BP1 cooperation with the REV7-shieldin complex underpins DNA structure-specific NHEJ. Nature. 2018;560 (7716):122–127.
[20] Dev H, Chiang T-WW, Lescale C, *et al.* Shieldin complex promotes DNA end-joining and counters homologous recombination in BRCA1-null cells. Nat Cell Biol. 2018;20(8):954–965.
[21] Tomida J, Takata K-I, Bhetawal S, *et al.* FAM 35A associates with REV 7 and modulates DNA damage responses of normal and BRCA 1-defective cells. Embo J. 2018;37(12). DOI:10.15252/embj.201899543.
[22] Findlay S, Heath J, Luo VM, *et al.* SHLD 2/ FAM 35A co-operates with REV 7 to coordinate DNA double-strand break repair pathway choice. Embo J. 2018;37(18). DOI:10.15252/embj.2018100158.
[23] Gao S, Feng S, Ning S, *et al.* An OB-fold complex controls the repair pathways for DNA double-strand breaks. Nat Commun. 2018;9(1):3925.
[24] Clairmont CS, Sarangi P, Ponnienselvan K, *et al.* TRIP13 regulates DNA repair pathway choice through REV7 conformational change. Nat Cell Biol. 2020;22 (1):87–96.
[25] Fradet-Turcotte A, Canny MD, Escribano-Díaz C, *et al.* 53BP1 is a reader of the DNA-damage-induced H2A Lys 15 ubiquitin mark. Nature. 2013;499(7456):50–54.
[26] Sanders SL, Portoso M, Mata J, *et al.* Methylation of histone H4 lysine 20 controls recruitment of Crb2 to sites of DNA damage. Cell. 2004;119(5):603–614.
[27] Botuyan MV, Lee J, Ward IM, *et al.* Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Crb2 in DNA repair. Cell. 2006;127(7):1361–1373.
[28] Mattiroli F, Vissers JA, van Dijk W, *et al.* RNF168 ubiquitinates K13-15 on H2A/H2AX to drive DNA damage signaling. Cell. 2012;150(6):1182–1195.
[29] Kolas NK, Chapman JR, Nakada S, *et al.* Orchestration of the DNA-damage response by the RNF8 ubiquitin ligase. Science. 2007;318(5856):1637–1640.
[30] Huen MS, Grant R, Manke I, *et al.* RNF8 transduces the DNA-damage signal via histone ubiquitylation and checkpoint protein assembly. Cell. 2007;131(5):901–914. .
[31] Mailand N, Bekker-Jensen S, Faustrup H, *et al.* RNF8 ubiquitylates histones at DNA double-strand breaks and promotes assembly of repair proteins. Cell. 2007;131(5):887–900.
[32] Doil C, Mailand N, Bekker-Jensen S, *et al.* RNF168 binds and amplifies ubiquitin conjugates on damaged chromosomes to allow accumulation of repair proteins. Cell. 2009;136(3):435–446.
[33] Stewart GS, Panier S, Townsend K, *et al.* The RIDDLE syndrome protein mediates a ubiquitin-dependent signaling cascade at sites of DNA damage. Cell. 2009;136 (3):420–434.
[34] Pesavento JJ, Yang H, Kelleher NL, et al. Certain and progressive methylation of histone H4 at lysine 20 during the cell cycle. Mol Cell Biol. 2008;28 (1):468–486.
[35] Drane P, Brault M-E, Cui G, *et al.* TIRR regulates 53BP1 by masking its histone methyl-lysine binding function. Nature. 2017;543(7644):211–216.
[36] Zhang A, Peng B, Huang P, et al. The p53-binding protein 1-Tudor-interacting repair regulator complex participates in the DNA damage response. J Biol Chem. 2017;292(16):6461–6467.
[37] Pellegrino S, Michelena J, Teloni F, et al. Replication-coupled dilution of H4K20me2 guides 53BP1 to pre-replicative chromatin. Cell Rep. 2017;19 (9):1819–1831.
[38] Simonetta M, de Krijger I, Serrat J, *et al.* H4K20me2 distinguishes pre-replicative from post-replicative chromatin to appropriately direct DNA repair pathway choice by 53BP1-RIF1-MAD2L2. Cell Cycle. 2018;17(1):124–136.
[39] Nelson G, Buhmann M, von Zglinicki T. DNA damage foci in mitosis are devoid of 53BP1. Cell Cycle. 2009;8 (20):3379–3383.

[40] van Vugt MA, Gardino AK, Linding R, *et al.* A mitotic phosphorylation feedback network connects Cdk1, Plk1, 53BP1, and Chk2 to inactivate the G(2)/M DNA damage checkpoint. PLoS Biol. 2010;8(1):e1000287.
[41] Orthwein A, Fradet-Turcotte A, Noordermeer SM, *et al.* Mitosis inhibits DNA double-strand break repair to guard against telomere fusions. Science. 2014;344 (6180):189–193.
[42] Peterson SE, Li Y, Chait BT, *et al.* Cdk1 uncouples CtIP-dependent resection and Rad51 filament formation during M-phase double-strand break repair. J Cell Biol. 2011;194(5):705–720.
[43] Aravind L, Koonin EV. The HORMA domain: a common structural denominator in mitotic checkpoints, chromosome synapsis and DNA repair. Trends Biochem Sci. 1998;23(8):284–286.
[44] Suzuki SW, Yamamoto H, Oikawa Y, *et al.* Atg13 HORMA domain recruits Atg9 vesicles during autophagosome formation. Proc Natl Acad Sci U S A. 2015;112(11):3350–3355.
[45] Jao CC, Ragusa MJ, Stanley RE, et al. A HORMA domain in Atg13 mediates PI 3-kinase recruitment in autophagy. Proc Natl Acad Sci U S A. 2013;110 (14):5486–5491.
[46] Michel M, Schwarten M, Decker C, *et al.* The mammalian autophagy initiator complex contains 2 HORMA domain proteins. Autophagy. 2015;11(12):2300–2308.
[47] Ye Q, Lau RK, Mathews IT, *et al.* HORMA domain proteins and a Trip13-like ATPase regulate bacterial cGAS-like enzymes to mediate bacteriophage immunity. Mol Cell. 2020;77(4):709–722 e707.
[48] Luo X, Tang Z, Xia G, *et al.* The Mad2 spindle checkpoint protein has two distinct natively folded states. Nat Struct Mol Biol. 2004;11(4):338–345.
[49] West AMV, Komives EA, Corbett KD. Conformational dynamics of the Hop1 HORMA domain reveal a common mechanism with the spindle checkpoint protein Mad2. Nucleic Acids Res. 2018;46(1):279–292.
[50] Chen RH, Shevchenko A, Mann M, et al. Spindle checkpoint protein XMAD1 recruits Xmad2 to unattached kinetochores. J Cell Biol. 1998;143(2):283–295.
[51] Luo X, Tang Z, Rizo J, et al. The Mad2 spindle checkpoint protein undergoes similar major conformational changes upon binding to either MAD1 or Cdc20. Mol Cell. 2002;9(1):59–71.
[52] Sironi L, Mapelli M, Knapp S, *et al.* Crystal structure of the tetrameric MAD1-Mad2 core complex: implications of a 'safety belt' binding mechanism for the spindle checkpoint. Embo J. 2002;21(10):2496–2506.
[53] De Antoni A, Pearson CG, Cimini D, *et al.* The MAD1/Mad2 complex as a template for Mad2 activation in the spindle assembly checkpoint. Curr Biol. 2005;15(3):214–225.
[54] Mapelli M, Massimiliano L, Santaguida S, et al. The Mad2 conformational dimer: structure and implications for the spindle assembly checkpoint. Cell. 2007;131(4):730–743.
[55] Qi S, Kim DJ, Stjepanovic G, et al. Structure of the Human Atg13-Atg101 HORMA Heterodimer: an Interaction Hub within the ULK1 Complex. Structure. 2015;23(10):1848–1857.
[56] Hwang LH, Lau LF, Smith DL, *et al.* Budding yeast Cdc20: a target of the spindle checkpoint. Science. 1998;279(5353):1041–1044.
[57] Luo X, Wagner G, Luo X, *et al.* Structure of the Mad2 spindle assembly checkpoint protein and its interaction with Cdc20. Nat Struct Biol. 2000;7(3):224–229.
[58] Hollingsworth NM, Byers B. HOP1: a yeast meiotic pairing gene. Genetics. 1989;121(3):445–462.
[59] Hollingsworth NM, Goetsch L, Byers B. The HOP1 gene encodes a meiosis-specific component of yeast chromosomes. Cell. 1990;61(1):73–84.
[60] Kim Y, Rosenberg S, Kugel C, *et al.* The chromosome axis controls meiotic events through a hierarchical assembly of HORMA domain proteins. Dev Cell. 2014;31(4):487–502.
[61] Tomida J, Takata K-I, Lange SS, *et al.* REV7 is essential for DNA damage tolerance via two REV3L binding sites in mammalian DNA polymerase ζ. Nucleic Acids Res. 2015;43(2):1000–1011.
[62] Hara K, Taharazako S, Ikeda M, *et al.* Dynamic feature of mitotic arrest deficient 2-like protein 2 (MAD2L2) and structural basis for its interaction with chromosome alignment-maintaining phosphoprotein (CAMP). J Biol Chem. 2017;292(43):17658–17667.
[63] Rizzo AA, Vassel F-M, Chatterjee N, *et al.* Rev7 dimerization is important for assembly and function of the Rev1/Polζ translesion synthesis complex. Proc Natl Acad Sci U S A. 2018;115(35):E8191–E8200.
[64] Wojtasz L, Daniel K, Roig I, *et al.* Mouse HORMAD1 and HORMAD2, two conserved meiotic chromosomal proteins, are depleted from synapsed chromosome axes with the help of TRIP13 AAA-ATPase. PLoS Genet. 2009;5(10):e1000702.
[65] Ye Q, Rosenberg SC, Moeller A, *et al.* TRIP13 is a protein-remodeling AAA+ ATPase that catalyzes MAD2 conformation switching. Elife. 2015;4. DOI:10.7554/eLife.07367.
[66] Ma HT, Poon RYC. TRIP13 regulates both the activation and inactivation of the spindle-assembly checkpoint. Cell Rep. 2016;14(5):1086–1099.
[67] San-Segundo PA, Roeder GS. Pch2 links chromatin silencing to meiotic checkpoint control. Cell. 1999;97 (3):313–324.
[68] Ho HC, Burgess SM, McKim KS. Pch2 acts through Xrs2 and Tel1/ATM to modulate interhomolog bias and checkpoint function during meiosis. PLoS Genet. 2011;7(11):e1002351.
[69] Li XC, Schimenti JC. Mouse pachytene checkpoint 2 (trip13) is required for completing meiotic recombination but not synapsis. PLoS Genet. 2007;3(8):e130.
[70] Chen C, Jomaa A, Ortega J, et al. Pch2 is a hexameric ring ATPase that remodels the chromosome axis protein Hop1. Proc Natl Acad Sci U S A. 2014;111(1): E44–53.

