

**AWARD NUMBER:** W81XWH-18-1-0384

**TITLE:** Mechanistic and Therapeutic Implications of Spliceosomal Gene Mutations in ER+ Breast Cancer

**PRINCIPAL INVESTIGATOR:** Sarat Chandarlapaty

**CONTRACTING ORGANIZATION:** Sloan Kettering Institute for Cancer Research

**REPORT DATE:** DECEMBER 2021

**TYPE OF REPORT:** Final Technical

**PREPARED FOR:** U.S. Army Medical Research and Development 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 22202-4302. 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.**

<b>1. REPORT DATE</b> DECEMBER 2021			<b>2. REPORT TYPE</b> Final		<b>3. DATES COVERED</b> 15 Aug 2018 - 14 Aug 2021	
<b>4. TITLE AND SUBTITLE</b>  Mechanistic and Therapeutic Implications of Spliceosomal Gene Mutations in ER+ Breast Cancer					<b>5a. CONTRACT NUMBER</b>	
					<b>5b. GRANT NUMBER</b> W81XWH-18-1-0384	
					<b>5c. PROGRAM ELEMENT NUMBER</b>	
<b>6. AUTHOR(S)</b>  Omar Abdel-Wahab; Sarat Chandarlapaty  E-Mail: <a href="mailto:abdelwao@mskcc.org">abdelwao@mskcc.org</a> ; <a href="mailto:chandars@mskcc.org">chandars@mskcc.org</a>					<b>5d. PROJECT NUMBER</b>	
					<b>5e. TASK NUMBER</b>	
					<b>5f. WORK UNIT NUMBER</b>	
<b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b>  Sloan Kettering Institute for Cancer Research 1275 York Avenue New York, NY 10065					<b>8. PERFORMING ORGANIZATION REPORT NUMBER</b>	
<b>9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b>  U.S. Army Medical Research and Development Command Fort Detrick, Maryland 21702-5012					<b>10. SPONSOR/MONITOR'S ACRONYM(S)</b>	
					<b>11. SPONSOR/MONITOR'S REPORT NUMBER(S)</b>	
<b>12. DISTRIBUTION / AVAILABILITY STATEMENT</b>  Approved for Public Release; Distribution Unlimited						
<b>13. SUPPLEMENTARY NOTES</b>						
<b>14. ABSTRACT</b> Recent evidence has revealed that altered mRNA splicing is a mechanism by which tumors can derive constitutive, tumor-promoting signals. Recurrent, somatic mutations in the core RNA splicing factor SF3B1 have been found in several malignancies. Through analyses of metastatic breast cancer patients at our center, we have noted hotspot mutations in SF3B1 in up to 6% of cases and these are associated with the ER+/HER2- subtype and inferior outcomes in patients. We therefore conducted studies to understand the potential implications of SF3B1 mutations on breast cancer pathogenesis and found that expression of the most common mutant, K700E, led to alterations in RNA splicing and promoted hormone-independent proliferation of ER+ cells. Based on these initial observations, we are now studying the consequences of SF3B1 mutation on breast tumorigenesis, cancer progression, and sensitivity to spliceosomal targeted therapy.						
<b>15. SUBJECT TERMS</b> Antisense oligonucleotides, breast cancer, estrogen receptor, E7107, E7820, PIK3CA, RBM39, RNA splicing, SF3B1.						
<b>16. SECURITY CLASSIFICATION OF:</b>			<b>17. LIMITATION OF ABSTRACT</b>	<b>18. NUMBER OF PAGES</b>	<b>19a. NAME OF RESPONSIBLE PERSON</b> USAMRDC	
<b>a. REPORT</b>	<b>b. ABSTRACT</b>	<b>c. THIS PAGE</b>			<b>19b. TELEPHONE NUMBER</b> (include area code)	
Unclassified	Unclassified	Unclassified	Unclassified	13		

## TABLE OF CONTENTS

	<u>Page</u>
1. Introduction	4
2. Keywords	4
3. Accomplishments	4
4. Impact	7
5. Changes/Problems	8
6. Products	9
7. Participants & Other Collaborating Organizations	11
8. Special Reporting Requirements	14
9. Appendices	14

**1. INTRODUCTION:** *Narrative that briefly (one paragraph) describes the subject, purpose and scope of the research.*

Recurrent, hotspot mutations in the RNA splicing factor SF3B1 occur in estrogen receptor positive (ER+) metastatic breast cancer (MBC). Here we aimed to (1) determine the mechanisms by which the SF3B1 K700E mutation promotes hormone-independent growth of breast cancer, (2) examine the effect of SF3B1 K700E mutation on mammary tumorigenesis and develop novel SF3B1 mutant-selective splicing reporters, and (3) determine the therapeutic benefit of modulating RNA splicing in genetically defined breast cancers with and without spliceosomal gene mutations.

**2. KEYWORDS:** *Provide a brief list of keywords (limit to 20 words).*

Antisense oligonucleotides, breast cancer, estrogen receptor, E7107, E7820, PIK3CA, RBM39, RNA splicing, SF3B1

**3. ACCOMPLISHMENTS:** *The PI is reminded that the recipient organization is required to obtain prior written approval from the awarding agency grants official whenever there are significant changes in the project or its direction.*

**What were the major goals of the project?**

*List the major goals of the project as stated in the approved SOW. If the application listed milestones/target dates for important activities or phases of the project, identify these dates and show actual completion dates or the percentage of completion.*

**Major Task 1:** Determine if SF3B1 K700E breast cancer cells depend on ER and/or ER signaling by testing the response of SF3B1 K700E models to inhibitors of ER or of its key targets (e.g. cyclin D1).

**Milestone(s):** Define the effects of SF3B1 K700E mutations on ER dependence and signaling as well as gene expression and splicing. Also, identify the global effects of SF3B1 K700E on splicing and gene expression in the context of breast cancer as well as those splicing events critical for estrogen-independent growth.

**Timeline:** Months 1-12. 100% complete.

**Major Task 2:** Examine the effect of *Sf3b1*K700E mutation on mammary tumorigenesis *in vivo*.

**Milestone(s):** Determine the effects of mutant *Sf3b1* as well as mutant *Pik3ca*, alone and together, on splicing and gene expression in mammary breast epithelial cells as well the tumor promoting potential for mutant SF3B1 *in vivo*.

**Timeline:** Months 1-36. 100% complete.

**Major Task 3:** Determine the therapeutic benefit of modulating RNA splicing in breast cancers with spliceosomal gene mutations.

**Milestone(s):** Determine if breast cancer cells with SF3B1 mutations and/or MYC overexpression are preferentially sensitive to H3B-8800 or E7070 and E7820 as well as the mechanistic basis for their sensitivity.

**Timeline:** Months 1-36. 100% complete.

## What was accomplished under these goals?

For this reporting period describe: 1) major activities; 2) specific objectives; 3) significant results or key outcomes, including major findings, developments, or conclusions (both positive and negative); and/or 4) other achievements. Include a discussion of stated goals not met. Description shall include pertinent data and graphs in sufficient detail to explain any significant results achieved. A succinct description of the methodology used shall be provided. As the project progresses to completion, the emphasis in reporting in this section should shift from reporting activities to reporting accomplishments.

### Major Task 1

Subtask 1. Generate additional tetracycline inducible models of SF3B1 K700E (or WT) in MCF7 or CAMA-1 ER+ models. (OAW and SC)

We have generated tetracycline inducible models of SF3B1 WT or K700E in MCF7 breast cancer cell lines. Western blots showed expression of WT or K700E at similar abundance induced by doxycycline treatment similar to what was observed in T47D cells (Figure 6D & Supp Fig S5 Liu et al, JCI 2021).

Subtask 2. Knock the *SF3B1* K700E allele into MCF7 using CRISPR-Cas9 (SC)

We successfully generated knockin of the K700E allele into MCF7 cells. After serial passage, the allele was lost. The same scenario occurred upon a repeated cloning effort – initial knockin of SF3B1K700E that was lost after serial passage. Based on this, we generated MCF10A cell lines with and without expression of SF3B1 K700E mutation as follows and part of our publication (Figure 8, Liu et al, JCI 2021):

-MCF10A cells with heterozygous knockin of SF3B1K700E allele.

-MCF10A cells with heterozygous knockin of PIK3CA H1047R allele.

-MCF10A cells with dual heterozygous knockin of SF3B1K700E and PIK3CA H1047R alleles.

These reagents were also distributed to other investigators seeking to further study *SF3B1* mutant breast cancer.

Subtask 3. Characterize the dependence of the mutant versus WT expressing models on hormones, ER level, and levels of key ER target genes as well as sensitivity to combination endocrine regimens. (SC)

We found no significant difference in the expression of ER and its key target genes in MCF7 and T47D cells expressing mutant SF3B1 by western blot (Supp Fig S5 Liu et al, JCI 2021). Moreover, mutant expressing cells remained dependent on estrogen as they did not grow in media supplemented with 10% charcoal stripped serum (css) in which hormones were depleted (Supp Fig S5 Liu et al, JCI 2021). Similarly, the mutant versus wild type models were equally sensitive to ER antagonism with fulvestrant or tamoxifen inhibiting growth at equivalent IC50s and GR50s ((Supp Fig S5 Liu et al, JCI 2021).

Subtask 4. Perform and analyze mRNA sequencing (RNA-seq) of T47-D, MCF-7, and CAMA-1 cells with inducible expressing vector, SF3B1 WT, and SF3B1 K700E (equal expression to endogenous WT) to characterize gene expression as well as RNA splicing events. (OAW)

We have performed RNA-seq analysis on MCF10A (WT or SF3B1 K700E knock-in), MCF7 and T47D doxycycline-inducible cells (Figure 4&6, Liu et al, JCI 2021). The expression of mutant SF3B1 caused extensive usage of alternative 3' splice sites that are conserved in all three cell lines. For differential gene expression analysis, the major variance was dominated by cell lines. Only a small fraction of genes were commonly regulated in all three cell lines (Figure 4, Liu et al, JCI 2021).

Interestingly, this list includes genes that are alternatively spliced, such as MAP3K7 and PPP2R5A. We investigated the consequences of MAP3K7 and PPP2R5A extensively and found them to be key mediators of invasiveness of the SF3B1 mutant breast tumors (Figure 7, Liu et al, JCI 2021).

## Major Task 2

Subtask 1 Assess *Mmtv*-cre *Sf3b1*<sup>K700E/WT</sup>, *Mmtv*-cre *R26R-LSL-Pik3ca*<sup>H1047R</sup>, *Mmtv*-cre *Sf3b1*<sup>K700E/WT</sup> *R26R-LSL-Pik3ca*<sup>H1047R</sup>, along with WT controls for tumor development. (OAW and SC)

We have generated four group of mice: *Mmtv*-cre, *Mmtv*-cre *Sf3b1*<sup>K700E/WT</sup>, *Mmtv*-cre *R26R-LSL-Pik3ca*<sup>H1047R</sup> and *Mmtv*-cre *Sf3b1*<sup>K700E/WT</sup> *R26R-LSL-Pik3ca*<sup>H1047R</sup> double mutant. While extensive growth of mammary ducts was found in *Pik3ca* single mutant and *Sf3b1*<sup>K700E</sup> *Pik3ca* double mutant mice, the *Sf3b1* K700E single mutant mice showed reduced mammary ductal branching. Hastened death was observed in *MMTV-Cre Sf3b1*<sup>K700E/WT</sup> *R26-Pik3ca*<sup>H1047R</sup> double mutant mice, where ER<sup>+</sup>/PR<sup>+</sup>, high Ki67 mammary gland adenocarcinoma developed (Figure 2, Liu et al. JCI 2021).

Subtask 3. We have now performed and analyzed RNA-seq of purified luminal (DAPI-CD24<sub>low</sub>) and basal (DAPI-CD24<sub>high</sub>) epithelial cells from the mouse models above (Supp Figure 2 and Figure 4). This has revealed the same global change in aberrant, intron-proximal 3' splice site usage in SF3B1 mouse epithelial cells and breast tumors as was seen in SF3B1 mutant breast cancer cell lines. This prominently includes mis-splicing of *Map3k7* and *Ppp2r5a*, as seen in human breast cancer cells with mutant SF3B1.

## Major Task 3

Subtask 1. Test the effects of 2 different inhibitors of SF3b complex in a panel of ER dependent cell lines isogenic for SF3B1<sup>K700E</sup> or WT. (OAW and SC)

We have tested the effect of SF3B1 inhibitor E7107 on in vitro proliferation in MCF10A and T47D cells expressing mutant SF3B1 and no significant difference was observed when compared to WT control cells. E7107 is very potent and toxic to all these cells under in vitro culture condition with IC<sub>50</sub> less than 0.3nM. Based upon the findings in Major Task 1, we identified AKT inhibition as potential pharmacologic strategy for targeting dual SF3B1 and PIK3CA mutant cancers and have found models with both alterations to be hypersensitive to AKT inhibitors (Figure 8, Liu et al, JCI 2020).

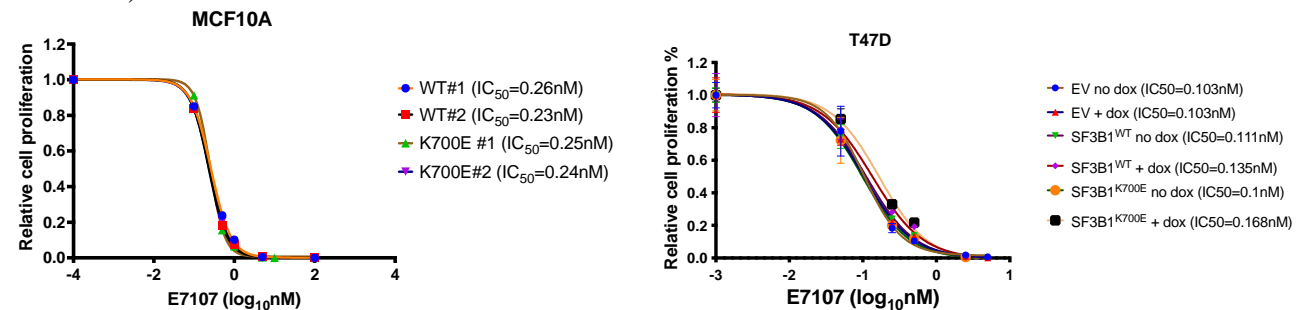
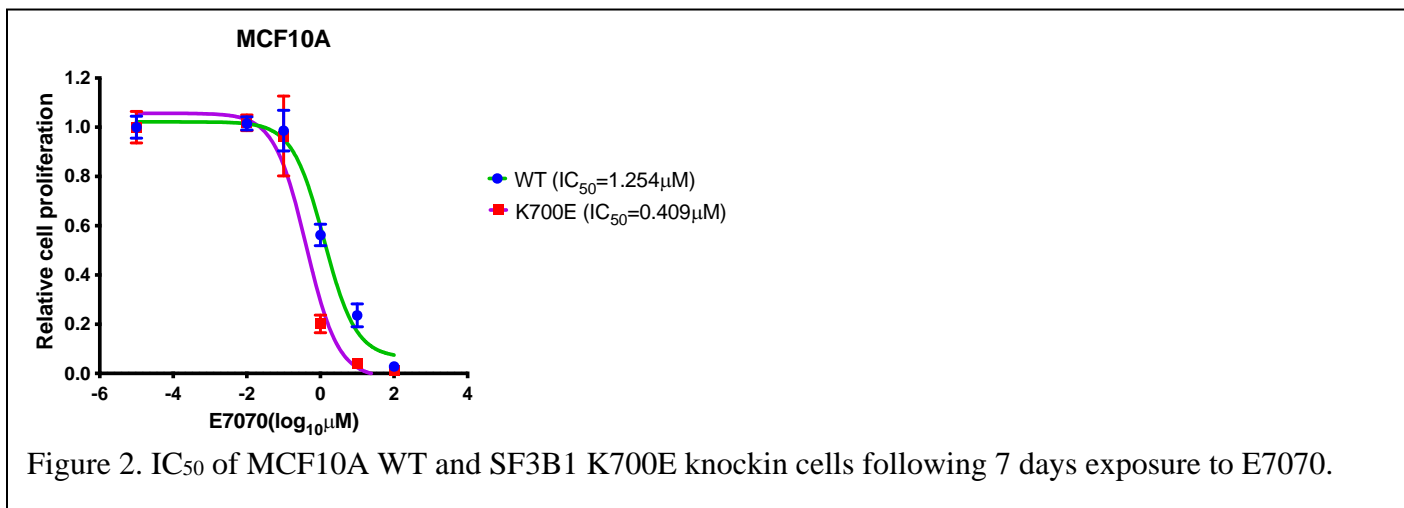


Figure 1. IC<sub>50</sub> of MCF10A WT and SF3B1 K700E knockin cells (left) or T47D cells expressing empty vector (EV), SF3B1 WT or SF3B1 K700E cDNA (right) following 7 days exposure to E7107.

Subtask 2: Test the effects of 2 different inhibitors of SF3b complex in a panel of genetically annotated ER/PR<sup>+</sup> and TNBC cell lines defined by MYC expression status (OAW and SC)

Subtask 4: Evaluate the sensitivity of human and murine breast cancer cells with and without an SF3B1 mutation to pharmacologic or genetic loss of the accessory splicing factor RBM39 (OAW and SC)

We have tested the effect of E7070, a sulfonamide inducing RBM39 degradation, on MCF10A WT or SF3B1 K700E knock-in cells. SF3B1 mutant cells were more sensitive to E7070 treatment in proliferation experiment with lower IC<sub>50</sub>.



### What opportunities for training and professional development has the project provided?

*If the project was not intended to provide training and professional development opportunities or there is nothing significant to report during this reporting period, state “Nothing to Report.”*

*Describe opportunities for training and professional development provided to anyone who worked on the project or anyone who was involved in the activities supported by the project. “Training” activities are those in which individuals with advanced professional skills and experience assist others in attaining greater proficiency. Training activities may include, for example, courses or one-on-one work with a mentor. “Professional development” activities result in increased knowledge or skill in one’s area of expertise and may include workshops, conferences, seminars, study groups, and individual study. Include participation in conferences, workshops, and seminars not listed under major activities.*

Dr. Bo Liu (shared postdoctoral fellow of the Abdel-Wahab and Chandarlapaty labs) presented his work at the 2019 RNA Processing in Cancer meeting at Yale University and at a large Departmental seminar series in 2020

### How were the results disseminated to communities of interest?

*If there is nothing significant to report during this reporting period, state “Nothing to Report.”*

*Describe how the results were disseminated to communities of interest. Include any outreach activities that were undertaken to reach members of communities who are not usually aware of these project activities, for the purpose of enhancing public understanding and increasing interest in learning and careers in science, technology, and the humanities.*

Nothing additional to report.

### What do you plan to do during the next reporting period to accomplish the goals?

*If this is the final report, state “Nothing to Report.”*

*Describe briefly what you plan to do during the next reporting period to accomplish the goals and objectives.*

None

4. **IMPACT:** Describe distinctive contributions, major accomplishments, innovations, successes, or any change in practice or behavior that has come about as a result of the project relative to:

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

If there is nothing significant to report during this reporting period, state “Nothing to Report.”

Describe how findings, results, techniques that were developed or extended, or other products from the project made an impact or are likely to make an impact on the base of knowledge, theory, and research in the principal disciplinary field(s) of the project. Summarize using language that an intelligent lay audience can understand (Scientific American style).

One of the main findings from our work thus far is that cells which contains mutations in genes that regulate RNA splicing can cause breast cancer cells to be come more invasive and lethal. Part of this effect appears to be through a common growth pathway known as the PI3K/AKT pathway and so these tumors may be treated with AKT inhibitors. We have also identified drugs which alter splicing by inhibiting arginine methylation of splicing proteins (by enzymes known as PRMTs) may also target splicing factor mutant cells. Inhibitors of PRMT5 are currently in phase I clinical trials in solid tumors and our data suggest that mutations in RNA splicing factors (such as SF3B1) may serve as a biomarker for responsiveness to these therapies.

**What was the impact on other disciplines?**

If there is nothing significant to report during this reporting period, state “Nothing to Report.”

Describe how the findings, results, or techniques that were developed or improved, or other products from the project made an impact or are likely to make an impact on other disciplines.

Nothing to report.

**What was the impact on technology transfer?**

If there is nothing significant to report during this reporting period, state “Nothing to Report.”

Describe ways in which the project made an impact, or is likely to make an impact, on commercial technology or public use, including:

- transfer of results to entities in government or industry;
- instances where the research has led to the initiation of a start-up company; or
- adoption of new practices.

Nothing to Report

**What was the impact on society beyond science and technology?**

If there is nothing significant to report during this reporting period, state “Nothing to Report.”

Describe how results from the project made an impact, or are likely to make an impact, beyond the bounds of science, engineering, and the academic world on areas such as:

- *improving public knowledge, attitudes, skills, and abilities;*
- *changing behavior, practices, decision making, policies (including regulatory policies), or social actions; or*
- *improving social, economic, civic, or environmental conditions.*

*Nothing to Report*

- 5. CHANGES/PROBLEMS:** *The PD/PI is reminded that the recipient organization is required to obtain prior written approval from the awarding agency grants official whenever there are significant changes in the project or its direction. If not previously reported in writing, provide the following additional information or state, "Nothing to Report," if applicable:*

**Changes in approach and reasons for change**

*Describe any changes in approach during the reporting period and reasons for these changes. Remember that significant changes in objectives and scope require prior approval of the agency.*

*Nothing to Report*

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

*Describe problems or delays encountered during the reporting period and actions or plans to resolve them.*

*Nothing to Report*

**Changes that had a significant impact on expenditures**

*Describe changes during the reporting period that may have had a significant impact on expenditures, for example, delays in hiring staff or favorable developments that enable meeting objectives at less cost than anticipated.*

*None.*

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

*Describe significant deviations, unexpected outcomes, or changes in approved protocols for the use or care of human subjects, vertebrate animals, biohazards, and/or select agents during the reporting period. If required, were these changes approved by the applicable institution committee (or equivalent) and reported to the agency? Also specify the applicable Institutional Review Board/Institutional Animal Care and Use Committee approval dates.*

**Significant changes in use or care of human subjects**

*None.*

**Significant changes in use or care of vertebrate animals**

*None.*

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

None

**6. PRODUCTS:** *List any products resulting from the project during the reporting period. If there is nothing to report under a particular item, state “Nothing to Report.”*

- **Publications, conference papers, and presentations**

*Report only the major publication(s) resulting from the work under this award.*

**Journal publications.** *List peer-reviewed articles or papers appearing in scientific, technical, or professional journals. Identify for each publication: Author(s); title; journal; volume: year; page numbers; status of publication (published; accepted, awaiting publication; submitted, under review; other); acknowledgement of federal support (yes/no).*

Bo Liu, Zhaoqi Liu, Michelle Ki, Caroline Erickson, Jorge Reis-Filho, Raul Rabadan, Omar Abdel-Wahab (co-corresponding), Sarat Chandarlapaty (co-corresponding). (2021) Mutant SF3B1 promotes AKT and NF- $\kappa$ B driven mammary tumorigenesis. *Journal of Clinical Investigation* Jan4;131(1):e138315. PMID: PMC7773370. Federal support acknowledged.

**Books or other non-periodical, one-time publications.** *Report any book, monograph, dissertation, abstract, or the like published as or in a separate publication, rather than a periodical or series. Include any significant publication in the proceedings of a one-time conference or in the report of a one-time study, commission, or the like. Identify for each one-time publication: author(s); title; editor; title of collection, if applicable; bibliographic information; year; type of publication (e.g., book, thesis or dissertation); status of publication (published; accepted, awaiting publication; submitted, under review; other); acknowledgement of federal support (yes/no).*

None.

**Other publications, conference papers and presentations.** *Identify any other publications, conference papers and/or presentations not reported above. Specify the status of the publication as noted above. List presentations made during the last year (international, national, local societies, military meetings, etc.). Use an asterisk (\*) if presentation produced a manuscript.*

2019 RNA Processing in Cancer Presentation entitled “Role of SF3B1 Mutations in Breast Cancer”, Yale University, New Haven, CT.

2019 AACR Annual Meeting Presentation entitled “Hotspot mutations in the core spliceosomal protein SF3B1 promote breast tumorigenesis”, Atlanta, GA.

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

*List the URL for any Internet site(s) that disseminates the results of the research activities. A short description of each site should be provided. It is not necessary to include the publications already specified above in this section.*

None.

- **Technologies or techniques**

*Identify technologies or techniques that resulted from the research activities. Describe the technologies or techniques were shared.*

We have generated a novel technique to identify aberrant 3' splice site usage from RNA-seq data. In addition, we have performed Morpholino based anti-sense oligonucleotides to modulate RNA splicing in breast cancer cells.

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

*Identify inventions, patent applications with date, and/or licenses that have resulted from the research. Submission of this information as part of an interim research performance progress report is not a substitute for any other invention reporting required under the terms and conditions of an award.*

None.

- **Other Products**

*Identify any other reportable outcomes that were developed under this project. Reportable outcomes are defined as a research result that is or relates to a product, scientific advance, or research tool that makes a meaningful contribution toward the understanding, prevention, diagnosis, prognosis, treatment and /or rehabilitation of a disease, injury or condition, or to improve the quality of life.*

We have generated substantial deep RNA-seq data on isogenic breast cancer cell lines with and without mutations in SF3B1. These include MCF10, MCF7, and T47D cells. These data will be deposited in GEO at the time of publication of the manuscript describing our splicing and gene expression analyses of these data.

## 7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

### What individuals have worked on the project?

<i>Name:</i>	<i>Omar Abdel-Wahab</i>
<i>Project Role:</i>	<i>Initiating PI</i>
<i>Researcher Identifier (e.g. ORCID ID):</i>	<i>0000-0002-3907-6171</i>
<i>Nearest person month worked:</i>	<i>1</i>
<i>Contribution to Project:</i>	<i>Overseeing all experiments in conjunction with Dr. Chandarlapaty.</i>
<i>Funding Support:</i>	<i>N/A</i>
<i>Name:</i>	<i>Sarat Chandarlapaty</i>
<i>Project Role:</i>	<i>Partnering PI</i>
<i>Researcher Identifier (e.g. ORCID ID):</i>	<i>0000-0003-4532-8053</i>
<i>Nearest person month worked:</i>	<i>1</i>
<i>Contribution to Project:</i>	<i>Overseeing all experiments in conjunction with Dr. Abdel-Wahab.</i>
<i>Funding Support:</i>	<i>N/A</i>
<i>Name:</i>	<i>Bo Liu</i>
<i>Project Role:</i>	<i>Senior Research Scientist</i>
<i>Researcher Identifier (e.g. ORCID ID):</i>	<i>None</i>
<i>Nearest person month worked:</i>	<i>8</i>
<i>Contribution to Project:</i>	<i>Dr. Liu generated all of the data presented here.</i>
<i>Funding Support:</i>	<i>N/A</i>
<i>Name:</i>	<i>Hana Cho</i>
<i>Project Role:</i>	<i>Research Assistant</i>
<i>Researcher Identifier (e.g. ORCID ID):</i>	<i>None</i>
<i>Nearest person month worked:</i>	<i>1</i>
<i>Contribution to Project:</i>	<i>Ms. Cho assisted Dr. Liu in generating data presented here.</i>
<i>Funding Support:</i>	<i>N/A</i>

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

*If there is nothing significant to report during this reporting period, state “Nothing to Report.”*

*If the active support has changed for the PD/PI(s) or senior/key personnel, then describe what the change has been. Changes may occur, for example, if a previously active grant has closed and/or if a previously pending grant is now active. Annotate this information so it is clear what has changed from the previous submission. Submission of other support information is not necessary for pending changes or for changes in the level of effort for active support reported previously. The awarding agency may require prior written approval if a change in active other support significantly impacts the effort on the project that is the subject of the project report.*

Nothing to report.

### What other organizations were involved as partners?

*If there is nothing significant to report during this reporting period, state “Nothing to Report.”*

*Describe partner organizations – academic institutions, other nonprofits, industrial or commercial firms, state or local governments, schools or school systems, or other organizations (foreign or domestic) – that were involved with the project. Partner organizations may have provided financial or in-kind support, supplied facilities or equipment, collaborated in the research, exchanged personnel, or otherwise contributed.*

*Provide the following information for each partnership:*

*Organization Name:*

*Location of Organization: (if foreign location list country)*

*Partner's contribution to the project (identify one or more)*

- *Financial support;*
- *In-kind support (e.g., partner makes software, computers, equipment, etc., available to project staff);*
- *Facilities (e.g., project staff use the partner's facilities for project activities);*
- *Collaboration (e.g., partner's staff work with project staff on the project);*
- *Personnel exchanges (e.g., project staff and/or partner's staff use each other's facilities, work at each other's site); and*
- *Other.*

## **8. SPECIAL REPORTING REQUIREMENTS**

**COLLABORATIVE AWARDS:** *For collaborative awards, independent reports are required from BOTH the Initiating Principal Investigator (PI) and the Collaborating/Partnering PI. A duplicative report is acceptable; however, tasks shall be clearly marked with the responsible PI and research site. A report shall be submitted to <https://ers.amedd.army.mil> for each unique award.*

**QUAD CHARTS:** *If applicable, the Quad Chart (available on <https://www.usamraa.army.mil>) should be updated and submitted with attachments.*

- 9. APPENDICES:** *Attach all appendices that contain information that supplements, clarifies or supports the text. Examples include original copies of journal articles, reprints of manuscripts and abstracts, a curriculum vitae, patent applications, study questionnaires, and surveys, etc.*

# Mutant SF3B1 promotes AKT- and NF- $\kappa$ B-driven mammary tumorigenesis

Bo Liu,<sup>1</sup> Zhaoqi Liu,<sup>2,3,4,5</sup> Sisi Chen,<sup>1</sup> Michelle Ki,<sup>1</sup> Caroline Erickson,<sup>1</sup> Jorge S. Reis-Filho,<sup>6</sup> Benjamin H. Durham,<sup>1,6</sup> Qing Chang,<sup>7</sup> Elisa de Stanchina,<sup>7</sup> Yiwei Sun,<sup>4,5</sup> Raul Rabadan,<sup>4,5</sup> Omar Abdel-Wahab,<sup>1,8,9</sup> and Sarat Chandarlapaty<sup>1,9,10</sup>

<sup>1</sup>Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, New York, USA. <sup>2</sup>Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China. <sup>3</sup>China National Center for Bioinformation, Beijing, China. <sup>4</sup>Program for Mathematical Genomics, <sup>5</sup>Department of Systems Biology and Department of Biomedical Informatics, Columbia University, New York, New York, USA. <sup>6</sup>Department of Pathology, <sup>7</sup>Antitumor Assessment Core and Molecular Pharmacology Department, and <sup>8</sup>Leukemia Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New York, USA. <sup>9</sup>Weill-Cornell Medicine, New York, New York, USA. <sup>10</sup>Breast Medicine Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New York, USA.

**Mutations in the core RNA splicing factor *SF3B1* are prevalent in leukemias and uveal melanoma, but hotspot *SF3B1* mutations are also seen in epithelial malignancies such as breast cancer. Although hotspot mutations in *SF3B1* alter hematopoietic differentiation, whether *SF3B1* mutations contribute to epithelial cancer development and progression is unknown. Here, we identify that *SF3B1* mutations in mammary epithelial and breast cancer cells induce a recurrent pattern of aberrant splicing leading to activation of AKT and NF- $\kappa$ B, enhanced cell migration, and accelerated tumorigenesis. Transcriptomic analysis of human cancer specimens, *MMTV-cre Sf3b1*<sup>K700E/WT</sup> mice, and isogenic mutant cell lines identified hundreds of aberrant 3' splice sites (3'ss) induced by mutant SF3B1. Consistently between mouse and human tumors, mutant SF3B1 promoted aberrant splicing (dependent on aberrant branchpoints as well as pyrimidines downstream of the cryptic 3'ss) and consequent suppression of PPP2R5A and MAP3K7, critical negative regulators of AKT and NF- $\kappa$ B. Coordinate activation of NF- $\kappa$ B and AKT signaling was observed in the knockin models, leading to accelerated cell migration and tumor development in combination with mutant PIK3CA but also hypersensitizing cells to AKT kinase inhibitors. These data identify hotspot mutations in *SF3B1* as an important contributor to breast tumorigenesis and reveal unique vulnerabilities in cancers harboring them.**

## Introduction

Comprehensive genomic analyses of cancer have identified change-of-function mutations in RNA splicing factors to be recurrent across a variety of disparate cancer types (1). Akin to driver oncoproteins, these mutations occur as heterozygous mutations recurrently affecting the same nucleotide positions (so-called hotspot mutations). Hotspot mutations occur in both the protein (2–5) as well as RNA components of the spliceosome (6, 7) and globally alter pre-mRNA recognition by the spliceosome in a

manner distinct from loss of function (8–10). Malignancies of the hematopoietic system, including myeloid neoplasms and chronic lymphocytic leukemia (CLL), harbor a high frequency of these alterations and have been subject to detailed mechanistic investigations into the consequences of mutation on splicing, transcription, and cancer phenotypes (11–14).

*SF3B1* is the most commonly mutated RNA splicing factor, and studies of hotspot mutations in *SF3B1* have demonstrated that these mutations promote usage of aberrant branchpoint nucleotides (8, 15). Recent functional investigations of aberrant splicing driven by mutant SF3B1 in the context of myelodysplastic syndromes (MDS), CLL, and uveal melanoma (UVM) have begun to elucidate specific aberrant splicing events required for the maintenance of *SF3B1* mutant cancers (11). In addition, a number of studies in the context of myeloid leukemias have identified that *SF3B1* mutations confer therapeutic vulnerabilities to further modulation of splicing (16) as well as specific metabolic perturbations (17). However, to date, the biological consequences of expression of the same hotspot mutations in *SF3B1* in epithelial-derived malignancies are largely unknown and make for an intriguing counterpoint. While kinase oncoproteins like BRAF or NTRK function as targetable drivers in different tissue types (18–21), it is unknown whether large-scale modification of RNA splicing in different cell types is similarly oncogenic and uses the same pathways within distinct tissues to derive tumor phenotypes.

**Authorship note:** BL and ZL contributed equally to this work.

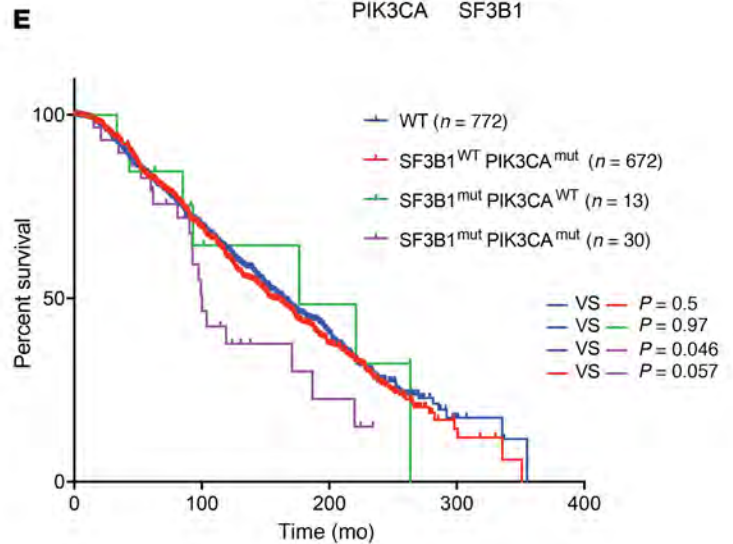
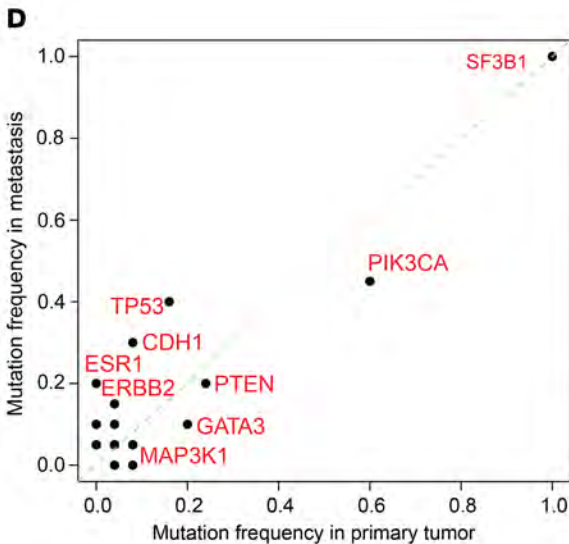
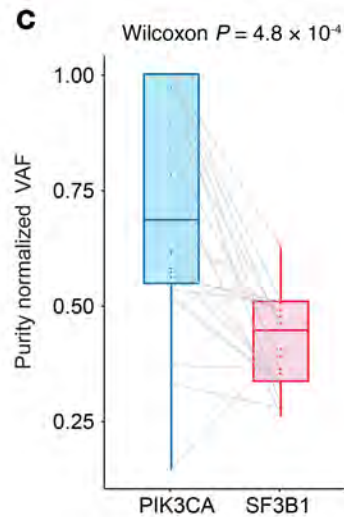
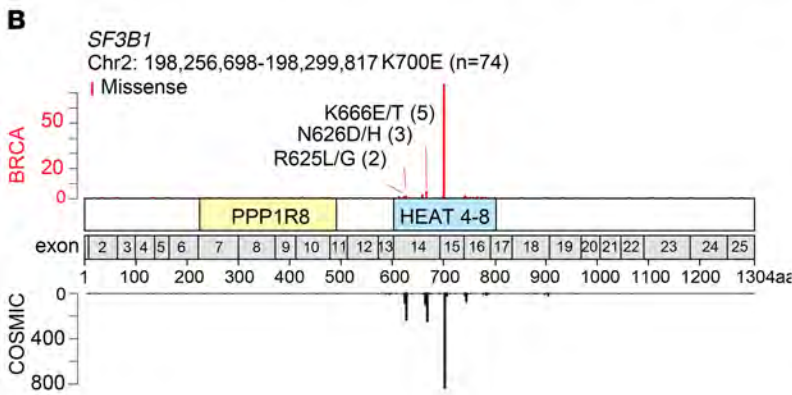
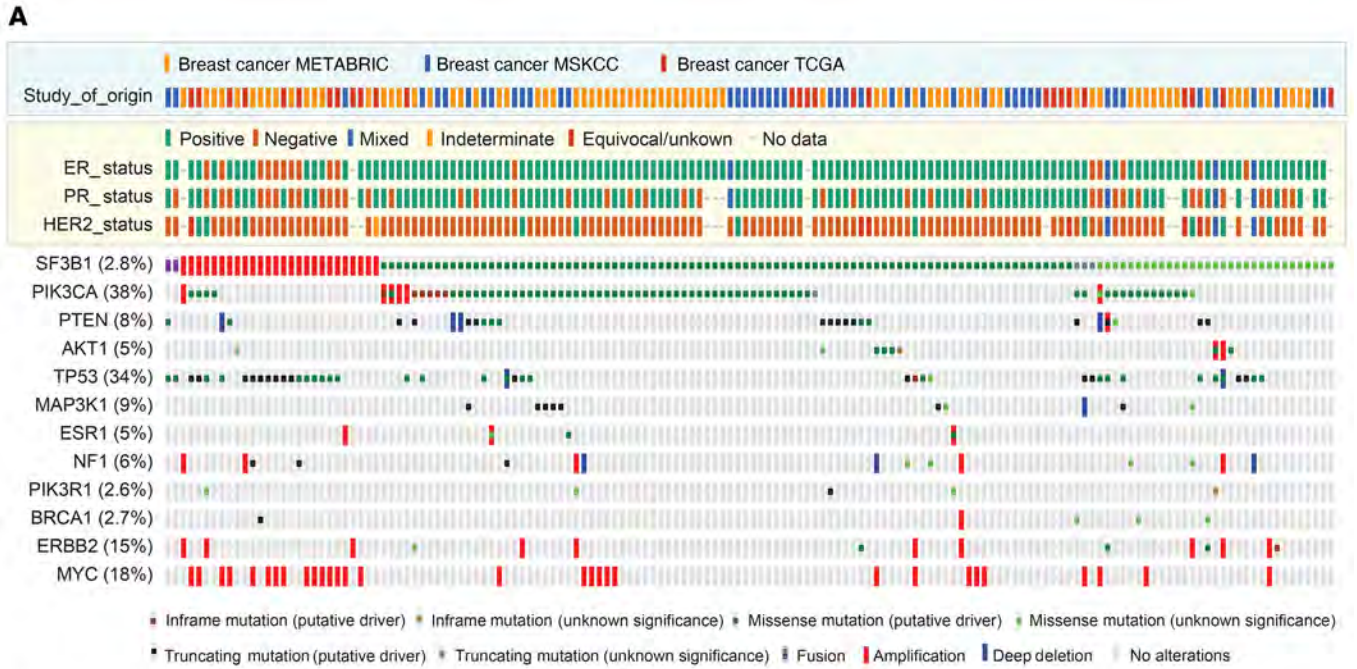
**Conflict of interest:** OAW has served as a consultant for H3B Biomedicine, Foundation Medicine Inc., Merck, and Janssen, and is on the scientific advisory board of Envisagenics Inc. and AlChem. He has received prior research funding from H3B Biomedicine and LOXO Oncology. JSRF has received personal/consultancy fees from Goldman Sachs, Paige AI, and REPARE Therapeutics. He is a member of the scientific advisory boards of VolitionRx, REPARE Therapeutics, and Paige AI; the Board of Directors of Grupo Oncoclínicas; and he is an ad hoc member of the scientific advisory boards of Roche Tissue Diagnostics, Ventana Medical Systems, Novartis, Genentech, and InVivo. RR is founder of GenoTwin and is a member of the scientific advisory board of AimedBio. SC has served as a consultant for Eli Lilly, Novartis, Context Therapeutics, Sermonix, Paige AI, and Bristol Myers Squibb. He has received prior research funding from Daiichi-Sankyo.

**Copyright:** © 2021, American Society for Clinical Investigation.

**Submitted:** March 19, 2020; **Accepted:** September 29, 2020; **Published:** January 4, 2021.

**Reference information:** *J Clin Invest.* 2021;131(1):e138315.

<https://doi.org/10.1172/JCI138315>.