[71] Ye Q, Kim DH, Dereli I, *et al*. The AAA + ATP ase TRIP 13 remodels HORMA domains through N-terminal engagement and unfolding. Embo J. 2017;36(16):2419–2434.

[72] Eytan E, Wang K, Miniowitz-Shemtov S, *et al*. Disassembly of mitotic checkpoint complexes by the joint action of the AAA-ATPase TRIP13 and p31 (comet). Proc Natl Acad Sci U S A. 2014;111 (33):12019–12024.

[73] Yang M, Li B, Tomchick DR, *et al*. p31comet blocks Mad2 activation through structural mimicry. Cell. 2007;131(4):744–755.

[74] Ji J, Tang D, Shen Y, *et al*. P31 comet, a member of the synaptonemal complex, participates in meiotic DSB formation in rice. Proc Natl Acad Sci U S A. 2016;113(38):10577–10582.

[75] van Hooff JJ, Tromer E, van Wijk LM, et al. Evolutionary dynamics of the kinetochore network in eukaryotes as revealed by comparative genomics. EMBO Rep. 2017;18(9):1559–1571.

[76] Kaisari S, Shomer P, Ziv T, *et al*. Role of Polo-like kinase 1 in the regulation of the action of p31(comet) in the disassembly of mitotic checkpoint complexes. Proc Natl Acad Sci U S A. 2019;116(24):11725–11730. .

[77] Nelson CR, Hwang T, Chen PH, et al. TRIP13PCH-2 promotes Mad2 localization to unattached kinetochores in the spindle checkpoint response. J Cell Biol. 2015;211(3):503–516.

[78] Joshi N, Barot A, Jamison C, et al. Pch2 links chromosome axis remodeling at future crossover sites and crossover distribution during yeast meiosis. PLoS Genet. 2009;5(7):e1000557.

[79] Wang K, Sturt-Gillespie B, Hittle JC, *et al*. Thyroid hormone receptor interacting protein 13 (TRIP13) AAA-ATPase is a novel mitotic checkpoint-silencing protein. J Biol Chem. 2014;289(34):23928–23937.

[80] Kurita K, Maeda M, Mansour MA, *et al*. TRIP13 is expressed in colorectal cancer and promotes cancer cell invasion. Oncol Lett. 2016;12(6):5240–5246.

[81] Li W, Zhang G, Li X, *et al*. Thyroid hormone receptor interactor 13 (TRIP13) overexpression associated with tumor progression and poor prognosis in lung adenocarcinoma. Biochem Biophys Res Commun. 2018;499(3):416–424.

[82] Banerjee R, Russo N, Liu M, *et al*. TRIP13 promotes error-prone nonhomologous end joining and induces chemoresistance in head and neck cancer. Nat Commun. 2014;5(1):4527.

[83] Mellacheruvu D, Wright Z, Couzens AL, *et al*. The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. Nat Methods. 2013;10(8):730–736.

[84] Yost S, de Wolf B, Hanks S, *et al*. Biallelic TRIP13 mutations predispose to Wilms tumor and chromosome missegregation. Nat Genet. 2017;49(1):1148–1151.

[85] Marks DH, Thomas R, Chin Y, *et al*. Mad2 overexpression uncovers a critical role for TRIP13 in mitotic exit. Cell Rep. 2017;19(9):1832–1845.

[86] Chao WC, Kulkarni K, Zhang Z, et al. Structure of the mitotic checkpoint complex. Nature. 2012;484 (7393):208–213.

[87] Kikuchi S, Hara K, Shimizu T, et al. Structural basis of recruitment of DNA polymerase ζ by interaction between REV1 and REV7 proteins. J Biol Chem. 2012;287(40):33847–33852.

[88] Dai Y, Zhang F, Wang L, *et al*. Structural basis for shieldin complex subunit 3-mediated recruitment of the checkpoint protein REV7 during DNA double-strand break repair. J Biol Chem. 2020;295(1):250–262.

[89] Alfieri C, Chang L, Barford D. Mechanism for remodelling of the cell cycle checkpoint protein MAD2 by the ATPase TRIP13. Nature. 2018;559(7713):274–278.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.23.111658; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Polymerase Theta Inhibition Kills Homologous Recombination Deficient Tumors

Jia Zhou[1], Camille Gelot[2], Constantia Pantelidou[3], Adam Li[1], Hatice Yücel[2], Rachel E. Davis[4], Anniina Farkkila[1], Bose Kochupurakkal[1], Aleem Syed[5], Geoffrey I. Shapiro[3,6], John A. Tainer[5], Brian S. J. Blagg[4], Raphael Ceccaldi[2,7*] and Alan D. D'Andrea[1,6,7*]

[1]Department of Radiation Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA 02215, USA.

[2]Inserm U830, PSL Research University, Institut Curie, 75005, Paris, France.

[3]Department of Medical Oncology, Dana-Farber Cancer Institute and Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA.

[4]Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, IN 46556, USA.

[5]Departments of Cancer Biology and of Molecular and Cellular Oncology, University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA.

[6]Center for DNA Damage and Repair, Dana-Farber Cancer Institute, Boston, MA, USA.

[7]Co-senior authors.

* Co-corresponding authors.

*** Corresponding authors:**

Alan D. D'Andrea, M.D.
Director, Susan F. Smith Center for Women's Cancers (SFSCWC)
Director, Center for DNA Damage and Repair
Dana-Farber Cancer Institute
The Fuller-American Cancer Society Professor
Harvard Medical School
Phone: 617-632-2080
Email: Alan_Dandrea@dfci.harvard.edu

Raphael Ceccaldi
Institut Curie, 75005, Paris, France
Phone: +33 (0)1 56 24 69 49
Email: raphael.ceccaldi@curie.fr

Key Words: Novobiocin, Polymerase theta (POLθ), Homologous Recombination, PARP inhibitor resistance.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.23.111658; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

PARP inhibitors (PARPi) have become a new line of therapy for Homologous Recombination (HR)-deficient cancers. However, resistance to PARPi has emerged as a major clinical problem. DNA polymerase theta (POLθ) is synthetic lethal with HR and a druggable target in HR-deficient cancers. Here, we identified the antibiotic Novobiocin (NVB) as a specific POLθ inhibitor that selectively kills HR-deficient tumor cells *in vitro* and *in vivo*. NVB directly binds to the POLθ ATPase domain, inhibits its ATPase activity, and phenocopies POLθ depletion. BRCA-deficient tumor cells and those with acquired PARPi resistance are sensitive to NVB *in vitro* and *in vivo*. Increased POLθ expression levels predict NVB sensitivity. The mechanism of NVB-mediated cell death in PARPi resistant cells is the accumulation of toxic RAD51 foci, which also provides a pharmacodynamic biomarker for NVB response. Our results demonstrate that NVB may be useful alone or in combination with PARPi in treating HR-deficient tumors, including those with acquired PARPi resistance.

One Sentence Summary: We identified Novobiocin as a specific POLθ inhibitor that selectively kills naïve and PARPi resistance HR-deficient tumors *in vitro* and *in vivo.*

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.23.111658; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Since the discovery of the synthetic lethal relationship between PARP inhibition and homologous recombination (HR) repair deficiency (*1, 2*), PARP inhibitors (PARPi) have become a new line of therapy for HR-deficient cancers (*3-7*). However, resistance to PARPi has emerged as a major obstacle to their clinical effectiveness (*8*). While several mechanisms of PARPi resistance are known (*9-14*), a rational and effective method for overcoming such resistance is still lacking. DNA polymerase theta (POLθ or POLQ), a key enzyme in microhomology-mediated end joining (MMEJ), is synthetic lethal with HR and a druggable target in HR-deficient cancers (*15-18*). A small-molecule screen identified novobiocin (NVB) as a specific POLQ inhibitor. NVB inhibited POLQ *in vitro* and *in vivo*, Moreover, we demonstrated that NVB can be used alone or in combination with PARPi in treating HR-deficient tumors, even after they have acquired PARPi resistance.

RESULTS

A small-molecule screen identified NVB as a specific POLθ inhibitor

Polymerase theta (POLθ, encoded by the *POLQ* gene) is a critical enzyme for repairing DNA double-strand breaks (DSBs) through a microhomology-mediated end joining (MMEJ) mechanism (*19, 20*). POLθ expression is elevated in several types of cancers where it correlates with poor prognosis (*21, 22*). HR-deficient tumors upregulate POLθ-mediated MMEJ as a survival strategy, and loss of POLθ in these tumors results in cell death (*15, 17*). This synthetic lethality with HR defines the rationale for developing POLθ inhibitors for the treatment of HR-deficient cancers (*16*).

The ATPase activity of POLθ is required for the survival of HR-deficient cells (*17*). We took advantage of the strong *in vitro* ATPase activity of purified POLθ (**Fig. S1A**) and designed a large-scale small-molecule screen to identify POLθ inhibitors based on the ADP-Glo luminescent assay (**Fig. 1A, B**). The hits from the primary screen were evaluated in a

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.23.111658; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

secondary screen (**Fig. S1B-D**). The four most potent POLθ ATPase inhibitors were Mitoxantrone (MTX), Suramin (SUR), Novobiocin (NVB) and Aurintricarboxylic acid (AUR), which were individually verified in a ^{32}P-based radiometric ATPase assay (**Fig. S1E**). To test the specificity of the most potent compounds, we purified two other known DNA repair-related ATPases, namely SMARCAL1 and CHD1 (*23, 24*), and performed ATPase inhibition assays. Only NVB showed high specific inhibition of POLθ with little effect on SMARCAL1 or CHD1 activity, while MTX, SUR and AUR were nonspecific pan-ATPase inhibitors that inhibited all three ATPases (**Fig. 1C and S1F, G**).