**Figure 1. *SF3B1* hotspot mutations are recurrent in breast cancer and are significantly associated with mutations activating PI3K signaling and shortened survival.** (A) Oncoprint of somatic alterations in *SF3B1* and other breast cancer drivers across 5366 patients from the METABRIC (23, 65), MSK-IMPACT (24), and TCGA (22) breast cancer cohorts. ER, estrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor 2. (B) *SF3B1* mutation maps showing the counts, amino acid change, position, and evidence of mutational hotspots, based on COSMIC database information. The y axis counts at the bottom of the maps reflect the number of identified mutations in the COSMIC database. (C) Purity normalized variant allele frequency (VAF) of *PIK3CA* and *SF3B1* mutations among 51 double-mutated samples in the MSK-IMPACT cohort. (D) Frequency of somatic mutations in patients from the MSK-IMPACT cohort ( $n = 94$ ) harboring *SF3B1* hotspot mutations. Mutation frequency was calculated for each reported gene in 57 primary samples (x axis) and 45 metastasis samples (y axis). (E) Kaplan-Meier curve of disease-free survival in *SF3B1* hotspot mutant ( $n = 13$ ), *SF3B1* WT *PIK3CA* mutant ( $n = 672$ ), and *SF3B1/PIK3CA* double-mutant ( $n = 30$ ) versus WT ( $n = 772$ ) ER<sup>+</sup> breast cancer patients from METABRIC. *P* values were derived from log-rank test. See also Supplemental Figure 1.

In this study, we investigated the consequences of *SF3B1* mutations in breast cancer, where across a cohort of more than 5000 patients, *SF3B1* alterations are observed in approximately 3% of unselected cases. The effect of mutation upon global splicing, RNA expression, tumorigenesis, and tumor phenotypes highlights how aberrant splicing patterns are conserved but lead to lineage-specific effectors and phenotypes as well as novel therapeutic opportunities. Our data identify that mutations in *SF3B1* promote breast cancer development and progression via aberrant splicing and expression of intermediary signaling proteins that normally negatively regulate AKT and NF- $\kappa$ B signaling in mammary epithelial cells.

## Results

*SF3B1* mutations are enriched in estrogen receptor-positive (ER<sup>+</sup>) breast cancer and associate with poor outcomes. To systematically establish the prevalence and significance of *SF3B1* mutations in breast cancer, we performed a large-scale analysis of genomic/exomic sequencing data from 5366 patients with breast cancer, including prior data from the METABRIC, TCGA, and MSK-IMPACT databases (22–24) (Figure 1A and Supplemental Table 1; supplemental material available online with this article; <https://doi.org/10.1172/JCI138315DS1>). Genetic alterations in *SF3B1*, including mutations, gene fusions, and copy number amplification, were identified in 152 patients (2.83%) (Figure 1A), among which 90 patients (1.68%) carried hotspot mutations previously identified as genomic hotspots (25). The K700E ( $n = 74$ ) substitution in *SF3B1* was the dominant *SF3B1* mutation in patients with breast cancer, followed by hotspot mutations at K666 ( $n = 5$ ), N626 ( $n = 3$ ), and R625 ( $n = 2$ ) residues (Figure 1B). Among the patients with hotspot *SF3B1* mutations, ER status was available for 89 patients, only 2 of which were ER<sup>-</sup> (Figure 1A and Supplemental Table 2). These 2 patients both had hormone receptor positive primary cancer and later developed metastatic ER<sup>+</sup> tumors. Within the METABRIC and TCGA cohorts where Pam50 and claudin low subtyping is annotated, we found 84% (45/53) of *SF3B1* mutations occurred in luminal A or B subtypes, and 60% (32/53) of the cases were significantly enriched in luminal A breast cancer ( $P = 0.002$ )

(Supplemental Figure 1). In terms of other genomic alterations, *SF3B1* hotspot mutations significantly co-occurred with *PIK3CA* mutations ( $n = 55$ ; 2.76% in patients with *PIK3CA* mutations;  $\log_2$  odds ratio = 1.382;  $q < 0.001$ ) (Supplemental Figure 1). Interestingly, most *SF3B1* mutant samples that did not carry *PIK3CA* mutations harbored mutations in *PTEN* or *AKT1*, each of which converges on activating PI3K signaling.

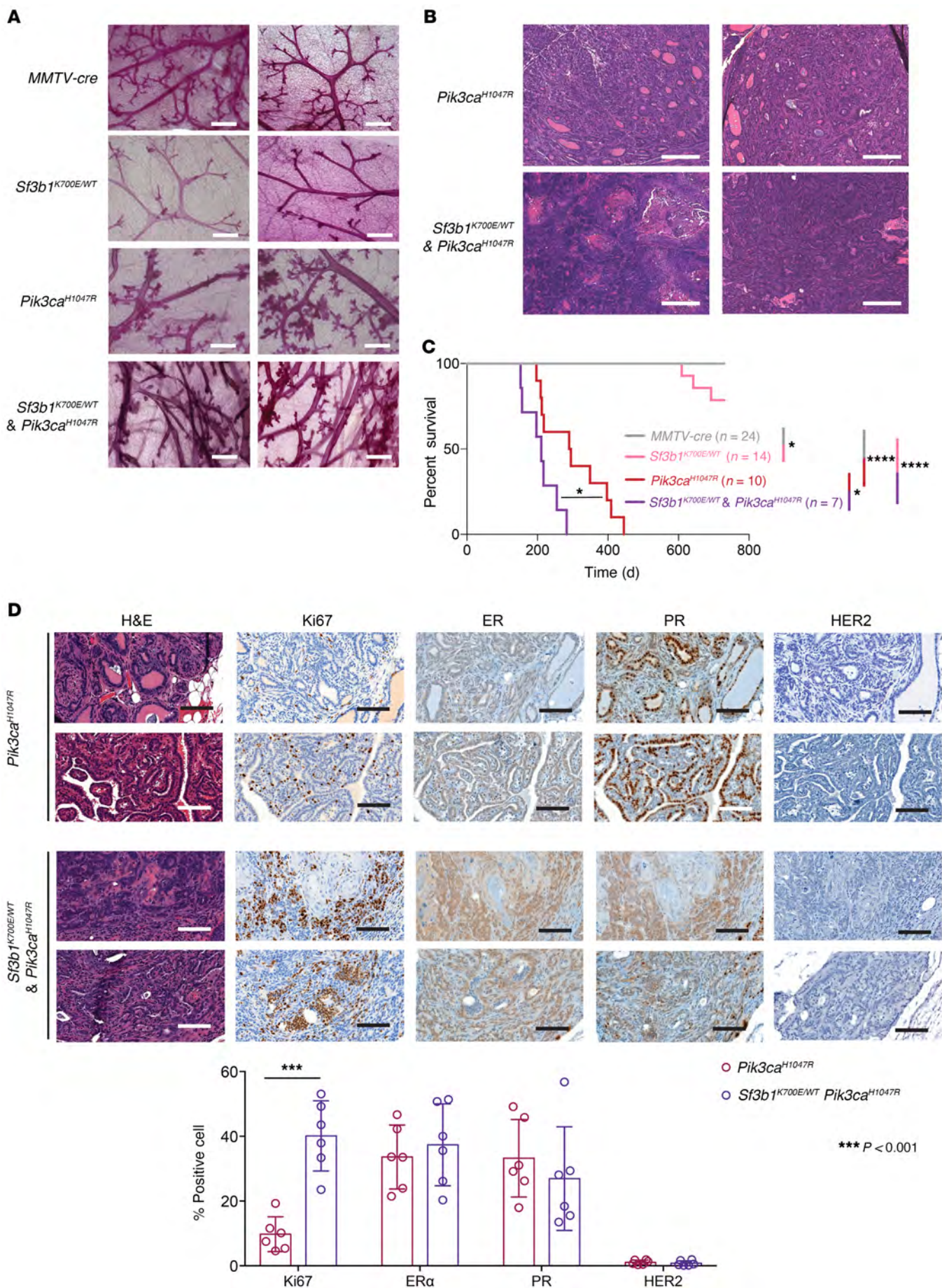
Given recent identification that ER<sup>+</sup> breast cancers frequently harbor several subclonal mutations that arise under the selection of endocrine-targeted therapy (24, 26, 27), we next evaluated the clonality of *SF3B1* mutations in our MSK-IMPACT cohort. Across tumors, mutations in *SF3B1* were clonal with a median allele frequency of about 50%, without manifesting allelic imbalances characteristic of mutant *PIK3CA* (Figure 1C). This latter finding is consistent with recent work from our group and others identifying essentiality of the WT allele in cells bearing hotspot change-of-function mutations in *SF3B1* and other RNA splicing factors (28–30). In considering the progression of tumors from primary to metastatic disease, genetic analysis of primary tumors and metastases from a series of patients with the *SF3B1* mutant identified frequent co-occurrence of *SF3B1* and *PIK3CA* mutations across primary tumors and metastases cases (Figure 1D).

We further analyzed the clinical outcomes of patients with *SF3B1* hotspot mutation and *PIK3CA* mutations (Figure 1E). The median survival time of patients with ER<sup>+</sup> breast cancer with no somatic alteration in either *SF3B1* or *PIK3CA* was 164 months, whereas that of patients with ER<sup>+</sup> disease harboring both *SF3B1* and *PIK3CA* mutations was 100.1 months (hazard ratio = 1.787,  $P = 0.046$ ). Neither *SF3B1* nor *PIK3CA* mutation alone was associated with significantly shorter survival. These findings point to *SF3B1* mutation as a recurrent alteration in ER<sup>+</sup> and *PIK3CA* mutant breast cancer that appears to be an early event and linked to poor prognosis.

### *Sf3b1*<sup>K700E</sup> promotes accelerated mammary tumor progression.

To determine the functional role of *SF3B1* hotspot mutations in breast tumorigenesis and progression, we generated a conditional knockin mouse model to express the most common *SF3B1* mutation found in patients with breast cancer (the K700E mutant) from the endogenous *Sf3b1* locus in the mouse mammary gland (31). Heterozygous *MMTV-cre Sf3b1*<sup>K700E/WT</sup> mice expressed mutant *Sf3b1* transcripts in mammary epithelial cells (sorted live, CD45<sup>-</sup>CD24<sup>+</sup> cells) at a 50% allelic frequency in RNA (similar to allelic frequency observed in human breast cancers; Figure 1C and Supplemental Figure 2, A–C). Due to the significant co-occurrence of *SF3B1* mutation and alterations activating the PI3K pathway, particularly *PIK3CA* mutations in patients with breast cancer (Supplemental Figure 1), we further crossed *MMTV-cre Sf3b1*<sup>K700E/WT</sup> mice with mice bearing a transgenic kinase domain mutation in *Pik3ca* from the *Rosa26* locus (*Rosa26 lox-stop-lox Pik3ca*<sup>H1047R</sup> (*R26-LSL-Pik3ca*<sup>H1047R</sup>)) (32) to evaluate the combinatorial effects of these mutations on mammary tumor development and progression.

Whole-mount examination of mammary glands from 4-month-old nulliparous mice of each genotype revealed ductal hyperplasia in mice expressing *Pik3ca*<sup>H1047R</sup> alone (*MMTV-cre R26-LSL-Pik3ca*<sup>H1047R</sup>) and in combination with mutant *Sf3b1* (Figure 2A and Supplemental Figure 2D). Interestingly, reduced branching of mammary ducts was observed in *MMTV-cre Sf3b1*<sup>K700E/WT</sup> mice (Supplemental Figure 2D). With further age,

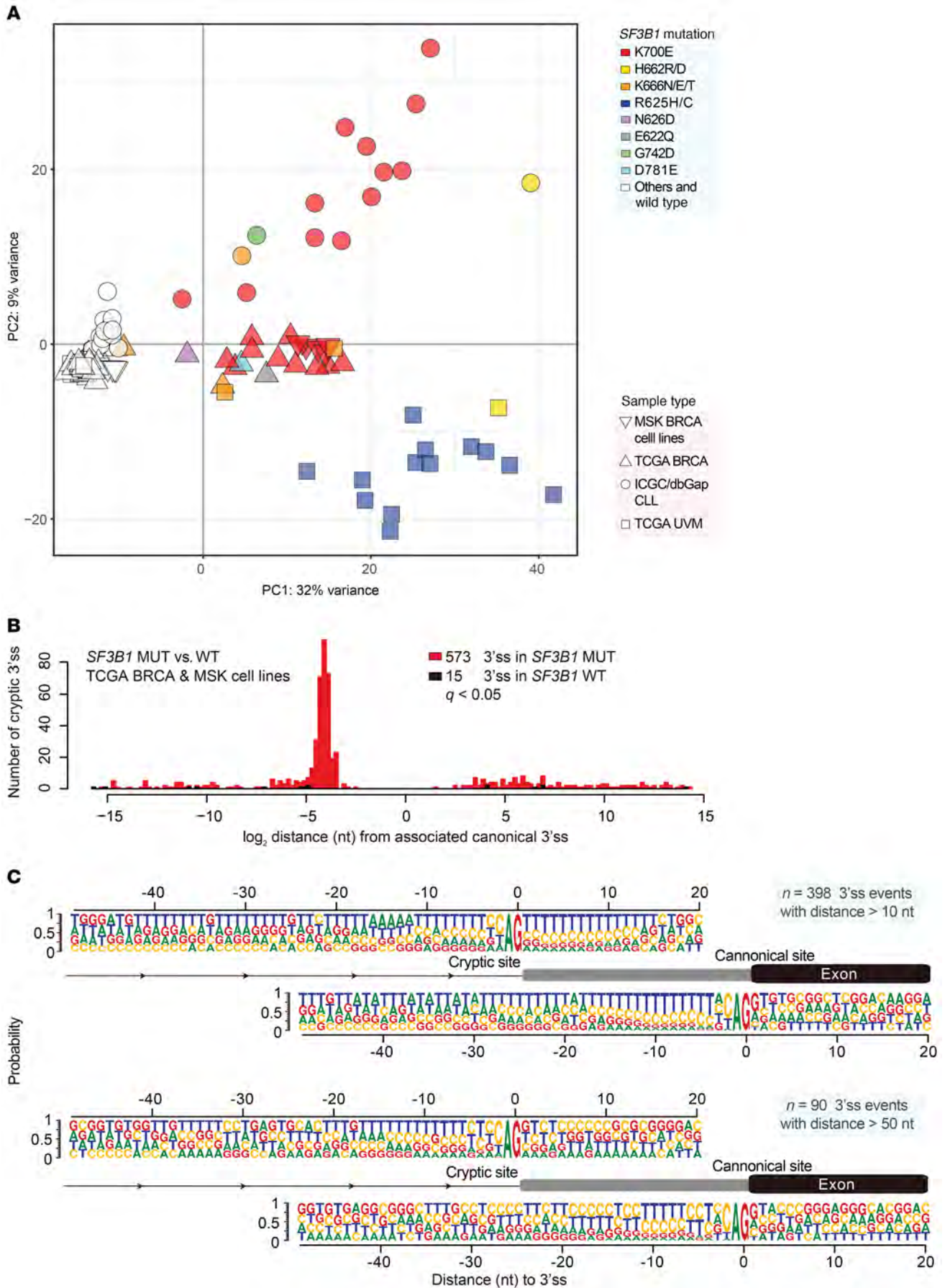


**Figure 2. *Sf3b1*<sup>K700E</sup> mutation promotes mammary tumorigenesis in conjunction with mutant *Pik3ca*.** (A) Whole-mount hematoxylin staining of mammary glands from *MMTV-cre* control, *MMTV-cre Sf3b1*<sup>K700E/WT</sup>, *MMTV-cre R26-LSL-Pik3ca*<sup>H1047R</sup>, and *MMTV-cre Sf3b1*<sup>K700E/WT</sup> *R26-LSL-Pik3ca*<sup>H1047R</sup> at the age of 4 months. Two representative examples are shown. Scale bars: 500  $\mu$ m. (B) Representative sections of H&E-stained mammary glands from the mice shown in A at time of death. Two representative examples are shown. Scale bars: 250  $\mu$ m. (C) Kaplan-Meier survival curve of the mice shown in A. The endpoint criterion was presence of tumors at least 1 cm<sup>3</sup>. *P* values calculated by log-rank (Mantel-Cox) test in Prism 7 software \**P* < 0.05, \*\*\*\**P* < 0.0001. (D) H&E and immunohistochemical staining of breast tumors developing in *MMTV-cre R26-LSL-Pik3ca*<sup>H1047R</sup> transgenic and compound *Sf3b1*<sup>K700E/WT</sup> *Pik3ca*<sup>H1047R</sup> transgenic mice. ER, estrogen receptor; PR, progesterone receptor. Scale bars: 100  $\mu$ m. Quantitation of percentage of positively stained cells (*n* = 6 tumors/group) is shown below. \*\*\**P* < 0.001 (Student's *t* test). See also Supplemental Figures 2–4.

*PIK3CA* single-mutant and *Sf3b1*<sup>K700E/WT</sup>/*Pik3ca*<sup>H1047R</sup> double-mutant mice (*MMTV-cre Sf3b1*<sup>K700E/WT</sup> *R26-LSL-Pik3ca*<sup>H1047R</sup>) went on to develop mammary gland tumors (Figure 2, B and C), while *MMTV-cre Sf3b1*<sup>K700E/WT</sup> mice did not exhibit any clear protumorigenic phenotype, which is consistent with previous studies with hematopoietic stem cells (13, 29) and B cells (33, 34) where *Sf3b1* K700E mutation alone was not sufficient to mediate tumorigenesis. Strikingly, the life span of double-mutant mice (median = 211 days) was significantly shorter than *Pik3ca* single-mutant mice (median = 293 days, *P* = 0.0125, Figure 2C). Akin to the tumors *MMTV-cre R26-LSL-Pik3ca*<sup>H1047R</sup> mice developed, *MMTV-cre Sf3b1*<sup>K700E/WT</sup> *R26-LSL-Pik3ca*<sup>H1047R</sup> mice develop lesions within the spectrum of tumors reported in *Pik3ca* conditional mouse models (35), including adenomyoepitheliomas and adenosquamous carcinomas (Supplemental Figure 3A). The adenomyoepitheliomas were composed of a proliferation of gland-like structures lined by a dual cell population of luminal epithelial (CK8/18 and Gata3 staining in Supplemental Figure 3A) and myoepithelial cells (Sma staining in Supplemental Figure 3A), whereas the adenosquamous carcinomas were composed of a complex admixture of gland-like structures, sheets of squamous cells (CK14 staining in Supplemental Figure 3A) and occasional squamous pearls; areas of necrosis were observed. No other histologic abnormalities were detected in the liver, lungs, kidneys, adrenals, spleen, heart, and central nervous system of these mice. Given that *MMTV-cre* has been reported to be expressed in tissues besides mammary epithelial cells (including expression in salivary gland and lymphocytes) (36), we examined both the salivary gland and performed a comprehensive analysis of hematopoietic tissues but did not identify overt malignancies in these tissues at the time of breast tumor formation (Supplemental Figure 3B and Supplemental Figure 4). To characterize the mammary gland tumors that developed in these mice, we probed ER, progesterone receptor (PR), HER2, and Ki67 in representative tumor sections. While all tumors collected were ER<sup>+</sup>, PR<sup>+</sup>, and HER2<sup>+</sup>, Ki67 staining was significantly higher in tumors from *MMTV-cre Sf3b1*<sup>K700E/WT</sup> *Pik3ca*<sup>H1047R</sup> mice than in those from *Pik3ca*<sup>H1047R</sup> single-mutant mice (Figure 2D). We next sought to understand the mechanistic basis by which *SF3B1* mutation could contribute to accelerated progression that was observed both in the mouse and human data.

*Pan-cancer and breast cancer-specific aberrant splicing events induced by mutant SF3B1.* Prior work in hematopoietic-derived cells and UVM has revealed that cancer-associated mutations in *SF3B1* result in splicing events that use aberrant branchpoint nucleotides, manifesting in increased use of aberrant (or so-called cryptic) intron-proximal 3' splice sites (3'ss) (15, 37). To understand the impact of *SF3B1* mutations on splicing in breast cancers compared with other malignancies, we performed a pan-cancer transcriptomic analysis to identify and quantify cryptic 3'ss usage across *SF3B1* WT versus mutant cancers (Figure 3). For breast cancer, RNA-seq data from 20 tumors harboring *SF3B1* mutations and 20 randomly selected, ER<sup>+</sup> WT tumors were collected from the TCGA breast cancer (BRCA) database (Supplemental Table 3). In this analysis, we also included an isogenic pair of parental and *SF3B1* K700E knockin MCF10A breast epithelial cells, as well as MCF7 and T47D breast cancer cells ectopically expressing *SF3B1* WT or *SF3B1* K700E mutant via a doxycycline-inducible expression system (Supplemental Figure 5). The expression of the *SF3B1* K700E mutant did not significantly affect cell proliferation and dependence on ER of these breast cancer cells under in vitro culture conditions (Supplemental Figure 5). For the comparison of *SF3B1* mutant breast cancer to other cancer types bearing *SF3B1* mutations, RNA-seq data of 15 patients with CLL and 16 patients with UVM together with same number of randomly chosen WT samples of the same histologic types of cancer were also included (Supplemental Table 3). Principal component analysis and unsupervised hierarchical clustering of cryptic 3'ss usage using the percent spliced in (PSI) matrix (Figure 3A, Supplemental Figure 6 and Supplemental Methods) robustly separated samples with *SF3B1* hotspot mutations from WT cases, with the exception of one breast cancer sample with an *SF3B1* K666N mutation (a result likely explained by the low allelic frequency [7%] of the mutation in the RNA-seq data from this sample). Samples with non-hotspot *SF3B1* mutations (e.g., S455V, E802Q) clustered together with *SF3B1* WT samples (Supplemental Figure 6 and Supplemental Table 3), suggesting that these mutations do not cause aberrant 3'ss usage. Within the *SF3B1* mutant group, samples clustered separately into subgroups by their cancer types. Breast cancer patient and cell line samples (triangles in Figure 3A) showed a very similar pattern of aberrant splicing, compared with CLL (circles in Figure 3A) or UVM samples (squares in Figure 3A), suggesting the aberrant 3'ss usage induced by *SF3B1* mutations in distinct cancers may be due to intrinsic gene expression patterns unique to cell lineage.

Consistent with previous studies in other cancer types (15, 37), most aberrant 3'ss in *SF3B1* mutant breast tumors were located 10 to 30 nucleotides upstream of the canonical 3'ss (Figure 3B). For most unaffected 3'ss, a stretch of pyrimidine nucleotides (known as a polypyrimidine tract) is present within the intron, near the canonical 3'ss. To understand the mechanism by which *SF3B1* mutations affect 3'ss recognition, we analyzed the sequences around cryptic 3'ss promoted in *SF3B1* mutant cells (see Methods; Figure 3C). As previously reported, this revealed that polypyrimidine tracts at cryptic 3'ss were interrupted by adenines, indicating a shorter distance between the cryptic 3'ss and branchpoint nucleotides (8, 15). At the same time, we also identified an enrichment of pyrimidines downstream of the aberrant 3'ss. Since most cryptic 3'ss were within 30 nucleotides upstream of the canonical 3'ss, this



**Figure 3. Differential aberrant 3' splicing events across *SF3B1* mutant cancers.** (A) Principal component analysis of differential 3' splice sites (3'ss) in *SF3B1* mutant and WT samples. The PSI input matrix contains 618 cryptic 3'ss events identified from the pan-cancer data set are shown. BRCA, breast cancer; CLL, chronic lymphocytic leukemia; UVM, uveal melanoma. (B) Histogram of statistically significant ( $q < 0.05$ ) cryptic 3'ss identified in BRCA samples: WT (black) versus mutant (red) for *SF3B1* and their distance in nucleotides (nt) from the associated canonical 3'ss. (C) Consensus 3'ss motif near the associated cryptic 3' AG dinucleotide (top) and canonical 3' AG dinucleotide (bottom) for 398 aberrant 3'ss events more than 10 nt from the canonical 3'ss as well as 90 aberrant 3'ss events more than 50 nt from the canonical 3'ss. See also Supplemental Figures 5 and 6.

enrichment could be due to a bias from the polypyrimidine track associated with the canonical 3'ss. To eliminate this bias, we performed a sequence analysis exclusively on cryptic sites 50 nucleotides away from the canonical 3'ss. This analysis continued to reveal significant enrichment of pyrimidines within these 90 aberrant 3'ss ( $P < 0.0001$ , Fisher's exact test, Figure 3C). These data indicate that recognition of cryptic 3'ss by mutant SF3B1 may require downstream pyrimidine-rich motifs, a finding which was evident across *SF3B1* mutant cancer types.

**Aberrant splicing mediated suppression of *PPP2R5A* and *MAP3K7* in *SF3B1*-mutant breast cancers.** To identify the key cellular processes that mutant SF3B1 alters in accelerating breast cancer progression, we performed a gene set enrichment analysis (GSEA) on the RNAs derived from WT and *SF3B1*-mutant breast cancer cell lines. This revealed that inflammatory response pathways ( $P = 0$ ), and NF- $\kappa$ B signaling ( $P = 0.002$ ) in particular, were significantly altered in *SF3B1* K700E-expressing breast cancer cell lines (Figure 4A). Moreover, GSEA analysis on the TCGA BRCA gene expression data also revealed upregulation of these same pathways in the patient samples bearing hotspot *SF3B1* mutations (Supplemental Figure 7A). To verify activation of NF- $\kappa$ B signaling as a result of mutant SF3B1 expression, we performed Western blot analysis to assess phosphorylation of p65 (RelA), a core subunit of NF- $\kappa$ B complex, in isogenic *SF3B1* K700E knockin MCF10A cells and MCF7 cells expressing mutant SF3B1 (Figure 4B and Supplemental Figure 7B). Phosphorylation of p65 was upregulated by 3- to 5-fold in mutant SF3B1-expressing cells, and the levels of phosphorylation were further enhanced following stimulation by TNF- $\alpha$  (Figure 4B and Supplemental Figure 7B). To further assess the significance of this activation of NF- $\kappa$ B, we examined the effect of *SF3B1* mutation on the expression of inflammatory cytokines known to be regulated by NF- $\kappa$ B. We measured the quantity of secreted cytokines in the culture media of MCF10A and MCF7 WT versus *SF3B1* K700E-expressing cells. IL-1 $\alpha$ , IL-8, and TGF- $\alpha$  were all increased in the media derived from *SF3B1* K700E MCF10A cells (Figure 4C), and mutant SF3B1-expressing MCF7 cells produced 10-fold more IL-8 than SF3B1 WT counterpart isogenic cells (Figure 4C), consistent with the observed upregulation of p65 phosphorylation in these cells. Upregulated IL-8 secretion in both MCF10A and MCF7 cells expressing mutant SF3B1 was also confirmed by ELISA (Figure 4D). These data together confirm the induction of NF- $\kappa$ B signaling observed in the GSEA analysis of *SF3B1* mutant tumors.

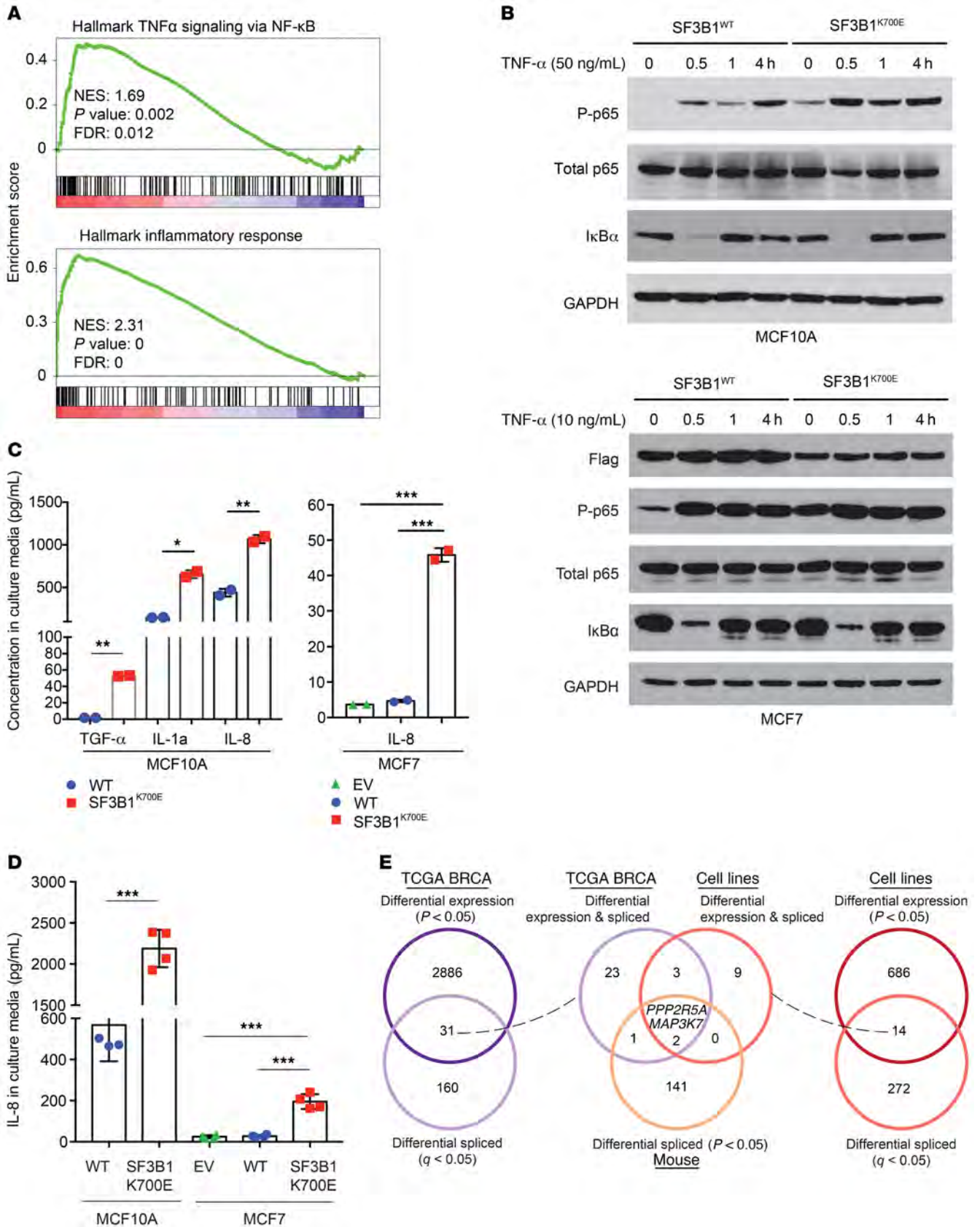
To understand the molecular mechanisms for NF- $\kappa$ B activation in *SF3B1* mutant cells, we examined the intersection of differentially expressed and differentially misspliced genes in human

tumors, cell lines, and our mouse model (Figure 4E and Supplemental Figure 7C). From this intersection, only 2 genes were identified: *MAP3K7* and *PPP2R5A*, both of which were significant reduced in expression in *SF3B1* mutants (Figure 4E and Supplemental Tables 4–6). Of note, these missplicing events in *MAP3K7* and *PPP2R5A* have been reported in hematopoietic malignancies with *SF3B1* mutations recently by Lee et al. (29) and Liu et al. (34).

**Mutant *SF3B1* causes aberrant splicing of *MAP3K7*, stimulating NF- $\kappa$ B signaling.** *MAP3K7* has known roles in regulating NF- $\kappa$ B signaling (38–40) and so we initially sought to verify its missplicing and biologic contribution to mutant SF3B1-mediated pathogenesis. To confirm a consistent effect of *SF3B1* mutation on *MAP3K7* splicing and expression, we examined the RNA-seq data of the isogenic cell lines, mouse mammary tissue, and human tumors at the candidate intron where missplicing was implicated (Figure 5, A–C). Across these samples, mutant SF3B1 led to aberrant splicing of *MAP3K7* via use of a cryptic 3'ss that was conserved in human breast cell lines and tumors and mouse mammary gland expressing *Sf3b1*<sup>K700E</sup> (Figure 5, A and B). As expected, NMD caused by missplicing led to reduced *MAP3K7* transcript and protein levels in mutant SF3B1-expressing cells and mouse mammary gland tumors (Figure 5, C–E, and Supplemental Figure 8A).

To understand this aberrant splicing event in *MAP3K7* in greater molecular detail, we mapped the branchpoints in the aberrantly spliced intron by lariat RT-PCR. Sequencing of the PCR products identified the branchpoints used for the canonical as well as cryptic 3'ss (Supplemental Figure 9A). The branchpoint for the canonical 3'ss was an adenine 22 nucleotides upstream of the intron-exon boundary, whereas the branchpoint for the cryptic 3'ss mapped to an adenine 14 nucleotides upstream of the aberrant intron-exon boundary (Figure 5F). To identify *cis* nucleotides required for mutant SF3B1-specific aberrant splicing of *MAP3K7*, we generated a minigene construct that contains sequences from the misspliced intron and flanking exons. As with the endogenous *MAP3K7* transcript, mutant SF3B1 also enhanced usage of a cryptic 3'ss in the minigene-derived *MAP3K7* RNA (Figure 5F). Substitution of adenines at the branchpoint with guanines blocked the usage of the cryptic 3'ss (Figure 5F). A pyrimidine-rich motif was located in the sequences downstream of the cryptic 3'ss. We mutated thymine nucleotides within this pyrimidine-rich motif to test if these residues were necessary for mutant SF3B1-specific splicing. Substitution of either 2 or 3 thymines to adenines completely abolished the usage of the cryptic 3'ss and aberrant inclusion of the 20-nucleotide intronic sequences (Figure 5F). These data identify the requirement of the aberrant branchpoint as well as pyrimidines downstream of the aberrant branchpoint in missplicing *MAP3K7*.

We next sought to establish the biologic role for *MAP3K7* in NF- $\kappa$ B signaling in breast cancer cells as its regulation of NF- $\kappa$ B signaling, which has been previously shown to be cell-type specific (38–40). To characterize the impact of *MAP3K7* on NF- $\kappa$ B activation in breast cells specifically, we manipulated *MAP3K7* activity in MCF7 and MCF10A cells. Knockdown of *MAP3K7* greatly increased phosphorylation of p65 and reduced p38 phosphorylation (Figure 5G and Supplemental Figure 8B), which was also observed with pharmacological inhibition of *MAP3K7* using 5Z-7-oxozeaenol (41) (Supplemental Figure 8C). Consistent with



**Figure 4. Increased inflammation and activation of NF- $\kappa$ B signaling in SF3B1 mutant mouse and human breast tumors.** (A) GSEA identifies enrichment of hallmark gene sets of TNF- $\alpha$  signaling via NF- $\kappa$ B (top) and inflammatory response (bottom) in SF3B1 mutant versus WT human breast cell lines. FDR, false discovery rate; NES, normalized enrichment score. (B) Activation of phospho-p65 (P-p65) upon TNF- $\alpha$  stimulation of MCF10A SF3B1<sup>K700E/WT</sup> mutant knockin cells (top) and MCF7 with transgenic expression of SF3B1<sup>K700E</sup> versus SF3B1<sup>WT</sup> cDNA (bottom). Quantitation of 3 independent experiments is shown in Supplemental Figure 7B. (C) Levels of TGF- $\alpha$ , IL-1 $\alpha$ , and IL-8 in culture media from MCF10A cells shown in B or IL-8 in culture media from MCF7 cells with expression of an empty vector (EV), SF3B1 WT, or SF3B1 K700E cDNAs measured by Milliplex cytokine/chemokine multiplex assay. (D) Levels of IL-8 in culture media shown in C measured by ELISA. Mean levels with standard deviation shown. (\* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.0001$ ; Student's  $t$  test for MCF10A and Dunnett's multiple comparisons test for MCF7). (E) Venn diagrams of genes significantly differentially expressed ( $P < 0.05$ ; DESeq2) and spliced ( $q < 0.05$ ;  $t$  test) across SF3B1 mutant versus WT human breast cancers from TCGA (22) (left), isogenic human breast cell lines (right), and overlapping across human tumors, human cell lines, and mouse mammary epithelial cells mutant for Sf3b1 (middle). See also Supplemental Figure 7.

this, restored expression of MAP3K7 (via expression of MAP3K7 cDNA) in SF3B1 K700E isogenic MCF10A cells or MCF7 expressing mutant SF3B1 greatly reduced NF- $\kappa$ B activation and promoted p38 phosphorylation (Figure 5H and Supplemental Figure 8D). Taken together, these data establish mutant SF3B1-directed missplicing of MAP3K7 to be a key mediator of enhanced NF- $\kappa$ B signaling in breast cancer.

*Mutant SF3B1 causes aberrant splicing of PPP2R5A, enhancing AKT activation and inflammatory signaling.* In addition to missplicing of MAP3K7, SF3B1 K700E mutation also induced usage of a cryptic 3' splice site 13 nucleotides upstream the normal 3' splice site of intron 4 of PPP2R5A in human and mouse cells (Figure 6, A and B). PPP2R5A encodes the alpha isoform of the B56 regulatory subunit of PP2A (the serine/threonine-protein phosphatase 2A). PP2A is 1 of 4 major serine/threonine phosphatases, and it is a heterotrimeric protein consisting of a catalytic C subunit, 1 structural A subunit, and a regulatory B subunit. Thirteen distinct genes encoding for the different regulatory B subunits and the specific B subunit expressed regulates substrate selectivity and catalytic activity of PP2A.

Mutant SF3B1 promoted the inclusion of 13 intronic nucleotides in exon 5 by using an upstream cryptic 3' splice site, resulting in a reading frame shift (Figure 6, A and B). This is predicted to cause NMD of the aberrantly spliced PPP2R5A transcript and, consistent with this, SF3B1 mutant patient breast tumors had significantly reduced PPP2R5A mRNA and protein expression compared with WT (Figure 6, C and D, and Supplemental Figure 10A). As with MAP3K7, missplicing of PPP2R5A by mutant SF3B1K700E could be replicated within a PPP2R5A minigene and this missplicing was abolished by disruption of branchpoint (Figure 6E and Supplemental Figure 9B) or interruption of the poly-pyrimidine tract with non-pyrimidines (Figure 6E).

PP2A has been reported to regulate phosphorylation of a variety of signaling proteins, including AKT, MEK1/2, and MYC, the downstream consequence of which could be NF- $\kappa$ B signaling. We therefore probed known PP2A substrates in the context of breast epithelial cells by Western blot and found increased

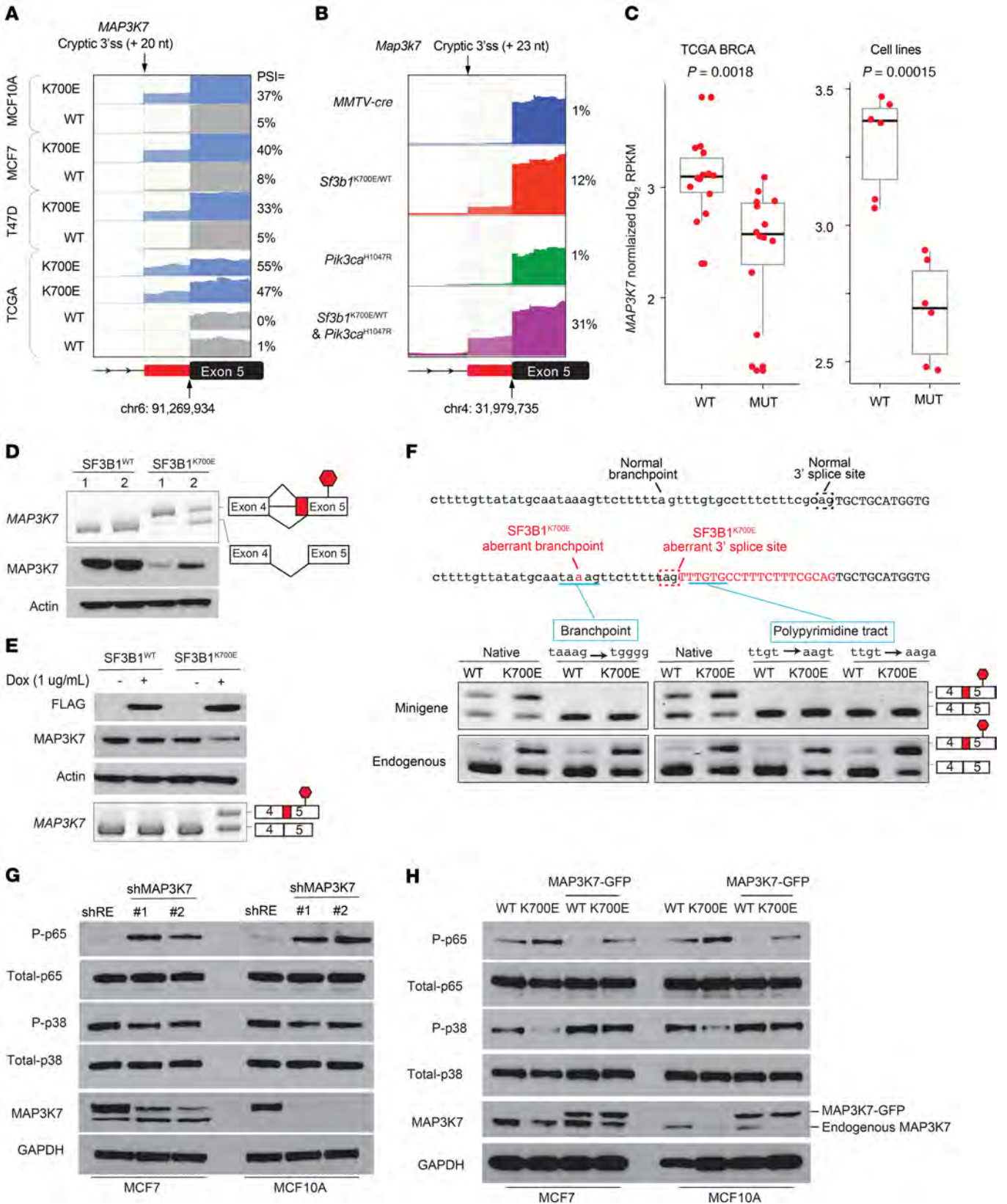
phosphorylation of AKT at residue T308 in breast cells expressing mutant SF3B1 (Figure 6D and Supplemental Figure 10A). Knockdown of PPP2R5A in MCF10A and MCF7 cells by shRNA also upregulated phosphorylation of p65 and AKT (at T308) (Figure 6F and Supplemental Figure 10B). Furthermore, depletion of PPP2R5A in cells also induced expression of inflammatory response genes, including IL-8, CDKN1A, IL-1A, and CXCL1 (Figure 6G). Restoration of PPP2R5A expression in MCF7 cells expressing mutant SF3B1 or isogenic SF3B1<sup>K700E/WT</sup> MCF10A cells showed consistently reduced phosphorylation of p65 and AKT (Supplemental Figure 10C). Given the known roles of AKT in regulating IKK and NF- $\kappa$ B activation, our data point to PPP2R5A downregulation as a second mechanism whereby mutant SF3B1 promotes NF- $\kappa$ B signaling in breast cancer.