To further evaluate the specificity of Novobiocin (NVB) for POLθ, we analyzed its binding capacity. NVB-conjugated beads pulled down the purified ATPase domain of POLθ but not of SMARCAL1, suggesting that NVB directly binds to the POLθ ATPase domain *in vitro* (**Fig. 1D, S1H**). Active NVB-beads also pulled down GFP-tagged full-length POLθ overexpressed in HEK293T cells (**Fig. S1I**). Excess free NVB competed the POLθ binding to the NVB-beads (**Fig. S1J**). Next, we performed protein thermal shift assays, an assay commonly used to demonstrate ligand-protein binding (*25, 26*). NVB significantly increased the thermal stability of POLθ when compared to DMSO (**Fig. 1E**), while the PARP1 inhibitor olaparib, used as an internal control, effectively stabilized the PARP1 protein (**Fig. S1K**). Finally, we performed molecular docking to find the NVB binding site using the published crystal structure of POLQ ATPase domain (*27*). The computational docking results revealed that a deep tunnel within the POLQ ATPase domain could accommodate NVB, and that one end of NVB could reach the active site to prevent ATP binding to POLQ and inhibit its ATPase activity (**Fig. 1F**). Taken together, these data demonstrate that NVB binds directly and specifically to the POLθ ATPase domain *in vitro*.

We next assessed whether cellular exposure to NVB has the same consequences as POLθ depletion. Both the ATPase and the polymerase domains contribute to efficient

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.23.111658; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

MMEJ repair activity (*28*), and the PARP-dependent recruitment of POLθ to DSB sites is a crucial step in this function (*15, 17*). The inhibition of POLθ by NVB prevented the recruitment of POLθ to laser micro-irradiated DNA damage sites in human cells (**Fig. 1G**). Consequently, NVB inhibited the MMEJ activity in U2OS cells, as demonstrated by the GFP-based EJ2 repair assay (**Fig. 1H**). In contrast, NVB had little effect on HR activity, as measured by the DR-GFP reporter assay (**Fig. S2A**). POLθ also functions as an anti-recombinase that antagonizes HR, and depletion of POLθ increases RAD51 (*17*) and γH2AX (*29, 30*) foci assembly. Again, cellular exposure to NVB recapitulated the RAD51 and γH2AX phenotypes of a POLθ knockdown. NVB activated more RAD51 foci (**Fig. S2B**) and γH2AX foci (**Fig. S2C**) in RPE1 cells, regardless of IR exposure. Taken together, these results establish that NVB-mediated POLθ inhibition impairs POLθ DNA repair function and phenocopies POLθ depletion in human cells.

NVB specifically Kills HR-deficient cells *in vivo* and *in vitro*

Depletion of POLθ is synthetically lethal in HR-deficient tumor cells (*15, 17*). Since NVB has been previously approved for human use, we decided to directly test whether NVB reduces the survival of HR-deficient tumor cells in animal models. First, we evaluated the efficacy of NVB in tumors derived from the K14-Cre-*Brca1*$^{f/f}$*;Trp53*$^{f/f}$ genetically-engineered mouse model (GEMM) of triple-negative breast cancer (TNBC), in which spontaneous mammary carcinomas develop after approximately 7 months (*31, 32*). Tumors from this model were transplanted into immunocompetent FVB/129P2 syngeneic mice that were subsequently treated with NVB. NVB strongly suppressed the growth of these *Brca1*-deficient GEMM-derived tumors (**Fig. 2A** and individual mouse data in **Fig. S3A**). The NVB-treated tumors were significantly smaller compared to the vehicle-treated after 7 days of treatment and beyond. NVB prolonged the overall survival of the tumor-bearing mice by 3-

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.23.111658; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

fold compared to vehicle, with median survival times of 29 and 10 days for NVB- and vehicle-treated mice, respectively (**Fig. 2B**).

Next we tested whether the efficacy of NVB *in vivo* was specific to HR-deficient tumors. We performed a mouse xenograft study with a FANCF-deficient, HR-deficient ovarian cancer cell line [TOV21G + empty vector (EV)] or with the FANCF-complemented, HR-proficient cell line (TOV21G + FANCF cDNA) (**Fig. 2C**). TOV21G + EV cells but not TOV21G + FANCF cDNA cells were sensitive to olaparib and NVB (**Fig. S3B, C**) (*33*). While vehicle treatment had no impact on the growth of FANCF-deficient or FANCF-complemented tumors, NVB specifically impaired the growth of the FANCF-deficient tumors, with no effect on FANCF-complemented tumors (**Fig. 2C**). We also measured RAD51 foci as a pharmacodynamic biomarker and showed that RAD51 foci were strongly induced in the treated tumors (**Fig. S3D-F**). These results demonstrate that NVB strongly and specifically suppresses the growth of HR-deficient tumors *in vivo*.

To confirm and strengthen the *in vivo* results, we studied isogenic pairs of HR-deficient and HR-proficient cells *in vitro*. We generated *BRCA1* (*34*) and *BRCA2* knockout RPE1 cells (referred as *BRCA1*$^{-/-}$ and *BRCA2*$^{-/-}$) in a *TP53*$^{-/-}$ background, which were sensitive to PARPi (**Fig. S4A**). Consistent with our *in vivo* data, clonogenic survival assays showed that NVB significantly reduced the survival of *BRCA1*$^{-/-}$ and *BRCA2*$^{-/-}$ cells, compared to the isogenic WT cells (**Fig. 2D, S4B and Fig. S4C**). We also tested the top 15 initial hits from our screen for their ability to differentially affect the survival of *BRCA1*$^{-/-}$ and WT cells. Among those, NVB had the highest efficacy in selectively killing the *BRCA1*$^{-/-}$ cells (**Fig. S4D**).

To demonstrate the cellular outcome of NVB exposure, we next showed that NVB exposure induces apoptosis in *BRCA1*$^{-/-}$ but not WT cells in a dose dependent manner (**Fig. S5A**). NVB induced chromosomal aberrations and radial chromosomes (a marker of

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.23.111658; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

genomic instability) in *BRCA1*$^{-/-}$ cells, even in the absence of the crosslinking agent mitomycin C (MMC) (**Fig. S5B-D**), confirming that NVB increases DNA damage. These results further demonstrate that NVB phenocopies POLθ depletion, and indicate that NVB-mediated POLθ inhibition is synthetically lethal with HR deficiency and specifically induces cell death in HR-deficient tumor cells.

POLθ is the major target of NVB in human cells

Off target effects are major concerns of biological small inhibitors. To further evaluate whether POLθ is the specific target of NVB and to demonstrate the specificity of NVB in cells, we knocked out the *POLQ* gene in U2OS and RPE1 cells using CRISPR-Cas9 and asked whether loss of POLθ alleviates the cytotoxicity of NVB (**Fig. 2E and S6**). POLQ knockout cells were confirmed by genomic sequencing, immunoblotting, increased RAD51 foci, increased sensitivity to irradiation, and increased micronuclei formation (**Fig. S6A-F**). Importantly, *POLQ*$^{-/-}$ cells showed increased tolerance to NVB, both in U2OS and RPE1 cells, as compared to the WT cells (**Fig. 2E and S6G**), indicating that POLθ is the major target of NVB in human cells.

NVB is a known coumarin antibiotic that inhibits DNA gyrase in bacteria (*35*). NVB has been reported to inhibit the ATPase of HSP90 and of the DNA gyrase homolog topoisomerase II (TOP2) in eukaryotes (*36*) (*37*) (*38*), suggesting that these enzymes may be off-target enzymes of NVB. However, other reports suggest that NVB does not inhibit TOP2 in eukaryotic cells (*39*). Additionally, the reported IC_{50} of these targets are very high (approximately 700 μM for HSP90 and 300 μM for TOP2), compared with the IC_{50} of POLθ. Nevertheless, we next examined whether the cytotoxic effect of NVB in HR-deficient cells resulted from off-target inhibition of HSP90 or TOP2. Unlike the well-characterized HSP90 inhibitor PU-H71, NVB treatment at a concentration that killed HR-deficient cells (i.e. 100

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.23.111658; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

μM) did not induce the degradation of HSP90 clients such as AKT1, BRCA1, or CDK6, or increase the expression of HSP70 (**Fig. S7A-C**). These results suggest that the cytotoxic effect of NVB is not related to HSP90 inhibition. Furthermore, the combination of NVB and the TOP2 inhibitor etoposide exhibited additive cytotoxicity in TOV21G and CAPAN1 cells (**Fig. S7D, E**), suggesting that NVB and etoposide do not share TOP2 as a target and that NVB cytotoxicity is unrelated to TOP2 inhibition.

NVB potentiates the effect of PARPi in HR-deficient tumors

PARPi are currently being evaluated in different combination therapy settings with cytotoxic chemotherapy, radiation, targeted therapies, and immunotherapies (*40*). We explored whether the combination of NVB and a PARPi may be more effective than PARP inhibition alone in killing HR-deficient tumors. Indeed, NVB further sensitized *BRCA1*$^{-/-}$ RPE1 cells to the PARPi rucaparib, but not BRCA-proficient WT cells in clonogenic assays (**Fig. 2F**). We studied the combination with a second PARPi, olaparib, where synergism of NVB and olaparib was also observed in HR-deficient TOV21G+EV cells, but not in HR-proficient TOV21G+FANCF cells (**Fig. S7F**). Importantly, NVB reduced the IC_{50} of rucaparib by more than 20-fold in *BRCA1*$^{-/-}$ RPE1 cells and the IC_{50} of olaparib more than 40-fold in TOV21G cells compared to their respective HR-proficient counterparts (**Fig. 2F and S7F**). One strategy to prevent acquired resistance during treatment is combination therapy. The observed synergy between NVB and PARPi prompted us to evaluate whether NVB could overcome PARPi resistance in HR-deficient tumors.