*SF3B1 mutations promote invasion and migration of breast cancer cells.* NF- $\kappa$ B signaling has been implicated in regulating cell migration of breast cancer cells (42–44). To test if SF3B1 mutation facilitates cell migration, we initially performed wound-healing assays. MCF7 cells expressing mutant SF3B1 traversed wounds significantly faster than the cells expressing WT SF3B1 or empty vector (Figure 7A) irrespective of stimulation by TNF- $\alpha$ . Comparable differences were also seen with MCF10A isogenic pairs (Figure 7B).

Since SF3B1 mutant cells secrete more inflammatory cytokines (including IL-8), we hypothesized that the higher migration might be mediated by cytokines within the culture media. To address this possibility, we harvested media from cells expressing WT or mutant SF3B1 and studied the effect of media addition upon migration of parental MCF7 cells in wound-healing assays. WT MCF7 cells exposed to media from mutant SF3B1-expressing cells healed wounds faster than controls (Figure 7C). In addition to cellular motility, NF- $\kappa$ B signaling has been linked to breast cancer invasiveness and we thus investigated the effects of SF3B1 mutation in Transwell Matrigel invasion assays. MCF10A breast epithelial cells hardly invaded Matrigel, whereas significantly more SF3B1 K700E knockin MCF10A cells migrated through the Matrigel layer (Figure 7D). Moreover, we assessed for enhanced invasiveness using a xenograft assay (Figure 7E), in which limiting quantities of WT and mutant SF3B1 expressing MCF7 cells were implanted into NSG mice, and observed that the SF3B1 mutant tumors grew to significantly larger sizes despite equivalent growth rates for the corresponding cell line models in 2D culture conditions.

As missplicing of MAP3K7 and PPP2R5A induced by mutant SF3B1 mediates dysregulation of NF- $\kappa$ B, we further tested the role of these proteins on cell migration. Knockdown of either MAP3K7 or PPP2R5A enhanced cell migration (Supplemental Figure 11, A and B and Supplemental Figure 12, A and B), and restoring these proteins in SF3B1 mutant cells restrain cell migration as assessed in wound-healing assays (Supplemental Figure 11, C and D, and Supplemental Figure 12, C and D). These data indicate that mutant SF3B1-mediated dysregulation of NF- $\kappa$ B signaling enhances breast cancer migration and invasion.

*SF3B1 K700E mutations sensitize breast cancer cells to AKT inhibitors.* While the independent splicing events identified above converge upon promoting NF- $\kappa$ B signaling and associated phenotypes, therapeutic approaches targeting NF- $\kappa$ B have proven challenging. By contrast, PI3K/AKT inhibition has shown some



**Figure 5. Aberrant splicing and downregulation of MAP3K7 by mutant SF3B1 promotes NF- $\kappa$ B activation in breast cancer.** Representative RNA-seq coverage plots showing cryptic 3' splice usage in *MAP3K7* in (A) human isogenic breast cancer cell lines (MCF10A, MCF7, T47D) expressing WT or SF3B1 K700E (K700E) and human breast tumor samples from TCGA (22) as well as (B) mouse mammary epithelial cells from MMTV-cre control, MMTV-cre *Sf3b1*<sup>K700E/WT</sup>, MMTV-cre *R26-LSL-Pik3ca*<sup>H1047R</sup> transgenic, and compound MMTV-cre *Sf3b1*<sup>K700E/WT</sup> *R26-LSL-Pik3ca*<sup>H1047R</sup> mice. PSI value is shown on the right of each coverage plot. (C) *MAP3K7* mRNA expression (normalized log<sub>2</sub>RPKM values) in *SF3B1* mutant (MUT) versus WT human breast tumors from TCGA and cell lines shown in A. Box indicates upper and lower quartiles; thick bar indicates the median value. Whiskers indicate the largest/smallest value no further than 1.5 times the interquartile range. (D) RT-PCR of *MAP3K7* splicing event (top) and *MAP3K7* protein levels in isogenic MCF10A cells with or without *SF3B1*<sup>K700E</sup> knockin. The 3' splice used is shown in the schematic. (E) *MAP3K7* protein levels in MCF7 cells expressing doxycycline-inducible FLAG-SF3B1<sup>WT</sup> or FLAG-SF3B1<sup>K700E</sup> vectors. RT-PCR of *MAP3K7* splicing is shown on bottom. (F) RT-PCR of *MAP3K7* splicing event derived from *MAP3K7* minigene with mutations at aberrant branchpoint or polypyrimidine region downstream of the aberrant 3' splice. RT-PCR of *MAP3K7* splicing event derived from endogenous *MAP3K7* is shown on bottom. A schematic of sequences around the normal and aberrant branchpoint and splice site is shown on top. (G) *MAP3K7* and phospho- and total p65 and p38 levels in MCF7 and MCF10A cells transduced with anti-*MAP3K7* shRNA. (H) *MAP3K7* and phospho- and total p65 and p38 levels in MCF7 and MCF10A cells expressing SF3B1 K700E mutant from D and E transduced with or without *MAP3K7-GFP* cDNA. Quantitation of 3 independent experiments is shown in Supplemental Figure 8. See also Supplemental Figure 9.

efficacy in breast cancer and so we sought to determine whether suppression of PPP2R5A might increase tumor dependence on AKT for growth and viability. Given the induction of AKT T308 phosphorylation in SF3B1 K700E knockin cells, we evaluated the effects of *SF3B1* mutation upon a set of established downstream substrates of activated AKT in breast cancer cells. In isogenic MCF7 breast cancer cells, expression of *SF3B1* K700E increased phosphorylation of FOXO1, PRAS40, and ribosomal protein S6 (Figure 8A and Supplemental Figure 13A), further confirming AKT activation. Recent studies have suggested that increased signaling output through PI3K/AKT can enforce heightened dependence and drug sensitivity. We therefore evaluated the effect of expression of mutant SF3B1<sup>K700E</sup> on selective inhibitors of AKT. We examined the effect of both an allosteric pan-AKT inhibitor, MK2206, as well as an ATP-competitive pan-AKT inhibitor, GDC0068, in MCF7 cells, which are *PIK3CA* mutant and sensitive to AKT inhibitors, and observed a 2.5 times to 4 times further reduction in the IC<sub>50</sub> as a result of *SF3B1* mutation (Figure 8B). Finally, to examine the sensitivity of MCF10A cells to AKT inhibitors, we created MCF10A cells with heterozygous knockin of *PIK3CA*<sup>H1047R</sup> in the endogenous locus of *PIK3CA* with or without combined *SF3B1*<sup>K700E/WT</sup> knockin (Supplemental Figure 13B). These cells exhibited comparable splicing alterations of *MAP3K7* and *PPP2R5A*, as well as activation of NF- $\kappa$ B signaling (Supplemental Figure 14A). Mutant *PIK3CA*<sup>H1047R</sup> induced expression of inflammatory response genes and promoted cell migration, which was further enhanced in the cells with both *SF3B1*<sup>K700E</sup> and *PIK3CA*<sup>H1047R</sup> mutations (Supplemental Figure 14). Knockin of *PIK3CA*<sup>H1047R</sup> relieved the dependency of MCF10A cells on epidermal growth factor (EGF) supplemented to culture media and activated the PI3K-AKT signaling pathway (Figure 8C and Supplemental Figure

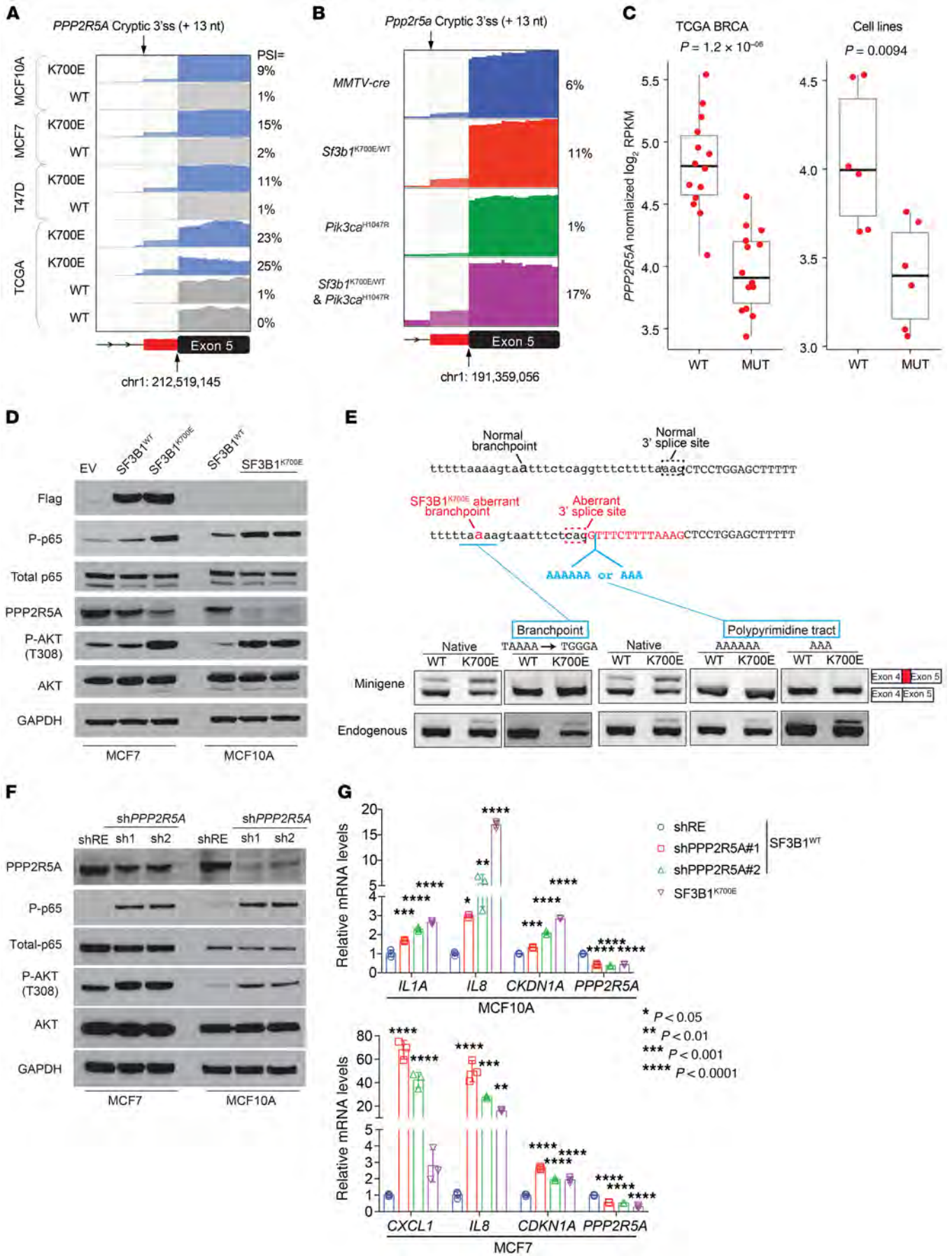
13C). Again, in these cells, expression of mutant SF3B1 enhanced the sensitivity of *PIK3CA* mutant cells to inhibition of AKT (Figure 8D). This enhanced sensitivity of *PIK3CA* mutant cells induced by *SF3B1* mutation was also confirmed by clonogenic survival assays (Figure 8, E and F), in which 2.5 times to 5 times reduction in IC<sub>50</sub> was associated with expression of the *SF3B1* mutation. Together, these data suggest that *SF3B1* mutant tumors are a unique subset of breast cancers wherein coordinate mutational activation of signaling together with suppression of key negative regulator PPP2R5A leads to enhanced sensitivity of AKT inhibitors, which are under current investigation in the clinic.

## Discussion

Identifying and targeting oncogenic pathways stimulated by gain-of-function genomic alterations has resulted in major clinical advances in outcomes for patients with breast cancer (45–49). Recent comprehensive genomic analyses of thousands of patients with breast cancer have identified a number of subgroups of patients with recurrent genomic alterations ranging from 1%–10% and for which deeper understanding of their contribution to tumor growth is lacking (22, 50, 51). As a result, functional investigation of the genetic alterations enriched in these specific subsets of patients with breast cancer has become paramount. Successful examples include the use of PD1 blockade in mismatch repair-deficient tumors (52), PARP inhibitors in germline *BRCA1/2*-mutant breast cancer (53), or anti-androgen therapy in androgen receptor-positive triple-negative breast cancer (54, 55). In this work, we identified what we believe is a novel class of recurrent genetic alterations that appear to promote breast tumorigenesis through modifying the function of the core RNA splicing factor SF3B1. Our data revealed *SF3B1* mutant breast cancers to be a unique transcriptomic subtype of ER<sup>+</sup> breast cancers with strong coenrichment with PI3K/AKT alterations, augmenting AKT and NF- $\kappa$ B signaling.

Given prior work identifying how *SF3B1* mutations alter splice site sequence preference with attendant effects of select genes (11, 29, 37), we sought to profile aberrant splicing across breast cancers and compare these results with other cancers where *SF3B1* mutations are found. Although the results demonstrated aberrant 3' splice usage shared across *SF3B1* mutant tumors, only a handful of events were shared between different breast cancer samples, isogenic ER<sup>+</sup> and ER<sup>-</sup> breast cancer cell lines, and our mouse breast cancer models. Specifically, our consensus overlap identified only 2 transcripts shared between cell lines and tumors with convergent alterations in 3' splice usage and mRNA downregulation — *PPP2R5A* and *MAP3K7*. Importantly, these 2 events are also observed in other cancer types, such as MDS and CLL, which identifies them as potentially critical mediators of the effects of mutant SF3B1 across diverse cancer lineages. For both of these mRNAs, mutant SF3B1 promoted usage of an aberrant branchpoint adenosine nucleotide that was required for missplicing, and the presence of a polypyrimidine tract downstream of the aberrant branchpoint was also necessary for mutant SF3B1 missplicing.

Our data demonstrated that *SF3B1* mutation can downregulate expression of both *PPP2R5A* and *MAP3K7* simultaneously and thereby promote AKT and NF- $\kappa$ B signaling output with result-



**Figure 6. Aberrant splicing and downregulation of PPP2R5A by mutant SF3B1 promotes AKT activation and inflammatory signaling in breast cancer.** Representative RNA-seq coverage plots showing cryptic 3' splice site usage in *PPP2R5A* in (A) human isogenic breast cancer cell lines (MCF10A, MCF7, T47D) expressing WT or SF3B1 K700E (K700E) and human breast tumor samples from TCGA (22) as well as (B) mouse mammary epithelial cells from *MMTV-cre* control, *MMTV-cre Sf3b1<sup>K700E/WT</sup>*, *MMTV-cre R26-LSL-Pik3ca<sup>H1047R</sup>* transgenic, and compound *MMTV-cre Sf3b1<sup>K700E/WT</sup> R26-LSL-Pik3ca<sup>H1047R</sup>* mice. PSI value of each event is shown on the right of each coverage plot. (C) *PPP2R5A* mRNA expression (normalized log<sub>2</sub>RPKM values) in SF3B1 mutant (MUT) versus WT primary human breast tumors from TCGA and cell lines shown in A. Box indicates upper and lower quartiles; thick bar indicates the median value. Whiskers indicate the largest/smallest value no further than 1.5 times the interquartile range. (D) *PPP2R5A*, phospho- and total AKT, and p65 levels in MCF7 and MCF10A with and without mutant SF3B1 expression (EV, empty vector). Quantitation of 3 independent experiments is shown in Supplemental Figure 10A. (E) RT-PCR of *PPP2R5A* splicing event derived from *PPP2R5A* minigene with mutations at aberrant branchpoint or polypyrimidine region downstream of the aberrant 3' splice site. RT-PCR of *PPP2R5A* splicing event derived from endogenous *PPP2R5A* is shown on the bottom. A schematic of sequences around the normal and aberrant branchpoint and splice site is shown on top. (F) *PPP2R5A*, phospho- and total AKT, and p65 levels in MCF7 and MCF10A with and without *PPP2R5A* shRNA knockdown. Quantitation of 3 independent experiments is shown in Supplemental Figure 10B. (G) Relative mRNA expression of *IL1A*, *CXCL1*, *IL8*, *CDKN1A*, and *PPP2R5A* in the cells from D and F. Mean ± standard deviation shown; asterisks represent *P* values of Dunnett's test compared with sh*Renilla* control (shRE). See also Supplemental Figures 9 and 10.

ing increases in breast cancer migration/invasion. Akin to *SF3B1* mutation in hematopoietic stem cells (13, 29) as well as B cells (33, 34), *SF3B1* mutation alone does not appear sufficient to mediate tumorigenesis in the mammary gland. However, when mutations in *SF3B1* and *PIK3CA* were paired, major acceleration in tumor phenotypes was observed, with the double mutants significantly accelerating tumor formation and lethality. Coinciding with this observation from the mice was the finding that patients with mutations in both *SF3B1* and *PIK3CA* had an inferior prognosis compared with WT patients.

We speculate that addition of the *SF3B1* mutation to a *PIK3CA* mutant model can serve to reinforce signals initially elicited by PI3K activation. It is well established that PI3K activation promotes feedback suppression of the pathway through a variety of mechanisms such as suppression of RTK signaling (56) as well as inhibition of parallel pathways such as the NF- $\kappa$ B pathway. Given the widespread mechanisms of feedback suppression of signaling as a result of oncogenic PI3K signaling, there may be stronger selection for genomic alterations that coordinately remodel multiple signaling pathways akin to changes in the epigenome or transcription factors.

Finally, our functional interrogation of the misspliced events in *SF3B1* mutant breast cancers nominated specific pathways for therapy of these tumors that might not emerge as obvious from the genotypes alone. We noted that the hyperactivation of AKT signaling that was induced as a result of combined *PIK3CA* and *SF3B1* mutation might engender further therapeutic significance. Recent work has demonstrated that multiple mutations in *PIK3CA* can lead to increases in AKT output that ultimately increases the sensitivity of these tumors to PI3K/AKT inhibition (57). The alteration in *PPP2R5A* results in changes that serve to further activate AKT signaling and, in like manner, appeared to

increase the sensitivity of these tumors to AKT kinase inhibitors. AKT inhibitors are currently under study in ER<sup>+</sup> breast cancers harboring AKT E17K mutations, and our data suggest that *SF3B1* mutation could provide an additional patient subset that may be sensitized to these therapies.

## Methods

**Human subjects.** A total of 1918 breast tumor specimens from 1756 patients with breast cancer who underwent prospective clinical genomic profiling were included for study. The demographic and clinical characteristics of the cohort are presented in Supplemental Table 2.