POLθ inhibition by NVB overcomes acquired PARPi resistance *in vivo* and *in vitro*

Multiple PARP inhibitors (PARPi) including olaparib, niraparib, rucaparib, and talazoparib have received approval for the treatment of ovarian and breast tumors with HR

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.23.111658; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

deficiency (*41-43*). Despite the remarkable initial tumor response and improved progression-free survival (PFS) in HR-deficient patients (*4-6, 44*), acquired PARPi resistance is emerging as an unmet medical need (*8*).

To evaluate whether POLθ inhibition could represent a novel therapeutic option for PARPi-resistant tumors, we first tested the efficacy *in vivo* of NVB in the patient-derived xenograft (PDX) model DF83, generated from an HR-deficient (loss of RAD51C expression), human high-grade serous ovarian carcinoma (*45*). NSG mice bearing the HR-deficient DF83 ovarian cancer PDX model were treated with olaparib, NVB, or a combination of the two, and tumor growth was monitored by bioluminescence imaging (BLI) (**Fig. 3A**). In order to determine the combined effect with NVB, olaparib was delivered at a submaximal dose (50 mg/kg daily), and we observed only a small degree of tumor growth inhibition. Treatment of DF83 with NVB alone (75 mg/kg, twice a day) led to tumor regression, while tumors in vehicle-treated mice showed exponential growth (**Fig. 3B**). Strikingly, when the DF83 tumor-bearing mice were given NVB and olaparib in combination, we observed complete tumor regression, and few tumor cells were detectable via BLI on day 28 in this group (**Fig. 3A-B**). These *in vivo* results suggest NVB may be useful in combination with PARPi to achieve better efficacy and prevent drug resistance in HR-deficient tumors. No significant toxicity to normal mouse tissue was observed from the drug combination.

Next, we evaluated the efficacy of NVB alone and in combination with olaparib in the PARPi-resistant PDX model DF59, which was isolated from a patient with a germline *BRCA1* mutation and acquired PARPi resistance. This model harbors biallelic mutations in the *TP53BP1* locus, providing a mechanism of the acquired PARPi resistance (*45*). In the DF59 PDX model, there was no response to olaparib monotherapy, but NVB monotherapy substantially reduced tumor growth, indicating a lack of cross-resistance between PARP inhibition and NVB. Importantly, the combination of NVB and olaparib further inhibited DF59

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.23.111658; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

tumor growth, with tumor regression in the first two weeks (**Fig. 3C-D**). In contrast to DF83 and DF59 models, the BRCA1 WT, HR proficient PDX model DF149 was resistant to both drugs and to the combination (**Fig. 3E-F**). These data demonstrate that NVB can inhibit the growth of a *BRCA1*-mutated, PARPi-resistant tumors *in vivo* and that the combination of NVB and a PARPi is particularly effective in treating at least some PARPi-resistant tumors.

RAD51 focus is a pharmacodynamic biomarker for NVB response *in vivo*

We previously showed that depletion of POLθ in HR-deficient tumors resulted in RAD51 accumulation (*17*). Similarly, 53BP1 and POLθ are synthetic lethal, and the double knockout cells accumulate large, toxic RAD51 foci (*46*). We thus reasoned that NVB may induce RAD51 foci in HR-deficient tumors with PARPi resistance *in vivo*, and the RAD51 foci may serve as a pharmacodynamic biomarker for NVB response. To test this hypothesis, we collected tumor cells from the PARPi and/or NVB treated PDX models and evaluated RAD51 foci by immunohistochemistry (**Fig. 3G)**. Since DF83 is an HR-deficient and PARPi sensitive model, RAD51 foci were not observed. The HR-restored and PARPi-resistant DF59 model, with biallelic 53BP1 mutations, exhibited a strong increase in RAD51 foci when treated with NVB alone or in combination with olaparib (**Fig. 3G)**. Similarly, TOV21G xenograft tumors also exhibited increased RAD51 foci by immunohistochemistry *in vivo* after NVB treatment, which was further enhanced by olaparib (**Fig. S3D-F**). Results from these two *in vivo* studies demonstrate that RAD51 focus can serve as a pharmacodynamic biomarker for the response of tumors to NVB or other POLθ inhibitors. Interestingly, we also observed that two NVB-responsive tumor models had elevated levels of POLθ mRNA expression, compared to the NVB-resistant DF149 model, suggesting that elevated POLθ may be a useful predictive biomarker for POLθ inhibitor response (**Fig. 3H**).

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.23.111658; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

NVB overcomes multiple PARPi resistance mechanisms but not perfect BRCA reversions

To better understand the mechanism by which NVB overcomes PARPi resistance and, we generated several PARPi-resistant clones from *BRCA1*$^{-/-}$ *RPE1* cells by gradually exposing them to increasing concentrations of PARPi (**Fig. 4A**). We selected four clones (named R1 to R4), all of which have acquired resistance to olaparib (**Fig. S8A-C**). Multiple PARPi resistance mechanisms were identified in these clones, corresponding to published mechanisms (*8*). Replication fork stabilization was evident in the R1 and R2 clones (**Fig. S8D**). The R1, R3, and R4 clones have restored RAD51 foci, a marker for restored HR activity (**Fig. S8E**). The restoration of HR repair in these clones may have resulted in part from their downregulation of NHEJ repair. Indeed, the R3 clone had decreased REV7 expression, and the R4 clone had decreased 53BP1 expression (**Fig.S8F-G**). None of the clones had re-expressed BRCA1, thereby excluding somatic reversion of BRCA1 as an underlying PARPi resistance mechanism (**Fig. 4B and S8H**). Strikingly, all four PARPi-resistant clones remained sensitive to NVB to a similar extent as the parental RPE-*BRCA1*$^{-/-}$ cells (**Fig. 4C-D**), suggesting that NVB may overcome multiple mechanisms of acquired resistance to PARPi. To confirm that POLθ inhibition was responsible for cell death in the PARPi-resistant *BRCA1*$^{-/-}$ clones treated with NVB, we knocked out *POLQ* in R1, R2, and the parental cells and measured their cell survival. Upon POLQ knockout, the PARPi-resistant clones showed reduced survival similar to the parental *BRCA1*$^{-/-}$ cells, demonstrating that POLQ inhibition was the mechanism for NVB cytotoxicity in these cells (**Fig. S8I**).

Next, we examined more clinically relevant PARPi-resistant, HR-deficient models, using tumor cells derived from patients. First, we evaluated the effect of NVB on a PARPi resistant clone (MDA-MB-436-R), previously derived from the BRCA1-mutated breast

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.23.111658; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

cancer cell line MDA-MB-436 (*47*). These resistant cells restored HR by two mechanism- namely, stabilization and increased expression of a mutant BRCA1 protein and loss of 53BP1- resulting in increased RAD51 loading. While highly resistant to olaparib, the MDA-MB-436-R cells were NVB sensitive, like the parental cells (**Fig. S9A-B**). In addition, we studied the PARPi-resistant clone UWB1.289-YSR12 (derived from the *BRCA1*-null human ovarian cancer cell line UWB1.289), which acquired resistance to PARPi via stabilization of replication forks (*48*). Although highly resistant to olaparib, these cells also remained sensitive to NVB, like the parental cells (**Fig. S9C-D**), once again suggesting that acquired resistance to PARPi does not determine cross-resistance to NVB.

Another known mechanism of PARPi resistance in *BRCA1/2*$^{-/-}$ tumors in the clinic is somatic reversion of the BRCA1/2 genes (*18*). We next determined whether NVB could overcome PARPi resistance caused by BRCA2 reversion. Unlike other PARPi resistant cells we analyzed, BRCA2-deficient tumor cells with near-perfect somatic reversion or CRISPR-edited reversion of BRCA2 (PEO4 and CAPAN1-CR) (*49, 50*) were not only resistant to PARPi but also to NVB, when compared to their HR-deficient parental cell lines (PEO1 and CAPAN1) **(Fig. S10A-D)**. These data suggest that NVB can overcome some but not all mechanisms of PARPi resistance.

POLθ expression is a predictive biomarker for NVB sensitivity

POLθ mRNA expression is increased in HR-deficient tumors and in several other cancers associated with poor prognosis (*17, 21*). We asked whether increased POLθ mRNA and protein expression may serve as a predictive biomarker for NVB sensitivity. As expected, CRISPR-mediated knockout of *BRCA1* in RPE-1 cells resulted in increased POLθ expression as compare to the parental cells (**Fig. 4E**). Interestingly, all PARPi-resistant clones (R1-R4) retain high POLθ expression, revealing a direct correlation between POLθ

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.23.111658; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

expression and NVB sensitivity (**Fig. 4E**). This correlation was also present in the BRCA1-deficient MDA-MB-436 cells. MDA-MB-436-R cells with acquired PARPi-resistance exhibited elevated POLθ mRNA expression and retained NVB sensitivity, while complementation of the parental cells with WT BRCA1 cDNA decreased POLθ expression and NVB sensitivity (**Fig. 4F and S9A-B**). More importantly, the direct correlation of POLθ expression and NVB sensitivity was observed in HR-deficient PDX models, where the elevated POLθ expression in DF83 and DF59 correlated with their high sensitivity to NVB *in vivo* (**Fig. 3H and 3A-F**).

Conversely, low POLθ expression was a predictive biomarker for NVB resistance. In the HR-proficient PDX cells (DF149) with wildtype BRCA1 and in the PARPi-resistant cell lines with BRCA2 reversion mutations (PEO4 and CAPAN1-CR), the cellular resistance to NVB was associated with low POLθ expression levels (**Fig. 3G and S10E**). Also, the loss of POLθ in *POLQ*$^{-/-}$ cells correlated with resistance to NVB (**Fig. 2E**). Taken together, in all cases examined, POLθ expression level correlated with cellular sensitivity to NVB, suggesting that POLθ expression in a tumor biopsy could provide a convenient predictive biomarker for patient enrollment in future clinical trials with NVB or other POLθ inhibitors.

Toxic RAD51 foci is a mechanism of NVB cytotoxicity in PARPi resistant cells

Depletion of POLθ in HR-deficient tumors resulted in RAD51 accumulation (*17*), which can be toxic if not regulated. Similarly, 53BP1 and POLθ are synthetic lethal, and the double knockout cells accumulate toxic RAD51 foci (*46*). We reasoned that accumulation of toxic RAD51 foci induced by NVB may be responsible for killing the PARPi resistant tumor cells. This is supported by the RAD51 pharmacodynamic results, where RAD51 foci were greatly increased in tumors after NVB treatment *in vivo* (**Fig. 3G and S3D-F**). We next determined RAD51 foci dynamics in our PARPi resistant cell line models. NVB induced

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.23.111658; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

accumulation RAD51 foci in PARPi-resistant RPE1-BRCA1$^{-/-}$ clones (R1 and R4) more readily than in WT RPE1 cells after irradiation (**Fig. 4G**). Similarly, NVB treatment alone strongly induced RAD51 foci in the PARPi resistant MDA-MB-436-R cells (**Fig. 4H**). In addition, knockdown of RAD51 sensitized RPE1 WT cells to NVB but rescued *BRCA1*$^{-/-}$ R1 clone from NVB cytotoxicity (**Fig. 4I**). Taken together, these results suggest toxic RAD51 foci, induced by NVB, is a mechanism for NVB cytotoxicity in HR-deficient cells with acquired PARPi resistance.

DISCUSSION

In summary, NVB is a specific and potent POLθ inhibitor which phenocopies POLθ depletion and is synthetic lethal in HR-deficient tumors. Many tumors types with PARPi resistance remain sensitive to NVB, depending on their mechanism of resistance. Tumors which acquire PARPi resistance by downregulating NHEJ and partially restoring HR repair remain sensitive to NVB. For these tumors, the persistent high level of POLθ expression is a predictive biomarker for their NVB responsiveness, and the NVB-induced RAD51 accumulation is both a pharmacodynamic biomarker and a mechanism of NVB cytotoxic activity.

Finally, our findings may translate to a potential new therapy for treating HR-deficient tumors in addition to PARP inhibition. Importantly, NVB, which is an oral and well-tolerated drug, has previously been investigated in a Phase 1 trial in which partial responses were documented, even though the enrolled patients were not preselected to have HR-deficient cancers (*51*). Results in this study provide a strong rationale for future clinical studies of NVB alone or in combination with PARPi to prevent or overcome PARPi resistance.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.23.111658; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

REFERENCES

1. H. Farmer *et al.*, Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* **434**, 917-921 (2005).

2. H. E. Bryant *et al.*, Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature* **434**, 913-917 (2005).

3. J. Ledermann *et al.*, Olaparib maintenance therapy in platinum-sensitive relapsed ovarian cancer. *N Engl J Med* **366**, 1382-1392 (2012).

4. R. L. Coleman *et al.*, Rucaparib maintenance treatment for recurrent ovarian carcinoma after response to platinum therapy (ARIEL3): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet* **390**, 1949-1961 (2017).

5. M. R. Mirza *et al.*, Niraparib Maintenance Therapy in Platinum-Sensitive, Recurrent Ovarian Cancer. *N Engl J Med* **375**, 2154-2164 (2016).

6. J. K. Litton *et al.*, Talazoparib in Patients with Advanced Breast Cancer and a Germline BRCA Mutation. *N Engl J Med* **379**, 753-763 (2018).

7. C. J. Lord, A. Ashworth, PARP inhibitors: Synthetic lethality in the clinic. *Science* **355**, 1152-1158 (2017).

8. C. J. Lord, A. Ashworth, Mechanisms of resistance to therapies targeting BRCA-mutant cancers. *Nat Med* **19**, 1381-1388 (2013).

9. W. Sakai *et al.*, Secondary mutations as a mechanism of cisplatin resistance in BRCA2-mutated cancers. *Nature* **451**, 1116-1120 (2008).

10. A. Ray Chaudhuri *et al.*, Replication fork stability confers chemoresistance in BRCA-deficient cells. *Nature* **535**, 382-387 (2016).

11. G. Xu *et al.*, REV7 counteracts DNA double-strand break resection and affects PARP inhibition. *Nature* **521**, 541-544 (2015).

12. S. M. Noordermeer *et al.*, The shieldin complex mediates 53BP1-dependent DNA repair. *Nature* **560**, 117-121 (2018).

13. H. Ghezraoui *et al.*, 53BP1 cooperation with the REV7-shieldin complex underpins DNA structure-specific NHEJ. *Nature* **560**, 122-127 (2018).

14. Y. J. He *et al.*, DYNLL1 binds to MRE11 to limit DNA end resection in BRCA1-deficient cells. *Nature* **563**, 522-526 (2018).

15. P. A. Mateos-Gomez *et al.*, Mammalian polymerase theta promotes alternative NHEJ and suppresses recombination. *Nature* **518**, 254-257 (2015).

16. G. S. Higgins, S. J. Boulton, Beyond PARP-POLtheta as an anticancer target. *Science* **359**, 1217-1218 (2018).

17. R. Ceccaldi *et al.*, Homologous-recombination-deficient tumours are dependent on Poltheta-mediated repair. *Nature* **518**, 258-262 (2015).

18. P. A. Konstantinopoulos, R. Ceccaldi, G. I. Shapiro, A. D. D'Andrea, Homologous Recombination Deficiency: Exploiting the Fundamental Vulnerability of Ovarian Cancer. *Cancer Discov* **5**, 1137-1154 (2015).

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.23.111658; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

19. D. M. Muzzini, P. Plevani, S. J. Boulton, G. Cassata, F. Marini, Caenorhabditis elegans POLQ-1 and HEL-308 function in two distinct DNA interstrand cross-link repair pathways. *DNA Repair (Amst)* **7**, 941-950 (2008).

20. S. H. Chan, A. M. Yu, M. McVey, Dual roles for DNA polymerase theta in alternative end-joining repair of double-strand breaks in Drosophila. *PLoS Genet* **6**, e1001005 (2010).

21. F. Lemee *et al.*, DNA polymerase theta up-regulation is associated with poor survival in breast cancer, perturbs DNA replication, and promotes genetic instability. *Proc Natl Acad Sci U S A* **107**, 13390-13395 (2010).

22. G. S. Higgins *et al.*, Overexpression of POLQ confers a poor prognosis in early breast cancer patients. *Oncotarget* **1**, 175-184 (2010).

23. J. Zhou *et al.*, Human CHD1 is required for early DNA-damage signaling and is uniquely regulated by its N terminus. *Nucleic Acids Res* **46**, 3891-3905 (2018).

24. T. Yusufzai, J. T. Kadonaga, HARP is an ATP-driven annealing helicase. *Science* **322**, 748-750 (2008).

25. M. Vedadi *et al.*, Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination. *Proc Natl Acad Sci U S A* **103**, 15835-15840 (2006).

26. D. Martinez Molina *et al.*, Monitoring drug target engagement in cells and tissues using the cellular thermal shift assay. *Science* **341**, 84-87 (2013).

27. J. A. Newman, C. D. O. Cooper, H. Aitkenhead, O. Gileadi, Structure of the Helicase Domain of DNA Polymerase Theta Reveals a Possible Role in the Microhomology-Mediated End-Joining Pathway. *Structure* **23**, 2319-2330 (2015).

28. K. Beagan *et al.*, Drosophila DNA polymerase theta utilizes both helicase-like and polymerase domains during microhomology-mediated end joining and interstrand crosslink repair. *PLoS Genet* **13**, e1006813 (2017).

29. M. J. Yousefzadeh, R. D. Wood, DNA polymerase POLQ and cellular defense against DNA damage. *DNA Repair (Amst)* **12**, 1-9 (2013).

30. G. S. Higgins *et al.*, A small interfering RNA screen of genes involved in DNA repair identifies tumor-specific radiosensitization by POLQ knockdown. *Cancer Res* **70**, 2984-2993 (2010).

31. X. Liu *et al.*, Somatic loss of BRCA1 and p53 in mice induces mammary tumors with features of human BRCA1-mutated basal-like breast cancer. *Proceedings of the National Academy of Sciences* **104**, 12111-12116 (2007).

32. C. Pantelidou *et al.*, PARP Inhibitor Efficacy Depends on CD8(+) T-cell Recruitment via Intratumoral STING Pathway Activation in BRCA-Deficient Models of Triple-Negative Breast Cancer. *Cancer Discov* **9**, 722-737 (2019).

33. T. Taniguchi *et al.*, Disruption of the Fanconi anemia-BRCA pathway in cisplatin-sensitive ovarian tumors. *Nat Med* **9**, 568-574 (2003).

34. K. S. Lim *et al.*, USP1 Is Required for Replication Fork Protection in BRCA1-Deficient Tumors. *Mol Cell* **72**, 925-941 e924 (2018).

35. A. Sugino, N. P. Higgins, P. O. Brown, C. L. Peebles, N. R. Cozzarelli, Energy coupling in DNA gyrase and the mechanism of action of novobiocin. *Proc Natl Acad Sci U S A* **75**, 4838-4842 (1978).

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.23.111658; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

36. M. G. Marcu, A. Chadli, I. Bouhouche, M. Catelli, L. M. Neckers, The heat shock protein 90 antagonist novobiocin interacts with a previously unrecognized ATP-binding domain in the carboxyl terminus of the chaperone. *J Biol Chem* **275**, 37181-37186 (2000).

37. T. Hsieh, D. Brutlag, ATP-dependent DNA topoisonmerase from D. melanogaster reversibly catenates duplex DNA rings. *Cell* **21**, 115-125 (1980).

38. T. Goto, J. C. Wang, Yeast DNA topoisomerase II. An ATP-dependent type II topoisomerase that catalyzes the catenation, decatenation, unknotting, and relaxation of double-stranded DNA rings. *J Biol Chem* **257**, 5866-5872 (1982).

39. M. J. Pocklington, J. R. Jenkins, E. Orr, The effect of novobiocin on yeast topoisomerase type II. *Mol Gen Genet* **220**, 256-260 (1990).

40. A. Drean, C. J. Lord, A. Ashworth, PARP inhibitor combination therapy. *Crit Rev Oncol Hematol* **108**, 73-85 (2016).

41. T. Popova *et al.*, Ploidy and large-scale genomic instability consistently identify basal-like breast carcinomas with BRCA1/2 inactivation. *Cancer Res* **72**, 5454-5462 (2012).

42. M. R. Mirza, S. Pignata, J. A. Ledermann, Latest clinical evidence and further development of PARP inhibitors in ovarian cancer. *Ann Oncol* **29**, 1366-1376 (2018).