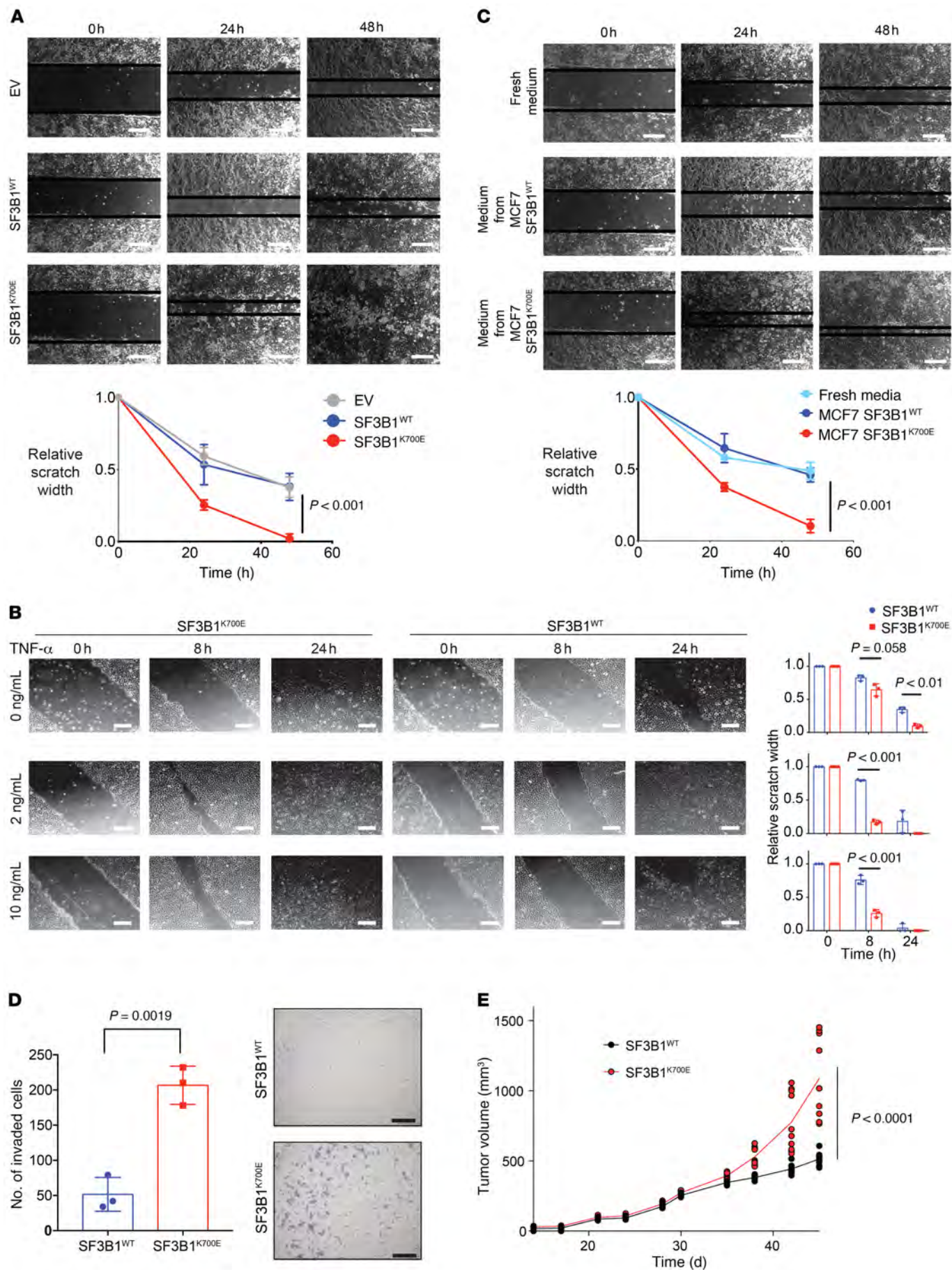
**Mutation oncoprint analysis.** The mutual mutational pattern of *SF3B1*, *PIK3CA*, *PTEN*, *AKT1*, *TP53*, *MAP3K1*, *ESR1*, *NF1*, *PIK3R1*, *BRCA1*, *ERBB3*, and *MYC* was investigated on 3 breast cancer cohorts: METABRIC, MSK-IMPACT, and TCGA (Figure 1A and Supplemental Figure 1). The analysis was performed using the online tool from the cBioportal website (<https://www.cbioportal.org/oncoprinter>). The raw input files are provided in Supplemental Table 1.

**Mutation and allele frequency analysis.** We compared the allele frequency of *SF3B1* and *PIK3CA* mutations in 23 double-mutated samples from the MSK-IMPACT breast cancer cohort. Variant allele frequency (VAF) was corrected by tumor purity estimated by immunohistochemistry. Wilcoxon rank sum test was applied to give statistics between the VAF of 2 mutations. We analyzed the frequency of somatic mutations in patients from the MSK-IMPACT cohort ( $n = 41$ ) harboring *SF3B1* hotspot mutations. Mutation frequency was calculated for each reported gene in 25 primary samples and 20 metastasis samples.

**Animals.** All mice were housed at Memorial Sloan Kettering Cancer Center (MSKCC). Generation and genotyping of the *Sf3b1<sup>K700E/WT</sup>* and *R26-LSL-Pik3ca<sup>H1047R</sup>* as well as the *MMTV-cre* transgenic mice have been previously described (13, 32). The mice were monitored closely for signs of disease or morbidity daily and were killed when they showed a volume of the visible tumor formation above 1 cm<sup>3</sup>, failed to thrive, experienced weight loss greater than 10% total body weight, or showed open skin lesions, bleeding, or any signs of infection.

**Whole mount staining of mouse mammary gland.** The fourth pair of mammary glands were excised from 4-month-old female mice and stained as previously described (58). Briefly, mammary glands were placed flattened on glass slides and fixed with Carnoy's fixative (ethanol/chloroform/acetic acid, 6:3:1) at room temperature for 4 hours. Then the glands were gradually rehydrated with 70%, 35%, 17.5% ethanol, and distilled water. The mammary glands were stained in carmine alum solution (0.2% carmine natural red [MilliporeSigma], 0.5% aluminum potassium sulfate [MilliporeSigma] in water) overnight at room temperature, and then dehydrated progressively in 70%, 95%, 100% ethanol. The mammary fat pads were then cleared in xylene overnight. The slides were mounted with Permount (Thermo Fisher Scientific) and scanned by Panoramic Scanner (3DHISTECH). The mammary epithelial area (MEA) and the number of branching points (center) were analyzed using FIJI software as previously described (59). The number of terminal buds were counted manually in 10 random 2 mm × 2 mm squares on each slide with QuPath software (60) and presented as number of terminal bud per square millimeter.

**Immunohistochemistry.** Mouse tissue was collected and fixed in 10% neutral buffered formalin for 48 hours. Samples were then



**Figure 7. *SF3B1* mutations promote migration and invasion of breast cancer cells.** (A) Representative photos (top) at 0, 24, and 48 hours following wound creation in cell culture of MCF7 cells expressing an empty vector (EV), *SF3B1*<sup>WT</sup>, or *SF3B1*<sup>K700E</sup> cDNA. Scale bars: 200  $\mu$ M. Relative scratch width over time shown below ( $n = 3$ ; mean value  $\pm$  standard deviation shown below;  $P$  values by ANOVA). (B) Quantification and representative photos (left) of wound at 0, 8, and 24 hours following wound creation and exposure to 0, 2, and 10 ng/mL TNF- $\alpha$  in cell culture of MCF10A with or without knockin of *SF3B1*<sup>K700E</sup>. Right, mean levels with standard deviation shown ( $n = 3$ ;  $P$  value by Student's  $t$  test). Scale bars: 500  $\mu$ M. (C) Same as A but for parental MCF7 cells exposed to medium from cells in A. Scale bars: 200  $\mu$ M. (D) Number of invaded MCF10A cells with or without knockin of *SF3B1*<sup>K700E</sup> migrating toward EGF in a Transwell invasion assay (left). Mean  $\pm$  standard deviation ( $n = 3$ ,  $P$  value by Student's  $t$  test) shown. Right, photos of hematoxylin staining of the migrated cells (photos taken using a microscope with a 10 $\times$  objective (total magnification  $\times$ 100); scale bars: 200  $\mu$ M). (E) Growth of MCF7 tumors expressing either WT or mutant *SF3B1* in NSG mice. Size of individual tumor ( $n = 10$  in each group) and the mean are shown ( $P$  value by ANOVA). See also Supplemental Figures 11 and 12.

processed, embedded in paraffin, sectioned at 4  $\mu$ m thickness, and stained with hematoxylin and eosin by the Laboratory of Comparative Pathology at MSKCC. Immunohistochemical staining was performed on a Ventana Stains System with primary antibodies (Supplemental Table 7) and OmniMap DAB anti-Rb Detection Kit or OmniMap DAB anti-Ms Detection Kit (Ventana). The slides were scanned by Panoramic Scanner (3DHISTECH) and positively stained cell number was counted with QuPath software (60).

**Cell lines.** Cell lines were maintained in 37°C incubator with 5% CO<sub>2</sub> in a humidified atmosphere. MCF7 and MCF10A cells were obtained from ATCC. HEK293T cells were cultured in DMEM with 10% FBS (Gibco). MCF7 cells were grown in DMEM/F12 supplemented with 10% FBS (Gibco), 100  $\mu$ g/mL penicillin, 100 mg/mL streptomycin (Corning), and 4 mM glutamine, and MCF10A cells were grown in DMEM/F12 supplemented with 5% horse serum (Gibco), 20 ng/mL EGF (MilliporeSigma), 10  $\mu$ g/mL insulin (MilliporeSigma), 0.5  $\mu$ g/mL hydrocortisone (MilliporeSigma), 0.1  $\mu$ g/mL cholera toxin (MilliporeSigma), 100  $\mu$ g/mL penicillin, and 100 mg/mL streptomycin (Corning). T47D cells were cultured in RPMI1640 with 10% FBS (Gibco), 100  $\mu$ g/mL penicillin, and 100 mg/mL streptomycin (Corning). For cell proliferation assay, 1000 cells were seeded in each well of 96-well plates, and proliferation was measured with resazurin reagent (Thermo Fisher Scientific).

MCF10A *SF3B1*<sup>K700E/WT</sup> knockin cell lines were generated by Horizon Discovery. MCF10A *PIK3CA*<sup>H1047R</sup> single and *PIK3CA*<sup>H1047R</sup> *SF3B1*<sup>K700E/WT</sup> double knockin cells were generated by CRISPR and homology-directed repair (HDR) as previously described (11). Briefly, guide RNA sequence was cloned into px458-GFP vector (Addgene 48138) and nucleofected into MCF10A WT or *SF3B1*<sup>K700E/WT</sup> cells together with a single-stranded donor oligonucleotide (ssODN) as a template for HDR using the Lonza Nucleofector V kit and program T-024 on the nucleofector device. The ssODN sequence was CGAAAGACCCTAGCCTTAGATAAACTGAGCAAGAGGCTTTG-GAGTATTTTCATGAAACAAAATGAACGACGCGCGTCACGGT-GGCTGGACAACAAAATGGATTGGATCTTCCACACAATTA-AACAGCATGCATGAACTGA. Additional silent mutations were introduced with H1047R to avoid Cas9 recognition on the knockin allele. GFP-positive nucleofected cells were single-cell sorted 48 hours after nucleofection. Clones were screened for knockin of *PIK3CA*<sup>H1047R</sup>

by PCR with primers GTTTCAGGAGATGTGTTACAAG and GTGT-GGAATCCAGAGTGAGC and subsequent Sanger sequencing.

**Plasmids.** The *SF3B1*-expressing lentiviral vector was generated by cloning cDNA of WT *SF3B1* or *SF3B1* K700E with flag tagged at N-terminal end into pInducer20 (Addgene 44012). Lentivirus was prepared in HEK293T cells with pInducer20 empty vector or the plasmid with *SF3B1* (WT or mutant) cDNA, and then used to transduce breast cell lines. Transduced cells were selected in medium supplemented with 1 mg/mL Geneticin (Invivogen) for 2 weeks before being used in the experiments.

shRNAs against PPP2R5A or renilla luciferase were cloned into MLS-E-GFP vector. The shRNA sequences were PPP2R5A-1: TTAGTTGAAACATACTCAACCA; PPP2R5A-2: TTTAATTATAT-TACTGATGA; Renilla: CAGGAATTATAATGCTTATCTA. Retrovirus was prepared in HEK293T cells and then used to transduce MCF7 or MCF10A cells. GFP<sup>+</sup> cells stably expressing shRNAs were selected by FACS 3 days after transduction.

shRNAs against MAP3K7 or renilla luciferase were cloned into LT3GEPiR vector. The shRNA sequences were MAP3K7-1: TTAGG-TAAATTTTTTATCAGTG and MAP3K7-2: TTTTCAACAATTTT-GATTCTAA. Lentivirus was prepared in HEK293T cells and then used to transduce MCF7 and MCF10A cells. Transduced cells were selected in medium supplemented with 1  $\mu$ g/mL puromycin (Thermo Fisher Scientific) for 1 week before being used in the experiments.

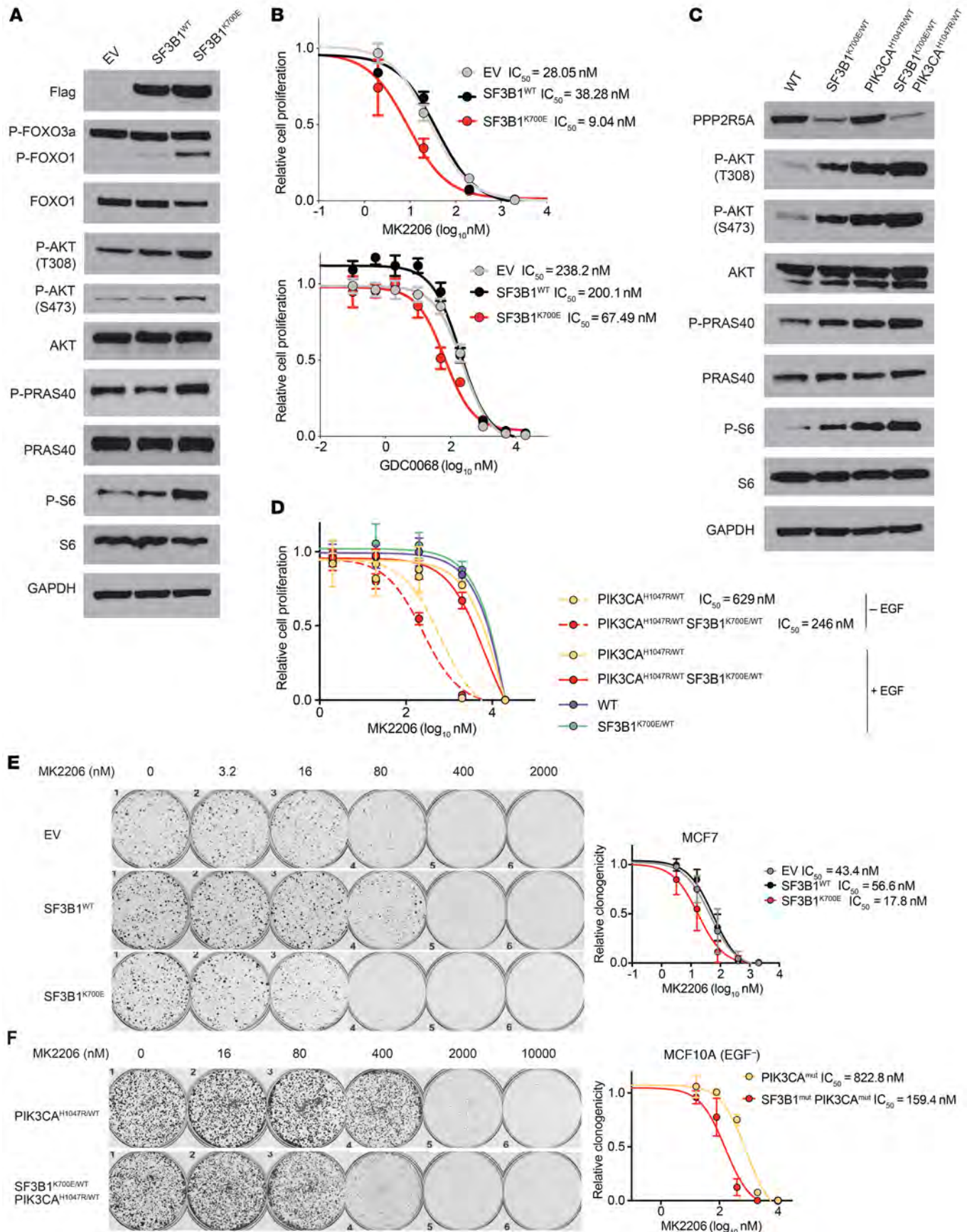
The MAP3K7 and PPP2R5A expression vector was generated by cloning *MAP3K7* or *PPP2R5A* cDNA into pMSCV-GFP retroviral vector to express MAP3K7 or PPP2R5A protein. Retrovirus was prepared in HEK293T cells and then used to transduce MCF7 or MCF10A cells. GFP<sup>+</sup> cells stably expressing MAP3K7 or PPP2R5A were selected by FACS 3 days after transduction.

**Antibodies.** The antibodies used are listed in Supplemental Table 7.

**Western blotting.** Protein samples were extracted from cultured cells with RIPA buffer (Thermo Fisher Scientific) and quantified by BCA assay. Protein fractionated on NuPAGE 4%–12% Bis-Tris gels (Life Technologies) was transferred onto nitrocellulose membranes. All primary antibodies (Supplemental Table 7) were used with 1:1000 dilution, except for GAPDH antibody, which was used at 1:5000 dilution in TBST containing 5% BSA and 0.02% sodium azide. Peroxidase-conjugated goat anti-rabbit IgG secondary antibody (MilliporeSigma) was used at 1:10000 dilution in TBST containing 2.5% skim milk. The blot was then visualized by chemiluminescence. Developed films were scanned and quantitated with FIJI software.

**RT-PCR and quantitative RT-PCR.** Total RNA was extracted with the QIAGEN RNeasy Mini Kit, and cDNA was generated with qScript cDNA SuperMix (Quantabio) according to the manufacturers' instructions. For RT-PCR, the amplification was performed with DreamTaq Hot Start Green PCT Master Mix (Thermo Fisher Scientific) and specific primers. PCR products were fractionated on 10% TBE gel (Life Technologies) and visualized with ethidium bromide. For quantitative RT-PCR, 0.5  $\mu$ L cDNA was added to 10  $\mu$  PCR reactions prepared with TaqMan Universal PCR Master Mix (Thermo Fisher Scientific) and gene-specific TaqMan probes. All quantitative RT-PCR analysis was performed on an Applied Biosystems ViiA 7 Real-Time PCR System. Relative gene expression levels were calculated using the comparative CT method. Primers and TaqMan probes used in this study are listed in Supplemental Table 8.

**Cell migration assay.** About  $1 \times 10^5$  cells were seeded into each well of 12-well plates and cultured until 100% confluence. The "scratch"



**Figure 8. *SF3B1*<sup>K700E</sup> mutation sensitizes breast cancer cells to AKT inhibitors.** (A) Western blot of MCF7 cells expressing an empty vector (EV), *SF3B1* WT, or *SF3B1* K700E cDNA. (B) IC<sub>50</sub> of MCF7 cells shown in A following 5 days of exposure to MK2206 (left) or GDC0068 (right). Data are represented as mean ± SD; n = 6. (C) Western blot of MCF10A WT cells as well as *SF3B1*<sup>K700E/WT</sup> knockin cells, *PIK3CA*<sup>H1047R/WT</sup> knockin cells, and *SF3B1*<sup>K700E/WT</sup> *PIK3CA*<sup>H1047R/WT</sup> double knockin cells cultured in the absence of EGF. (D) IC<sub>50</sub> to MK2206 of MCF10A cells shown in C in the presence or absence of EGF. Data are represented as mean ± SD; n = 6. (E) Clonogenic survival assay of MCF7 cells shown in A following 21 days of exposure to MK2206. IC<sub>50</sub> curve is shown at right. (F) Clonogenic survival assay of MCF710A cells shown in C following 14 days of exposure to MK2206 in absence of EGF. IC<sub>50</sub> curve is shown at right. See also Supplemental Figures 13 and 14.

was created with a p200 pipet tip on the cell monolayer through the center of the well. The debris were removed by washing the well with 1 mL culture medium and then 1 mL medium (with 2% serum) specific for the cell migration assay was added into each well. The plate was incubated at 37°C, and images of the scratch were captured at various time points with a Zeiss Axiovert Microscope. The width of scratch was measured with FIJI software.

**Cell invasion assay.** MCF10A cells were cultured to about 70% confluence and serum starved overnight. Harvested by trypsinization, 5 × 10<sup>4</sup> cells resuspended in serum-free medium (without EGF) were seeded to the Matrigel invasion chamber (8 μm pore size, Corning) in a 24-well plate. Complete culture medium for MCF10A was added to the bottom chambers of each well. After 24-hour incubation, the cells at the bottom side of the chamber were fixed and stained with hematoxylin. The membrane was then cut off from the chamber and the number of migrating cells were quantified by counting cells under microscope.

**Cytokine multiplex assay.** Forty-one cytokines and chemokines were analyzed simultaneously with bead-based multiplex assay using Milliplex Map Human Cytokine/Chemokine Magnetic Bead Panel (HCYTMAP-60K-PX41, Millipore) on the Luminex FlexMap 3D instrument according to the manufacturer's instruction. Briefly, cells were seeded in a T25 flask and media were refreshed when the cells approached 70% confluence. Cell culture supernatant was harvested 48 hours after media change and multiplex cytokine assay was performed on the Luminex FlexMap 3D instrument. The results were normalized to cell numbers in each flask.

**IL-8 ELISA.** IL-8 concentration in the culture supernatant was quantified by human IL-8 ELISA kit (RayBiotech) according to the manufacturer's instruction. Briefly, cells were seeded in a T25 flask and culture medium was changed when the cells approached 70% confluence. Cell culture supernatant was harvested 48 hours after media change and IL-8 ELISA was performed on SpectraMax M5 microplate reader (Molecular Devices). The results were normalized to cell numbers in each flask.

**Cell proliferation.** Cell proliferation was measured as previously described (61). Briefly, cells (1000/well) were seeded into 96-well plates and treated with different concentrations of indicated drugs. Media and drug were changed every week. On days 0, 3, 5, and 7, 25 μL Resazurin (R&D Systems) was added to each well and incubated for 4 hours. Fluorescence in the plate was measured using a SpectraMax M5 microplate reader (Molecular Devices). Cell growth data from day 5 were then plotted as percentage inhibition against the log concentration of indicated drugs. IC<sub>50</sub> was determined using a sigmoidal regression model using GraphPad Prism 7.0.