43. L. B. Alexandrov *et al.*, Signatures of mutational processes in human cancer. *Nature* **500**, 415-421 (2013).

44. E. Pujade-Lauraine *et al.*, Olaparib tablets as maintenance therapy in patients with platinum-sensitive, relapsed ovarian cancer and a BRCA1/2 mutation (SOLO2/ENGOT-Ov21): a double-blind, randomised, placebo-controlled, phase 3 trial. *Lancet Oncol* **18**, 1274-1284 (2017).

45. K. Parmar *et al.*, The CHK1 Inhibitor Prexasertib Exhibits Monotherapy Activity in High-Grade Serous Ovarian Cancer Models and Sensitizes to PARP Inhibition. *Clin Cancer Res*, (2019).

46. W. Feng *et al.*, Genetic determinants of cellular addiction to DNA polymerase theta. *Nat Commun* **10**, 4286 (2019).

47. N. Johnson *et al.*, Stabilization of mutant BRCA1 protein confers PARP inhibitor and platinum resistance. *Proc Natl Acad Sci U S A* **110**, 17041-17046 (2013).

48. S. A. Yazinski *et al.*, ATR inhibition disrupts rewired homologous recombination and fork protection pathways in PARP inhibitor-resistant BRCA-deficient cancer cells. *Genes Dev* **31**, 318-332 (2017).

49. L. J. Barber *et al.*, Comprehensive genomic analysis of a BRCA2 deficient human pancreatic cancer. *PLoS One* **6**, e21639 (2011).

50. W. Sakai *et al.*, Secondary mutations as a mechanism of cisplatin resistance in BRCA2-mutated cancers. *Nature* **451**, 1116-1120 (2008).

51. J. P. Eder, C. A. Wheeler, B. A. Teicher, L. E. Schnipper, A phase I clinical trial of novobiocin, a modulator of alkylating agent cytotoxicity. *Cancer Res* **51**, 510-513 (1991).

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.23.111658; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ACKNOWLEDGEMENTS

We thank the ICCB-Longwood Screening Facility at Harvard Medical School, for their help in small-molecule screening. We thank Prafulla Gokhale and Qing Zeng at Belfer Center for Applied Cancer Science for their help in PDX studies. We thank Connor Clairmont for providing RPE *BRCA1*$^{-/-}$ cells and Lisa Moreau for her help in chromosome aberration assays. We thank Drs. Huy Nguyen and Sandor Spisak for genomic analysis of the *POLθ* Knockout clones. We thank Dr. Jeremy Stark at Beckman Research Institute of the City of Hope for providing DR-GFP and EJ-2 repair substrates.

FUNDING

This research was supported by a Stand Up To Cancer (SU2C)-Ovarian Cancer Research Fund Alliance-National Ovarian Cancer Coalition Dream Team Translational Research Grant (SU2C-AACR-DT16-15) to A.D.D. SU2C is a program of the Entertainment Industry Foundation. Research grants are administered by the American Association for Cancer Research, the scientific partner of SU2C. This work was also supported by the NIH (grants R37HL052725, P01HL048546, P50CA168504), the US Department of Defense (grants BM110181 and BC151331P1), as well as grants from the Breast Cancer Research Foundation and the Fanconi Anemia Research Fund to A.D.D., as well as the ERC starting grant (N° 714162) and the Ville de Paris Emergences Program grant (N° DAE 137) to R.C. This work was also supported by Ann Schreiber Mentored Investigator Award from Ovarian Cancer Research Fund Alliance (457527) and Joint Center for Radiation Therapy Award from Harvard Medical School to J.Z. J.A.T. and A.S. are supported by NIH grants P01 CA092548, R35 CA220430, and by the Cancer Prevention and Research Institute of Texas and the Robert A. Welch Chemistry Chair.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.23.111658; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

COMPETING INTERESTS

A.D. D'Andrea reports receiving commercial research grants from Eli Lilly & Company, Sierra Oncology, and EMD Serono and is a consultant/advisory board member for Eli Lilly & Company, Sierra Oncology, and EMD Serono. G.I. Shapiro has received research funding from Eli Lilly, Merck KGaA/EMD-Serono, Merck, and Sierra Oncology, and he has served on advisory boards for Pfizer, Eli Lilly, G1 Therapeutics, Roche, Merck KGaA/EMD-Serono, Sierra Oncology, Bicycle Therapeutics, Artios, Fusion Pharmaceuticals, Cybrexa Therapeutics, Astex, Almac, Ipsen, Bayer, Angiex, Daiichi Sankyo, Boehringer Ingelheim, ImmunoMet and Asana.

DATA AND MATERIALS AVAILABILITY

All related data have been presented in the main manuscript or in supplementary data.

SUPPLEMENTARY MATERIALS

Figs. S1 to S10

Legends for Figs. S1 to S10

Materials and Methods

Supplementary References

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.23.111658; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

FIGURE LEGENDS

Figure 1. A small-molecule screen identifies Novobiocin (NVB) as a specific POLθ ATPase inhibitor. A, The domain structures of full length (FL) and ATPase domain (or ΔPol, a.a. 1-987) of POLθ, a Coomassie-stained gel of the purified POLθ ATPase domain, and a schematic of the small molecule screen for POLθ inhibitors. **B**, Results of the small-molecule screen. Shown is quantification of POLθ ATPase activity measured by the ADP-Glo assay in the presence of small-molecule libraries. We screened a total of 23,883 small-molecules in duplicate across multiple libraries enriched with known bioactive compounds, and we identified 72 compounds (0.3% of total) that significantly reduced POLθ ATPase activity (*z*-score < -4). Four top hits verified in the secondary screen (Fig. S1B-E) were labeled; Mitoxantrone (MTX), Suramin (SUR), Novobiocin (NVB) and Aurintricarboxylic acid (AUR). Data shown are mean ± s.e.m, n = 2. **C**, Quantification of POLθ and SMARCAL1 ATPase activity in the presence of indicated small-molecules. Mean ± s.d., n = 3. **D**, Pulldown experiments *in vitro* with NVB-conjugated beads and purified POLθ and SMARCAL1 ATPase domains (Flag-tag purification from SF9 cells). Input and pulldown proteins were detected by Coomassie staining on SDS gel. **E**, Thermal shift assay with NVB at indicated temperatures. Full length GFP-POLθ was detected by Western blot using the GFP antibody and PARP1 by a PARP1 antibody. Data in **D** and **E** are representative images of three independent experiments. **F**, The NVB binding tunnel in POLQ ATPase domain (shown in surface - PDB 5AGA) was predicted by extra precision glide docking and lowest binding free energy from prime MM-GBSA calculations to have multiple hydrogen bonds and close hydrophobic packing. Top and side views showing NVB docking into the tunnel. Binding modes of NVB (green sticks) and AMP-PNP (magenta sticks) in POLQ helicase domain (PDB 5AGA) with green sphere showing active site Mg^{2+} ion. **G**, Representative images and quantification of POLθ accumulation at sites of laser micro-irradiated DNA damage in cells overexpressing GFP-tagged, full-length POLθ in the presence of indicated inhibitors. Mean ± s.e.m, n = 3. **H**, MMEJ repair reporter assay in U2OS cells treated with indicated inhibitors. Mean ± s.d., n = 3. Statistics analysis between groups was performed using *t*-test with Welch's correction. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ns, not significant.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.23.111658; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 2. NVB kills HR-deficient tumors *in vivo* and *in vitro* and synergizes with PARPi. A, Tumor growth of the GEMM model (*Brca1*$^{-/-}$ TNBC) after treatment with vehicle (PBS) or NVB. Tumor chunks from GEMM mice were implanted in syngeneic FVB/129P2 mice, which were treated with PBS or 100 mg/kg NVB via IP injection twice a day for 5 weeks. n = 6 for PBS group and n = 7 for NVB group. Mean ± s.e.m. is shown. Statistical analyses were *t*-test. **B**, Survival plot of the experiment shown in (**A**). Median survival and *p* value are shown. Statistics analyses were *t*-test. **C**, Tumor growth of the xenograft model with TOV21G (FANCF deficient ovarian cancer) or FANCF-complemented TOV21G cells. The tumor bearing mice were treated with vehicle or 100 mg/kg NVB via IP injection twice a day for 4 weeks. Fold change of tumor was calculated using the formula Fold = T/T0, where T is the tumor size at given time and T0 is the initial tumor size. Data are Mean ± s.e.m., n = 10 in each group. Dunn's multiple comparisons test. Representative images are TOV21G xenograft tumors at the end of NVB treatment. **D,** Clonogenic survival of *BRCA1*$^{-/-}$ and WT RPE1 cells under increasing concentrations of the POLθ inhibitor NVB. Survival is normalized to the untreated (Unt) sample. Data shown are mean ± s.e.m., n = 4. **E**, Representative images and quantification of clonogenic survival of *POLQ*$^{-/-}$ and WT U2OS cells under increasing concentrations of NVB. Survival is shown as relative to the untreated (Unt) sample. Data shown are mean ± SD, n = 5. **F,** Clonogenic survival of *BRCA1*$^{-/-}$ and WT RPE1 cells under increasing concentrations of the PARPi Rucaparib alone or in combination with NVB. Quantifications, representative images and calculated IC50 values are shown. Data shown are mean ± s.e.m., n = 3.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.23.111658; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 3. NVB overrides PARPi resistance *in vivo* in PDX models and RAD51 is a dynamic pharmacodynamic biomarker. A-B, Efficacy of NVB in PDX model DF83 (RAD51C deficient). NSG mice (n = 7 mice/group) bearing luciferized DF83 cells derived from an ovarian cancer patient were treated with 50 mg/kg of olaparib (daily, orally), 75 mg/kg NVB (via IP, b.i.d.), or both for 4 weeks. Tumor growth was monitored weekly by bioluminescence imaging. Representative images of tumor burden on day 0 and day 28 are shown, and quantifications shown are mean ± s.e.m. Log_2(fold change in tumor size) was calculated using the formula log_2(T/T0), where T is the tumor volume at given time and T0 is the initial tumor volume. **C-D**, Efficacy of NVB in PARPi resistant PDX model DF59 (BRCA1-deficient, 53BP1 loss), n = 6 mice/group. Same conditions and analysis were applied as in **A-B**. **E-F**, Efficacy of NVB in PARPi resistant PDX model DF149 (BRCA1-WT, HR proficient), n = 7 mice/group. Same conditions and analysis were applied as in **A-B**. **G**, Representative IHC images and quantifications of RAD51 foci in DF-59 and DF-83 PDX models. Tumor bearing mice were dosed with indicated drugs and as in efficacy study in Fig 3, and tumor cells were collected from ascites 52 hours after first dose. FFPE sections of the isolated tumor cells were stained with a RAD51 antibody. The percentage of RAD51-foci positive cells (> 4 foci/cell) in 3 random 40X fields was estimated in DF-59. No RAD51 focus was observed in the DF-83 model. Statistical analysis method was one-way ANOVA Bonferroni's multiple comparisons test. *, $p < 0.05$; **, $p < 0.01$; ****, $p < 0.0001$. **H**. POLQ expression at mRNA level in PDX models DF53, DF83 and DF149.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.23.111658; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 4. POLQ expression level is a predictive biomarker for NVB sensitivity. A. A schematic shows how the PARPi resistant clones R1-R4 were generated. TP53 was knocked out in order to generate *BRCA1* knockout (*BRCA1*$^{-/-}$) RPE1 cells, which were then used for selection of PARPi resistant clones. **B**. A Western blot of WT and *BRCA1*$^{-/-}$ RPE1 cells and the PARPi resistant clones, using an anti-BRCA1 antibody (Millipore #OP92). No BRCA1 reversion was observed in the clones. **C-D.** NVB sensitivity of WT RPE1, *BRCA1*$^{-/-}$ RPE1, and PARPi-resistant *BRCA1*$^{-/-}$ clones (R1-R4). R1 was selected by olaparib and R2-R4 were selected by Niraparib. R1 and R2 were tested in clonogenic survival assay **(C)**. R3 and R4 were tested in CellTiter-Glo cell viability assay (**D**). Data are mean ± s.e.m., n = 3. **E**. POLQ expression at mRNA level and protein level in WT RPE1, *BRCA1*$^{-/-}$ RPE1, and PARPi-resistant *BRCA1*$^{-/-}$ clones (R1-R4). POLQ mRNA was measured by qRT-PCR and normalized to beta-Actin (ACTB). **F**. POLQ expression at mRNA level in parental (P), PARPi resistant (R), and empty vector (+EV) or BRCA1-cDNA (+BR1) complemented MDA-MB-436 cells. **G**, RAD51 foci assay with DMSO or NVB in RPE1-*BRCA1* WT RPE1, *BRCA1*$^{-/-}$ RPE1, and PARPi-resistant *BRCA1*$^{-/-}$ clones (R1-R4) 4 hours after 5 Gy of IR. Cells with more than 10 foci were counted as positive. Data were Mean ± SEM from two independent experiments. **H**, RAD51 foci assay with DMSO or NVB in MDA-MB-436 or the PARPi-resistant MDA-MB-436-R. Cells with more than 5 foci were counted as positive. Data were Mean ± SEM from two independent experiments. **I.** Novobiocin sensitivity of RPE1-WT and the R1 PARPi-resistant clone after being treated with Control or RAD51 siRNA. A Western blot shows the efficiency of RAD51 knockdown in RPE1-WT and the R1 PARPi-resistant clone.

Fig 1

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.23.111658; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

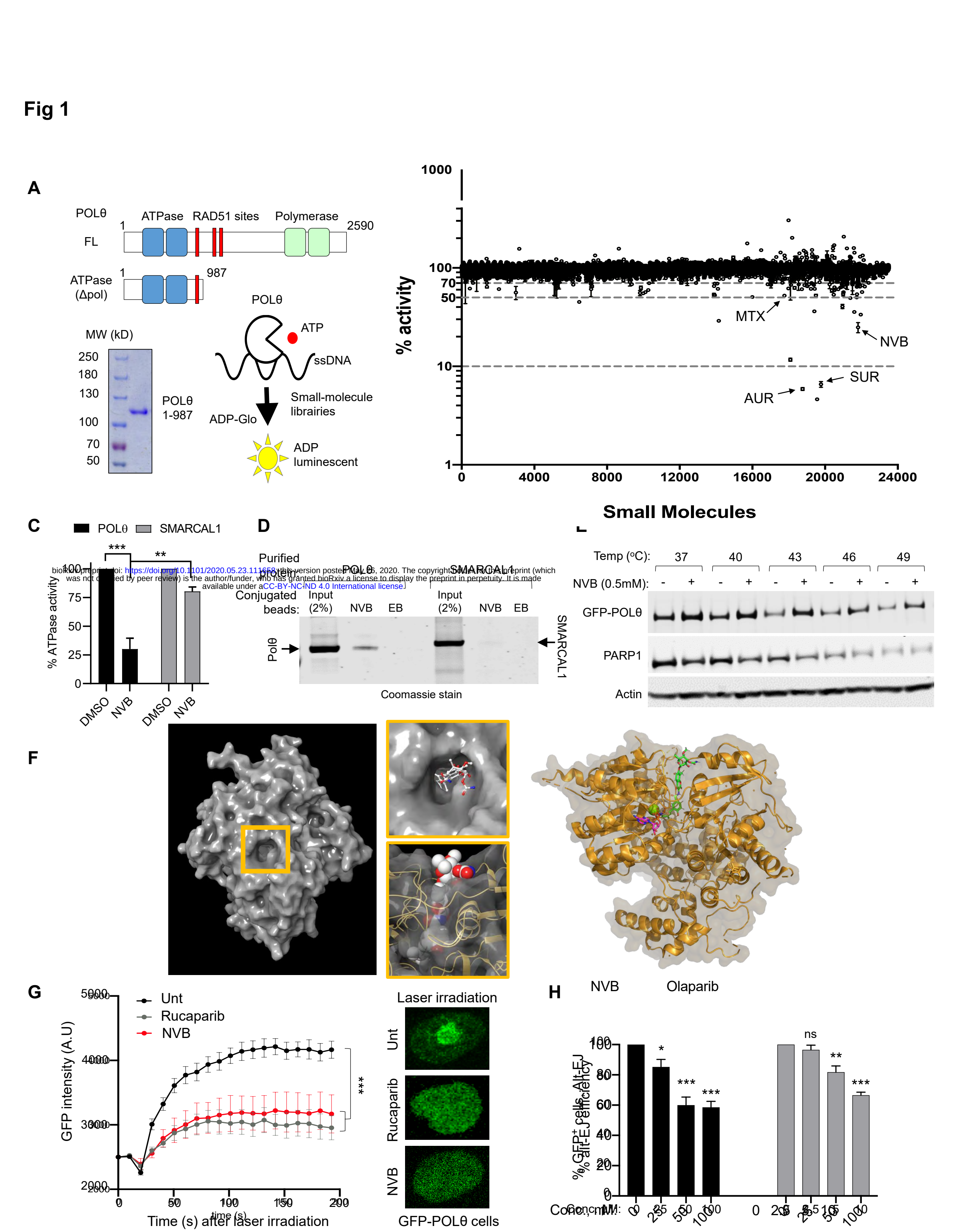
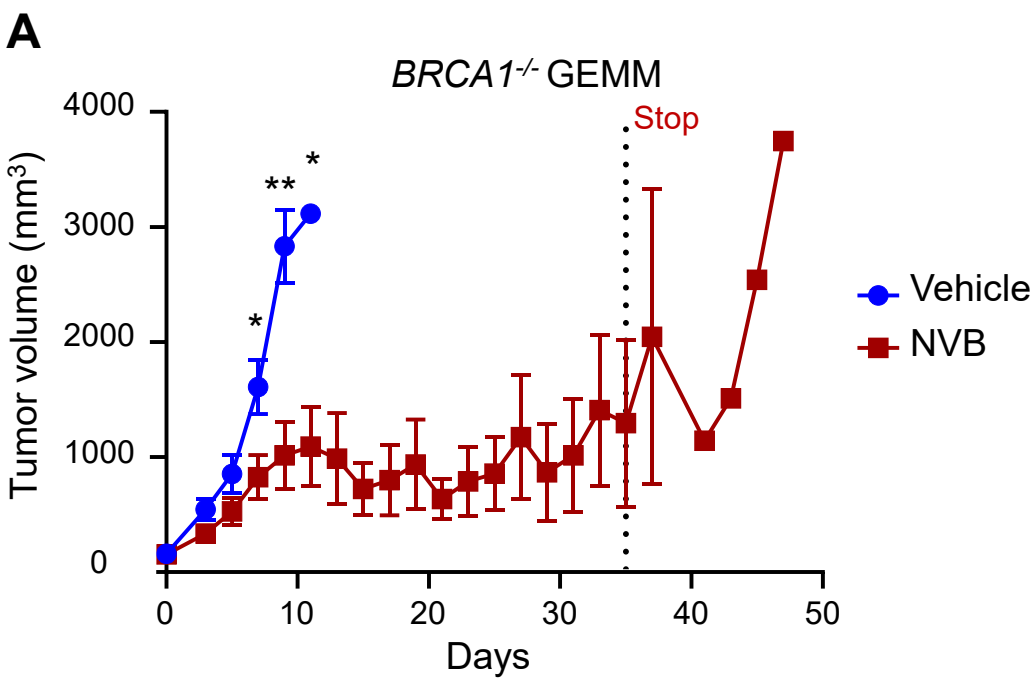


Fig 2



bioRxiv preprint doi: https://doi.org/10.1101/2020.05.23.111658; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