**Minigene assay.** The *PPP2R5A* minigene construct was generated by inserting the DNA fragment containing the *PPP2R5A* genomic sequence from exon 4 to exon 5 in between the KpnI and XhoI restriction sites of pcDNA3.1 vector. The *MAP3K7* minigene construct was generated by inserting the DNA fragment containing the *MAP3K7* genomic sequence from exon 4 to exon 5 at HindIII restriction site of pcDNA3.1 vector. The sequences of the inserted fragments were verified by Sanger sequencing. Mutagenesis of minigene constructs was performed with the Agilent QuikChange II site-directed mutagenesis kit with specific primers according to the manufacturer's directions. For transient transfection experiments, cells were seeded into a 24-well plate with culture medium containing doxycycline 48 hours before transfection of minigene constructs in the presence of X-tremeGENE HP DNA transfection reagent (Roche) according to the manufacturer's directions. Forty-eight hours after transfection, cells were collected and RNA was extracted using the QIAGEN RNeasy Mini Kit. Minigene-derived and endogenous *MAP3K7* and *PPP2R5A* transcripts were analyzed by RT-PCR using specific primers (Supplemental Table 8).

**Clonogenic survival assay.** Clonogenic survival assay was performed on a 6-well plate. Five hundred cells in single cell suspension were seeded into each well of a 6-well plate and treated with various doses of MK2206 on the next day, following which the plate was incubated at 37°C with 5% CO<sub>2</sub> in a humidified atmosphere for 14 days (for MCF10A cells) or 21 days (for MCF7 cells). After incubation, the colonies were fixed with formalin and stained with 0.01% crystal violet. The plate was then photographed and analyzed with the Oxford Gel-Count colony counter.

**Xenograft assay.** NOD-SCID gamma (NSG) female mice ages 6–8 weeks were obtained from Jackson Laboratory and maintained in pressurized ventilated caging. MCF7 xenograft tumors expressing WT or mutant *SF3B1* were established in NSG mice by subcutaneously implanting 0.72 mg sustained release 17β-estradiol pellets with a 10 g trocar into one flank followed by injecting 1 × 10<sup>6</sup> cells suspended 1:1 (vol/vol) with reconstituted basement membrane (Matrigel, Collaborative Research) on the opposite side 3 days later. On the same day of cell inoculation, the mice were fed with filter water containing 2 mg/mL doxycycline and 2% sucrose, which was refreshed twice a week. Tumor dimensions were measured with vernier calipers starting from 14 days after cell implantation and tumor volumes were calculated as (π/6 × larger diameter × (smaller diameter)<sup>2</sup>). The experiment was closed 7 weeks after implantation. Based on our previous work measuring the variability in size and growth of MCF7 xenografts, we estimated 10 mice per group would allow us to detect tumor size differences of greater than 200 mm<sup>3</sup>.

**RNA-seq of breast cancer cell lines and mouse tissues.** Total RNA was extracted with the Direct-Zol RNA miniprep kit (Zymo). RNA-seq libraries were prepared and sequenced by Genewiz with 150-bp paired reads with 100 million read depth.

**Heatmap and hierarchical clustering.** To explore potential tumor/hotspot specificity or similarity of 3' ss alteration, principal component analysis and unsupervised hierarchical clustering were performed using 618 significantly (*q* value < 0.001) differential splicing junctions between all *SF3B1* mutant and all WT samples (see Supplemental Methods for novel cryptic 3' ss identification). Command `prcomp` from R package `stats` was used for the principal component analysis. Command `heatmap.2` from R package `gplots` was adopted for the clus-

tering analysis using  $(1 - \text{Pearson correlation coefficient})/2$  as distance. This analysis was performed on TCGA BRCA, MSK breast cancer cell lines (MCF7, T47D and MCF10A), ICGC CLL, dbGaP CLL, and TCGA UVM samples.

**Sequence motif analysis.** We performed motif analysis at aberrant 3' ss as well as the associated canonical 3' ss. We first identified sets of cryptic 3' ss significantly differentially used between *SF3B1* mutant and WT samples by combining TCGA breast cancer samples and MSK cell line data. Next, motif logo was plotted based on the intron nucleotide composition up to 50 nt upstream and 20 nt downstream from the AG dinucleotide of cryptic 3' ss as well as the associated canonical 3' AG site, using R package motifStack. Fisher's exact test was performed to compare the enrichment of pyrimidines in the 10 nt downstream from the AG of cryptic 3' ss to the associated canonical 3' AG site.

**GSEA analysis.** We used GSEA software (62) with defined hallmark genesets (MSigDB, v6.1) to investigate molecular profiles enriched in *SF3B1* mutant breast cancer samples. Parameters used were as follows: collapse data: false; permutation type: gene set, 1000 permutations; chip platform: gene symbol. The input gene expression matrix contains 10,165 genes, which are significantly differentially expressed (DE) between *SF3B1* mutations and WT on cell line data (combining the DE lists from MCF7, T47D, and MCF10A).

**Expression data set and differential expression.** To generate mRNA expression matrix for transcriptome analysis, we used featureCounts (63) from package Subread to call read counts from STAR realigned bam files. Genes with low read depths across the cohort were removed. Then, read counts were transformed into RPKM values followed by  $\log^2$  transformation, and quantile normalized on the sample level. We also performed DESeq2 analysis (64) (using the raw read counts) to call DE genes between *SF3B1* mutant and WT TCGA breast cancer samples and MSK cell line data.

**Data availability.** The RNA sequencing data were deposited in the NCBI Sequence Read Archive (SRA) under accession number GSE145471.

**Statistics.** Statistical significance was determined by (a) unpaired or paired 2-sided Student's *t* test after testing for normal distribution,

(b) 1-way or 2-way ANOVA followed by Dunnett's or Tukey's multiple comparisons tests, or (c) Kruskal-Wallis tests with Dunn's test corrected for multiple comparisons. Data were plotted using GraphPad Prism 7 software as individual points with mean values and error bars representing standard deviation.

**Study approval.** This study was approved by the Memorial Sloan Kettering Cancer Center Institutional Review Board (IRB) and all patients provided written informed consent for tumor sequencing and review of patient medical records for detailed demographic, pathologic, and treatment information (NCT01775072). All mouse procedures and experiments were completed in accordance with the Guidelines for the Care and Use of Laboratory Animals and were approved by the Institutional Animal Care and Use Committees at MSKCC (MSKCC#12-10-016).

## Author contributions

BL, ZL, RR, OAW, and S. Chandralapaty designed the study. BL, S. Chen, MK, CE, QC, and EDS performed the animal experiments. BHD and JSRF assisted with pathological analysis. BL, ZL, YS, and RR performed RNA-seq analyses. BL, OAW, and S. Chandralapaty performed in vitro cell line experiments. BL, ZL, OAW, and S. Chandralapaty wrote the manuscript.

## Acknowledgments

OAW and S. Chandralapaty are supported by the Department of Defense Breast Cancer Research Program (W81XWH-18-1-0383), the Functional Genomics Initiative of Memorial Sloan Kettering Cancer Center, and the NCI Cancer Center Support Grant (CCSG, P30 CA08748). S. Chandralapaty acknowledges support by the Breast Cancer Research Foundation for this work.

Address correspondence to: Sarat Chandralapaty or Omar Abdel-Wahab, Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, 417 East 68th Street, New York, New York 10065, USA. Phone: 646.888.4311; Email: chandars@mskcc.org (SC). Phone: 646.888.3487; Email: abdelwao@mskcc.org (OAW).

- Seiler M, et al. Somatic mutational landscape of splicing factor genes and their functional consequences across 33 cancer types. *Cell Rep.* 2018;23(1):282–296.e4.
- Yoshida K, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature.* 2011;478(7367):64–69.
- Wang L, et al. *SF3B1* and other novel cancer genes in chronic lymphocytic leukemia. *N Engl J Med.* 2011;365(26):2497–2506.
- Papaemmanuil E, et al. Somatic *SF3B1* mutation in myelodysplasia with ring sideroblasts. *N Engl J Med.* 2011;365(15):1384–1395.
- Graubert TA, et al. Recurrent mutations in the *U2AF1* splicing factor in myelodysplastic syndromes. *Nat Genet.* 2011;44(1):53–57.
- Suzuki H, et al. Recurrent noncoding *U1* snRNA mutations drive cryptic splicing in *SHH* medulloblastoma. *Nature.* 2019;574(7780):707–711.
- Shuai S, et al. The *U1* spliceosomal RNA is recurrently mutated in multiple cancers. *Nature.* 2019; 574(7780):712–716.
- Darman RB, et al. Cancer-associated *SF3B1* hotspot mutations induce cryptic 3' splice site selection through use of a different branch point. *Cell Rep.* 2015;13(5):1033–10345.
- Kim E, et al. *SRSF2* mutations contribute to myelodysplasia by mutant-specific effects on exon recognition. *Cancer Cell.* 2015;27(5):617–630.
- Ilagan J, et al. *U2AF1* mutations alter splice site recognition in hematological malignancies. *Genome Research.* 2015;25(1):14–26.
- Inoue D, et al. Spliceosomal disruption of the non-canonical BAF complex in cancer. *Nature.* 2019;574(7778):432–436.
- Yoshimi A, et al. Coordinated alterations in RNA splicing epigenetic regulation drive leukaemogenesis. *Nature.* 2019;574(7777):273–277.
- Obeng EA, et al. Physiologic expression of *Sf3b1*(K700E) causes impaired erythropoiesis, aberrant splicing, and sensitivity to therapeutic spliceosome modulation. *Cancer Cell.* 2016;30(3):404–417.
- Wang L, et al. Transcriptomic characterization of *SF3B1* mutation reveals its pleiotropic effects in chronic lymphocytic leukemia. *Cancer Cell.* 2016;30(5):750–763.
- DeBoever C, et al. Transcriptome sequencing reveals potential mechanism of cryptic 3' splice site selection in *SF3B1*-mutated cancers. *PLoS Comput Biol.* 2015;11(3):e1004105.
- Seiler M, et al. *H3B-8800*, an orally available small-molecule splicing modulator, induces lethality in spliceosome-mutant cancers. *Nature Med.* 2018;24(4):497–504.
- Dalton WB, et al. Hotspot *SF3B1* mutations induce metabolic reprogramming and vulnerability to serine deprivation. *J Clin Invest.* 2019;129(11):4708–4723.
- Drilon A, et al. Efficacy of arotrectinib in *TRK* fusion-positive cancers in adults and children. *N Engl J Med.* 2018;378(8):731–739.
- Pratils CA, et al. (V600E)*BRAF* is associated with disabled feedback inhibition of *RAF-MEK* signaling and elevated transcriptional output of the pathway. *Proc Natl Acad Sci U S A.* 2009;106(11):4519–4524.
- Hyman DM, et al. Vemurafenib in multiple non-melanoma cancers with *BRAF* V600 mutations.

- N Engl J Med.* 2015;373(8):726–736.
21. Diamond EL, et al. Efficacy of MEK inhibition in patients with histiocytic neoplasms. *Nature.* 2019;567(7749):521–524.
  22. Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours. *Nature.* 2012;490(7418):61–70.
  23. Curtis C, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature.* 2012;486(7403):346–352.
  24. Razavi P, et al. The genomic landscape of endocrine-resistant advanced breast cancers. *Cancer Cell.* 2018;34(3):427–438.e6.
  25. Chang MT, et al. Accelerating discovery of functional mutant alleles in cancer. *Cancer Discov.* 2018;8(2):174–183.
  26. Toy W, et al. ESR1 ligand-binding domain mutations in hormone-resistant breast cancer. *Nat Genet.* 2013;45(12):1439–1445.
  27. Bertucci F, et al. Genomic characterization of metastatic breast cancers. *Nature.* 2019;569(7757):560–564.
  28. Bielski CM, et al. Widespread selection for oncogenic mutant allele imbalance in cancer. *Cancer Cell.* 2018;34(5):852–862.e4.
  29. Lee SC, et al. Synthetic lethal and convergent biological effects of cancer-associated spliceosomal gene mutations. *Cancer Cell.* 2018;34(2):225–241.e8.
  30. Zhou Q, et al. A chemical genetics approach for the functional assessment of novel cancer genes. *Cancer Res.* 2015;75(10):1949–1958.
  31. Obeng EA, et al. Mutant splicing factor 3b subunit 1 (SF3B1) causes dysregulated erythropoiesis and a stem cell disadvantage. *Blood.* 2014;124(21):828.
  32. Adams JR, et al. Cooperation between Pik3ca and p53 mutations in mouse mammary tumor formation. *Cancer Res.* 2011;71(7):2706–2717.
  33. Yin S, et al. A murine model of chronic lymphocytic leukemia based on B cell-restricted expression of sf3b1 mutation and atm deletion. *Cancer Cell.* 2019;35(2):283–296.e5.
  34. Liu Z, et al. Mutations in the RNA splicing factor SF3B1 promote tumorigenesis through MYC stabilization. *Cancer Discov.* 2020;10(6):806–821.
  35. Koren S, et al. PIK3CA(H1047R) induces multipotency and multi-lineage mammary tumours. *Nature.* 2015;525(7567):114–118.
  36. Wagner KU, et al. Cre-mediated gene deletion in the mammary gland. *Nucleic Acids Res.* 1997;25(21):4323–4330.
  37. Darman RB, et al. Cancer-associated SF3B1 hotspot mutations induce cryptic 3' splice site selection through use of a different branch point. *Cell Rep.* 2015;13(5):1033–1034.
  38. Sato S, et al. Essential function for the kinase TAK1 in innate and adaptive immune responses. *Nat Immunol.* 2005;6(11):1087–1095.
  39. Singh A, et al. TAK1 inhibition promotes apoptosis in KRAS-dependent colon cancers. *Cell.* 2012;148(4):639–650.
  40. Ajibade AA, et al. TAK1 negatively regulates NF- $\kappa$ B and p38 MAP kinase activation in Gr-1+CD11b+ neutrophils. *Immunity.* 2012;36(1):43–54.
  41. Wu J, et al. Mechanism and in vitro pharmacology of TAK1 inhibition by (5Z)-7-oxozeaenol. *ACS Chem Biol.* 2013;8(3):643–650.
  42. Helbig G, et al. NF-kappaB promotes breast cancer cell migration and metastasis by inducing the expression of the chemokine receptor CXCR4. *J Biol Chem.* 2003;278(24):21631–21638.
  43. Sovak MA, et al. Aberrant nuclear factor-kappaB/Rel expression and the pathogenesis of breast cancer. *J Clin Invest.* 1997;100(12):2952–2960.
  44. Cogswell PC, Guttridge DC, Funkhouser WK, Baldwin AS. Selective activation of NF-kappa B subunits in human breast cancer: potential roles for NF-kappa B2/p52 and for Bcl-3. *Oncogene.* 2000;19(9):1123–1131.
  45. Slamon DJ, et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med.* 2001;344(11):783–792.
  46. Baselga J, et al. Pertuzumab plus trastuzumab plus docetaxel for metastatic breast cancer. *N Engl J Med.* 2012;366(2):109–119.
  47. André F, et al. Alpelisib for PIK3CA-mutated, hormone receptor-positive advanced breast cancer. *N Engl J Med.* 2019;380(20):1929–1940.
  48. Baselga J, et al. Everolimus in postmenopausal hormone-receptor-positive advanced breast cancer. *N Engl J Med.* 2012;366(6):520–529.
  49. Finn RS, et al. Palbociclib and letrozole in advanced breast cancer. *N Engl J Med.* 2016;375(20):1925–1936.
  50. Ciriello G, et al. Comprehensive molecular portraits of invasive lobular breast cancer. *Cell.* 2015;163(2):506–519.
  51. Chang MT, et al. Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. *Nat Biotechnol.* 2016;34(2):155–163.
  52. Le DT, et al. PD-1 Blockade in tumors with mismatch-repair deficiency. *N Engl J Med.* 2015;372(26):2509–2520.
  53. Robson M, et al. Olaparib for metastatic breast cancer in patients with a germline BRCA mutation. *N Engl J Med.* 2017;377(6):523–533.
  54. Lehmann BD, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J Clin Invest.* 2011;121(7):2750–2767.
  55. Gucalp A, et al. Phase II trial of bicalutamide in patients with androgen receptor-positive, estrogen receptor-negative metastatic Breast Cancer. *Clin Cancer Res.* 2013;19(19):5505–5512.
  56. Chandraratna S. Negative feedback and adaptive resistance to the targeted therapy of cancer. *Cancer Discov.* 2012;2(4):311–319.
  57. Vasan N, et al. Double PIK3CA mutations in cis increase oncogenicity and sensitivity to PI3Ka inhibitors. *Science.* 2019;366(6466):714–723.
  58. Plante I, Stewart MK, Laird DW. Evaluation of mammary gland development and function in mouse models. *J Vis Exp.* 2011;(53):2828.
  59. Stanko JP, Fenton SE. Quantifying branching density in rat mammary gland whole-mounts using the sholl analysis method. *J Vis Exp.* 2017;(125):55789.
  60. Bankhead P, et al. QuPath: open source software for digital pathology image analysis. *Sci Rep.* 2017;7(1):16878.
  61. Li Z, et al. Loss of the FAT1 tumor suppressor promotes resistance to CDK4/6 inhibitors via the Hippo pathway. *Cancer Cell.* 2018;34(6):893–905.e8.
  62. Subramanian A, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005;102(43):15545–15550.
  63. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics.* 2014;30(7):923–930.
  64. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550.
  65. Pereira B, et al. The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. *Nat Commun.* 2016;7:11479.

## SUPPLEMENTAL METHODS

### Pan-cancer *SF3B1* mutated sample collection and sequencing

For all 1,756 patients, tumor and patient-matched normal DNA samples were extracted from either representative formalin-fixed paraffin embedded (FFPE) tumor biopsy samples or mononuclear cells from peripheral blood. All specimens underwent massively parallel next-generation sequencing in a CLIA-certified laboratory using MSK-IMPACT, a hybridization capture-based next-generation sequencing assay, which analyzes all protein-coding exons and selected intronic and regulatory regions of 341 to 468 cancer-associated genes, all as previously described (1-3). Average sequencing coverage across all tumors was 743-fold. Somatic mutations, DNA copy number alterations, and structural rearrangements were identified as previously described (4). Purity, average ploidy, and allele-specific integer-copy number for each segment were then determined by maximum likelihood. Loss of heterozygosity (LOH) was determined based on allele-specific integer-copy number for tumor suppressors.

To perform pan-cancer analysis of alternative 3'ss usage, we downloaded RNA sequencing of patient samples with or without *SF3B1* mutations from public dataset portals. 1) We downloaded raw fastq files of six cases of K700E, one G742D, one H662D, one K666N and nine randomly chosen WT samples, from ICGC Data Portal (<https://dcc.icgc.org>) with the project name 'CHRONIC LYMPHOCYTIC LEUKEMIA - ES'. 2) Six K700E and one K666N case of CLL patients and 7 WTs were accessed from dbGaP under the project of The UC San Diego CLL Study (phs000767.v1.p1). 3) 16 cases of TCGA UVM with *SF3B1* hotspot mutations (13 R625H/C, one H662R and two K666T) and equal number of UVM WTs were downloaded from cancer genomics cloud ([cgc.sbgenomics.com](http://cgc.sbgenomics.com)). 4) 20 cases of *SF3B1* mutant TCGA-BRCA samples and 20 BRCA WTs were downloaded from cancer genomics cloud. The 20 cases of *SF3B1* mutations include ten K700E, one K666N, one N626D, one K666E, one E622Q, one D781E as well as five non-hotspot changes (A633V, E802Q, G241\*, L646F and S455Vfs\*13) (**Table S3**).

### Isolation of mouse mammary gland epithelial cells

Mammary gland epithelial cells were isolated from 4-month old mice as described previously (5). Briefly, mammary glands dissected from 4-month-old female mice were minced and digested with collagenase/trypsin mixture for 4 hours at 37°C. After washed with L15 medium supplemented with 10% FBS, the red blood cells were lysed with Hybri-Max Red Blood Cell Lysing Buffer (Sigma). After washing with L15/10% FBS medium, the samples were incubated in serum-free Joklik's medium and then treated with 0.05% trypsin and DNase I (Fisher Scientific). Trypsin was neutralized by adding L15/10% FBS medium and the samples were filtered through a 40µm cell strainer to remove cell clumps and debris. Cells were pelleted and resuspended in L15/10% FBS medium. Fluorescence-

activated cell sorting (FACS) was used to isolate mammary epithelial cells. The cells were incubated with antibodies against CD24 and CD45 in L15/10% FBS medium on ice for 45 min. DAPI was used to exclude dead cells. The mammary epithelial cells were sorted as CD24+/CD45- cells on FACS Aria sorter. Data analysis was performed using the FCS Express 7 software.

### **Flow cytometry analyses of hematopoietic cells**

Peripheral blood and bone marrow samples collected from mice were first lysed with ACK lysis buffer to remove red blood cells and washed with ice-cold PBS. Cells were stained with monoclonal antibodies against cell surface markers (listed in **Table S7**) in PBS/2% BSA for 30 minutes on ice. DAPI was used to exclude dead cells. FACS analysis was performed on an LSRII Fortessa (BD Biosciences) and data analysis was performed using the FlowJo 10 software.