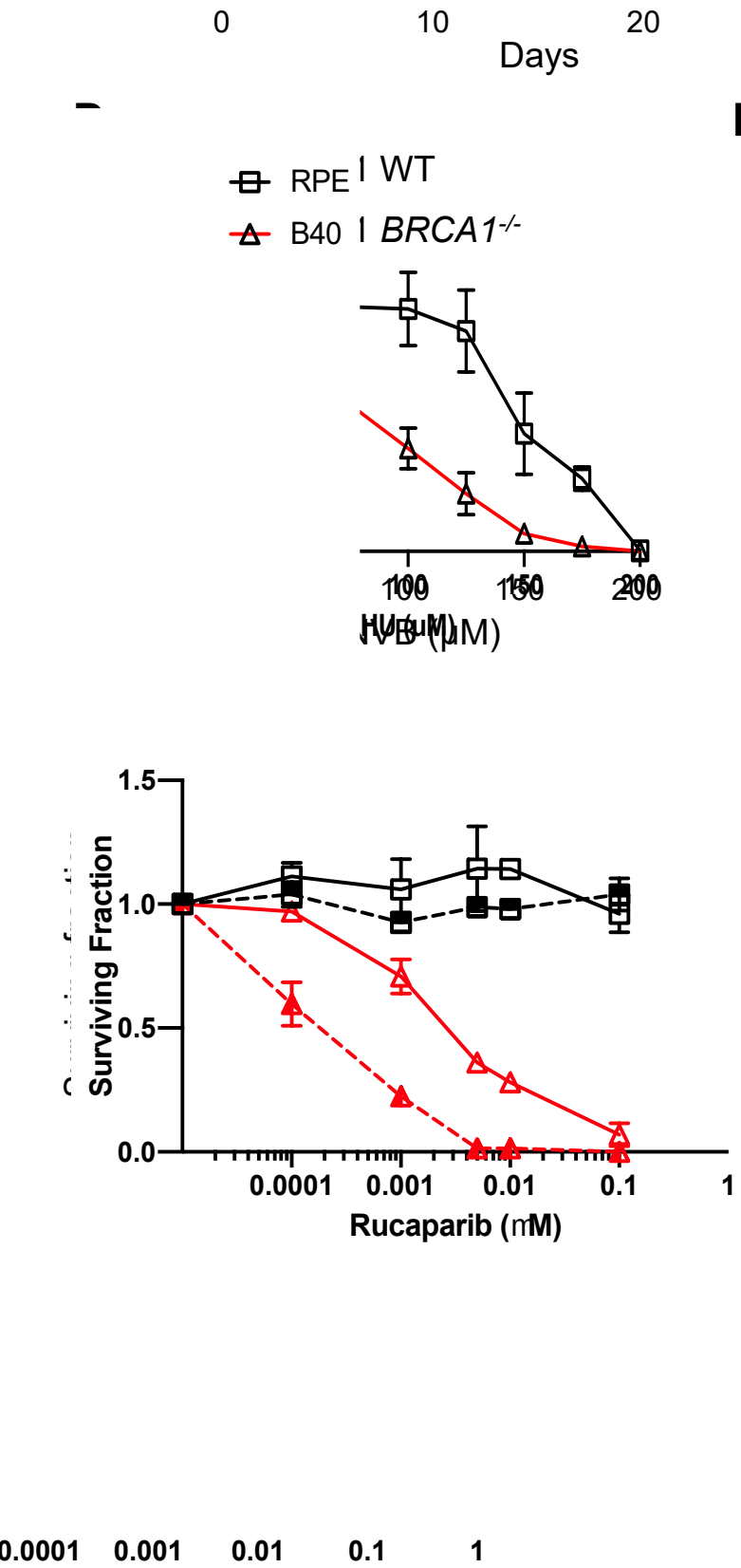
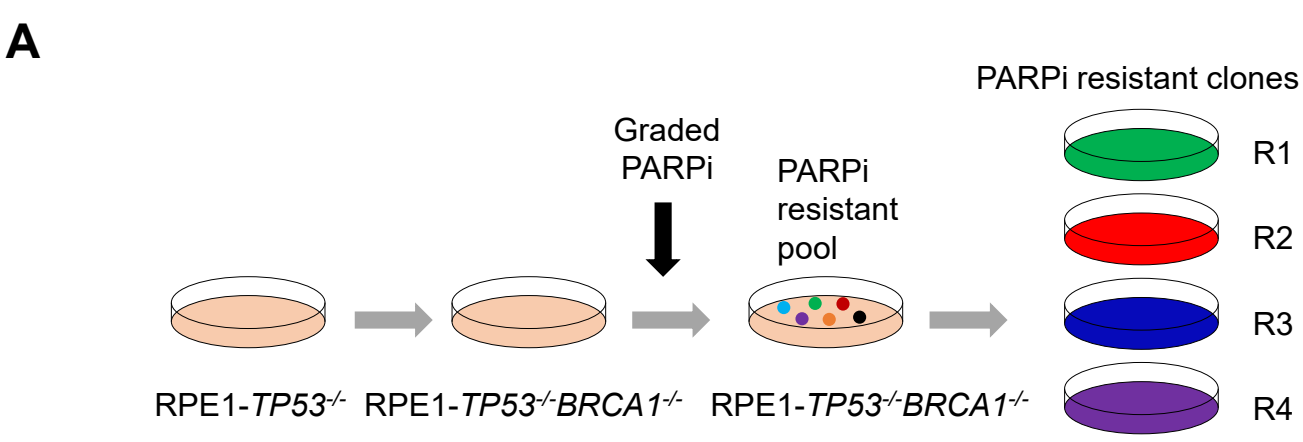
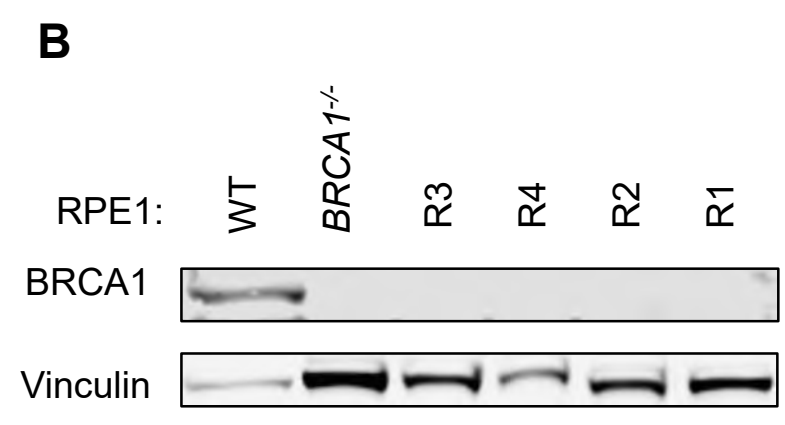


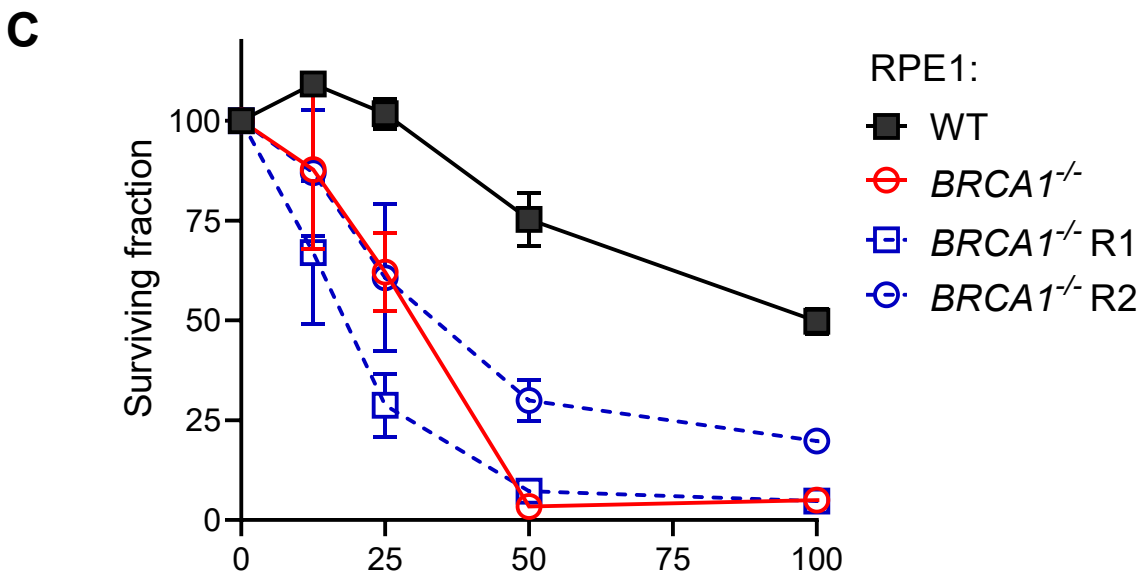
Fig 3

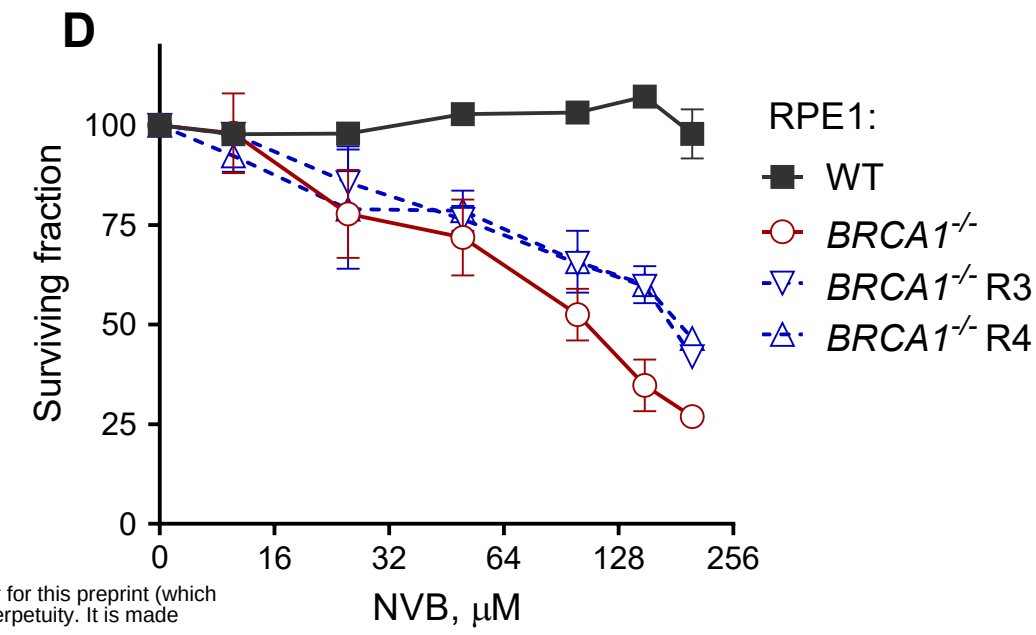
bioRxiv preprint doi: https://doi.org/10.1101/2020.05.23.111658; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fig 4









bioRxiv preprint doi: https://doi.org/10.1101/2020.05.23.111658; this version posted May 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