### **Identification of novel cryptic 3' ss usage**

In order to annotate novel 3'ss that are not present in current genome, we adopted the splice junction read output by STAR alignment (6). This preprocess is motivated by DeBoever et al. (7), to enable a more sensitive detection of novel sites, specifically designed for 3/5'ss identification. To this end, all the bam files were transformed back to .fastq format (samtools), and re-aligned using a splice junction database (7). Counts of junction reads from SJ.out.tab file (STAR output) were merged into a unique matrix, with each row indicates one splice junction and column for each sample. We further filtered out junctions with a read coverage of < 50 sum up across all samples. Then, for each junction, we defined it as a novel splice site if it only has one end (either 3' or 5') annotated in the provided known genome, while the other end is not. Corresponding canonical junction are found if it shares the same annotated end of the cryptic junction and the other end also annotated. If this kind of canonical junction did exist, we calculated the distance between the novel end and the corresponding canonical end. If there was more than one distance, we utilized the minimum. After adding the information of transcript's strand and relative position to the canonical site, we determined whether a cryptic site is on 3'ss or 5'ss, upstream or downstream. For mouse RNA-Seq, we reproduced this essential preprocess using a standard mm10 genome (Ensembl: Mus\_musculus. GRCm38.75). Junctions with both ends unannotated were discarded from the analysis.

Next, with clear relationship of each cryptic junction to its canonical one, we calculated the Percent-Spliced-In (PSI) using raw reads counts for each sample, transforming the reads matrix into a PSI matrix with values between 0-1. Then, for two groups of interested comparison, t-test was applied for all novel cryptic splice site using PSI, rather than reads counts, followed by BH multiple test correction. Difference of averaged PSI between two groups was also adopted as important metrics for potential differential

splicing. A bar plot of  $\log_2$  distance in base pair was used to demonstrate cryptic 3' splicing pattern on significantly differentially used novel sites.

### **Branchpoint mapping**

Lariat RT-PCR amplifying branchpoint-spanning fragments from lariat RNAs arising from splicing of alternative spliced introns of *PPP2R5A* or *MAP3K7*. Briefly, the RT reaction was done with SuperScript III reverse transcriptase (Invitrogen) and a primer complementary to the intronic sequences downstream of the 5' splice sites to generate cDNA from lariat RNAs. Nested PCR with pairs of outer primers (using RT primer as the reverse primer) and inner primers (**Table S8**) to amplify the branchpoint spanning fragments. The PCR products were cloned into the pGEM-T vector (Promega) and sequenced by Sanger sequencing.

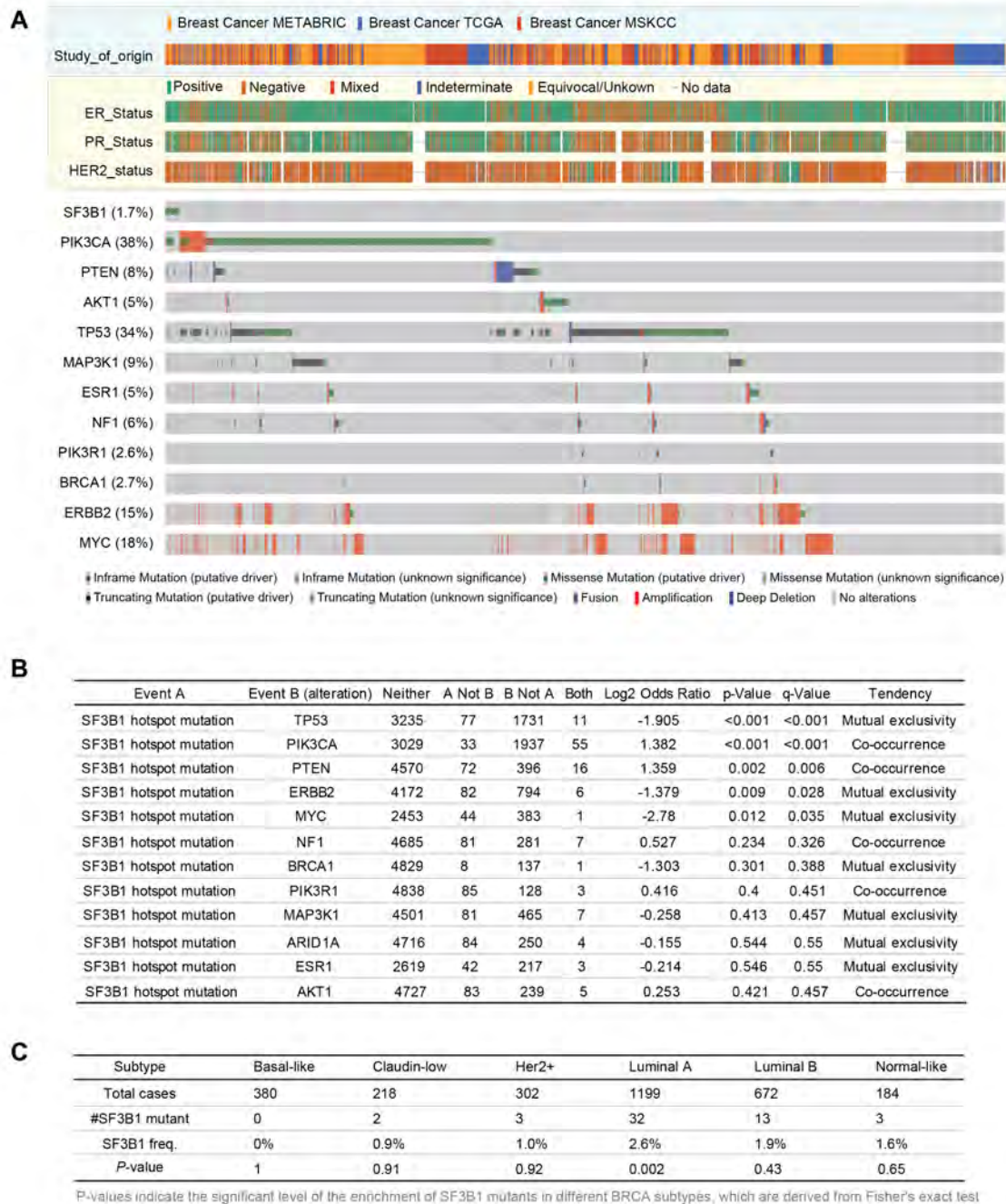
### **Soft agar assay**

Five thousand cells in single cell suspension in growth medium containing 0.35% low-melting-point agarose were overlaid onto a solidified normal medium containing 0.7% low-melting-point agarose. One ml growth media were added on top of the soft agar layers and refreshed twice a week. Cells were incubated at 37°C with 5% CO<sub>2</sub> for 3 weeks. Colonies were visualized by staining with crystal violet and counted with Oxford GelCount colony counter.

## SUPPLEMENTAL REFERENCES

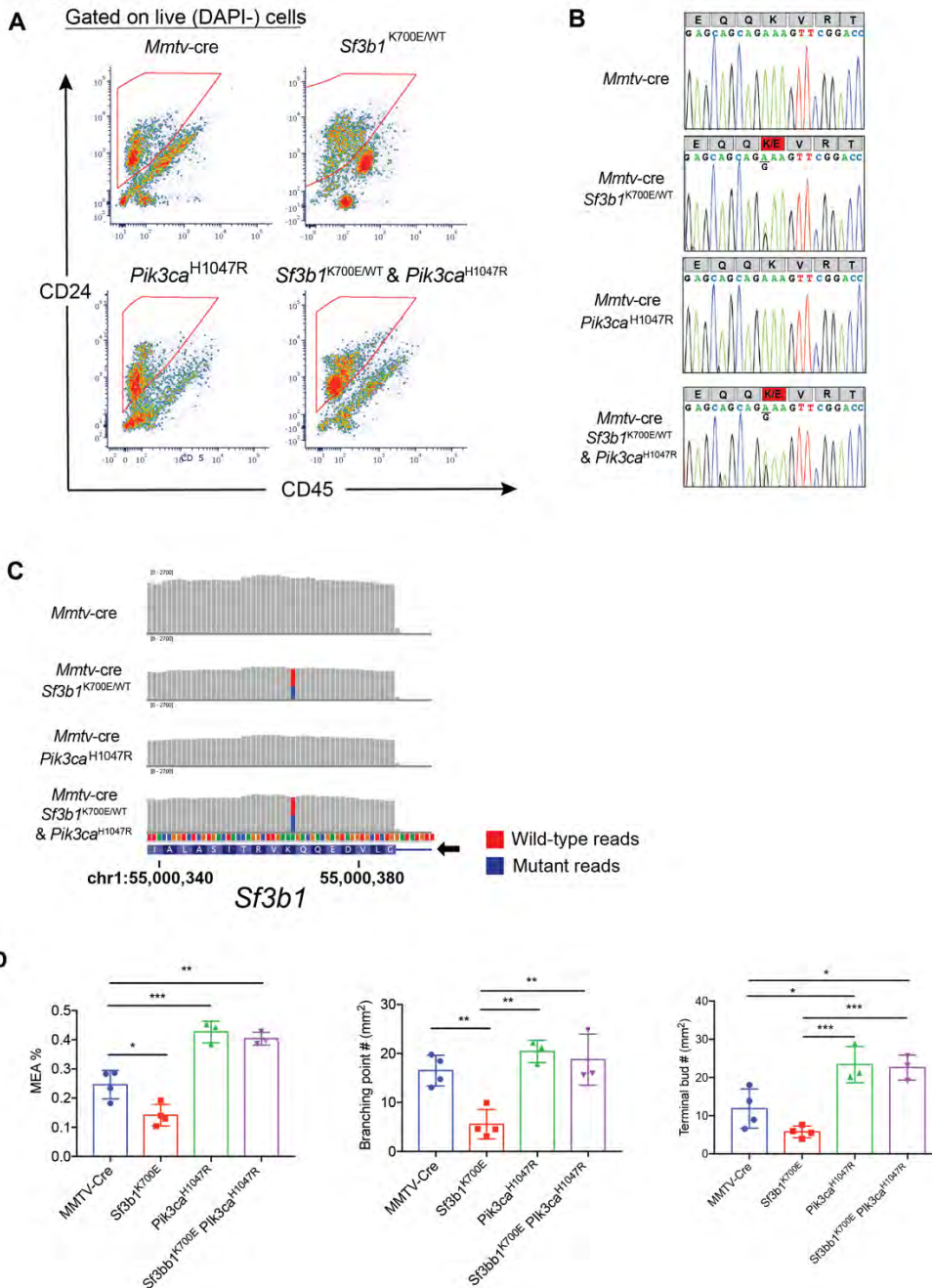
1. Cheng DT, Mitchell TN, Zehir A, Shah RH, Benayed R, Syed A, et al. Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT): A Hybridization Capture-Based Next-Generation Sequencing Clinical Assay for Solid Tumor Molecular Oncology. *J Mol Diagn.* 2015;17(3):251-64.
2. Razavi P, Chang MT, Xu G, Bandlamudi C, Ross DS, Vasan N, et al. The Genomic Landscape of Endocrine-Resistant Advanced Breast Cancers. *Cancer Cell.* 2018;34(3):427-38 e6.
3. Zehir A, Benayed R, Shah RH, Syed A, Middha S, Kim HR, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med.* 2017;23(6):703-13.
4. Li Z, Razavi P, Li Q, Toy W, Liu B, Ping C, et al. Loss of the FAT1 Tumor Suppressor Promotes Resistance to CDK4/6 Inhibitors via the Hippo Pathway. *Cancer Cell.* 2018;34(6):893-905 e8.
5. Smalley MJ. Isolation, culture and analysis of mouse mammary epithelial cells. *Methods Mol Biol.* 2010;633:139-70.
6. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29(1):15-21.
7. DeBoever C, Ghia EM, Shepard PJ, Rassenti L, Barrett CL, Jepsen K, et al. Transcriptome sequencing reveals potential mechanism of cryptic 3' splice site selection in SF3B1-mutated cancers. *PLoS computational biology.* 2015;11(3):e1004105.

Figure S1



**Figure S1. SF3B1 hotspot mutations in breast cancer. Related to Figure 1. (A)** Oncoprint of SF3B1 hotspot mutations and other breast cancer driver alterations across 5,366 patients including METABRIC, MSK-IMPACT, and TCGA breast cancer cohorts. **(B)** Contingency table of co-occurrence/mutual exclusivity of SF3B1 hotspot mutations and other breast cancer driver alterations. **(C)** Distribution of SF3B1 mutations across breast cancer subtypes (METABRIC and TCGA cohorts). 'Total cases' indicates the total number of samples assigned to each subtype. 'SF3B1 mutant' and 'SF3B1 freq.' indicate the number of SF3B1 mutated cases and frequency in each subtype, respectively. 'P-value' indicates the enrichment of SF3B1 mutants, calculated from the 2 by 2 contingency table test (SF3B1 mutation vs each subtype). One-side test is performed here.

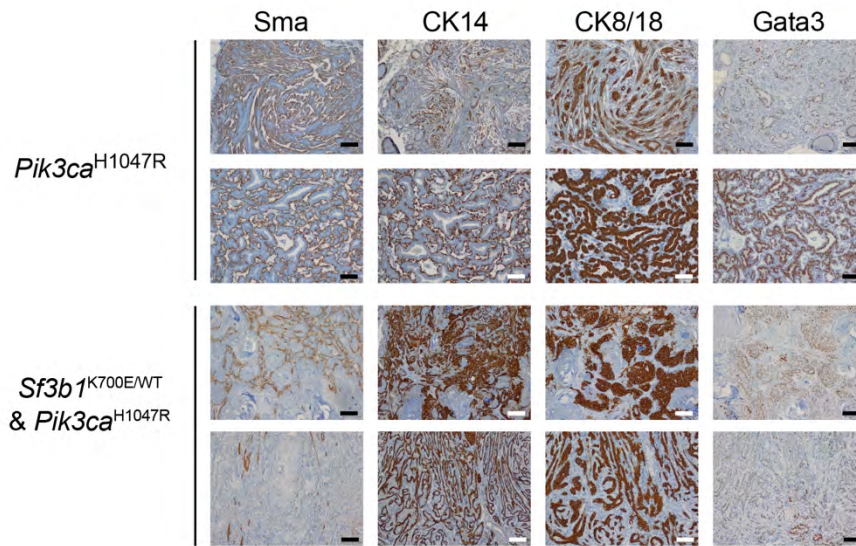
## Figure S2



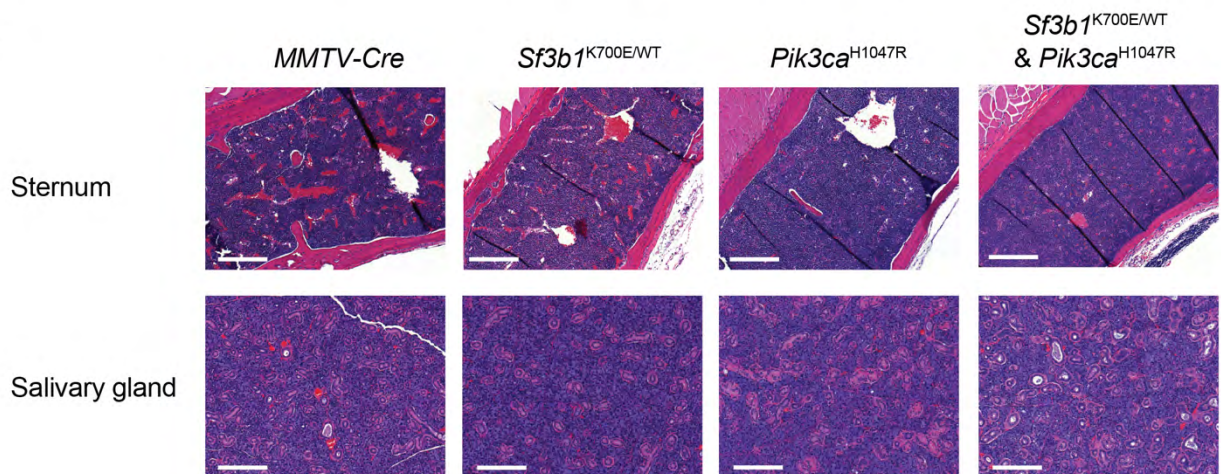
**Figure S2. Validation of *MMTV-cre Sf3b1*<sup>K700E/WT</sup> knockin mice and purification of mammary breast epithelial cells for RNA-seq. Related to Figure 2. (A) FACS plots used for purification of mammary breast epithelial cells from *MMTV-cre* control, *MMTV-cre Sf3b1*<sup>K700E/WT</sup>, *MMTV-cre R26-LSL-Pik3ca*<sup>H1047R</sup>, and *MMTV-cre Sf3b1*<sup>K700E/WT</sup> *Pik3ca*<sup>H1047R</sup> 4-month old mice. (B) Sanger sequencing of *Sf3b1* cDNA around sequences coding for K700 from live, CD45<sup>-</sup>, CD24<sup>+</sup> cells from mice in (A). (C) RNA-seq coverage plots of region of *Sf3b1*<sup>K700E</sup> mutant allele in material from mice in (A). (D) Quantitation of mammary epithelial area (MEA%, left) and the numbers of branching points (center) and terminal buds (right) on whole mount staining of mammary glands from *MMTV-cre* control (n=4), *Sf3b1*<sup>K700E/WT</sup> (n=4), *Pik3ca*<sup>H1047R</sup> (n=3) and double mutant (n=3) mice at age of 4 months. \* p<0.05, \*\* p<0.01, \*\*\* p<0.001 (by Tukey's multiple comparisons test).**

## Figure S3

A

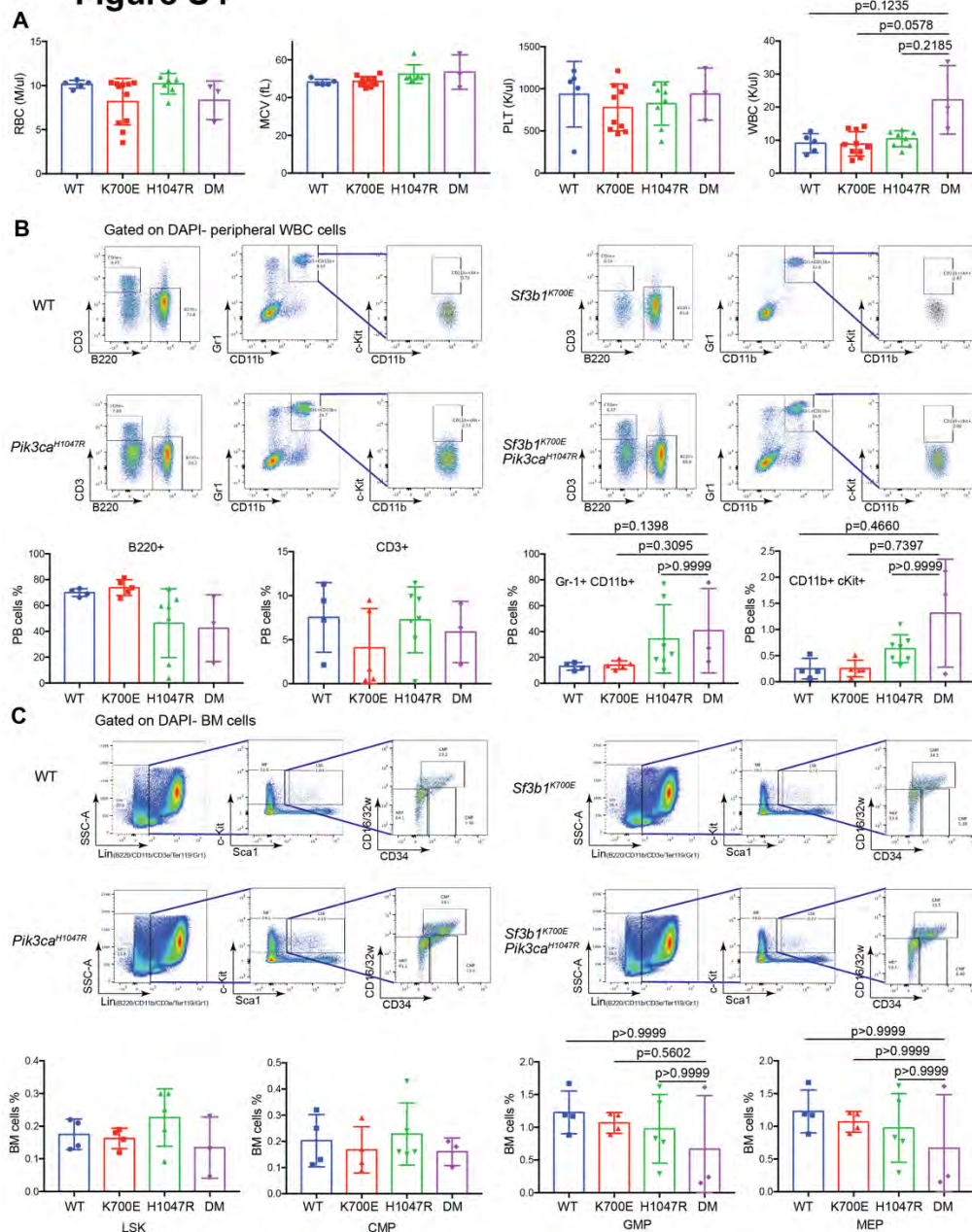


B



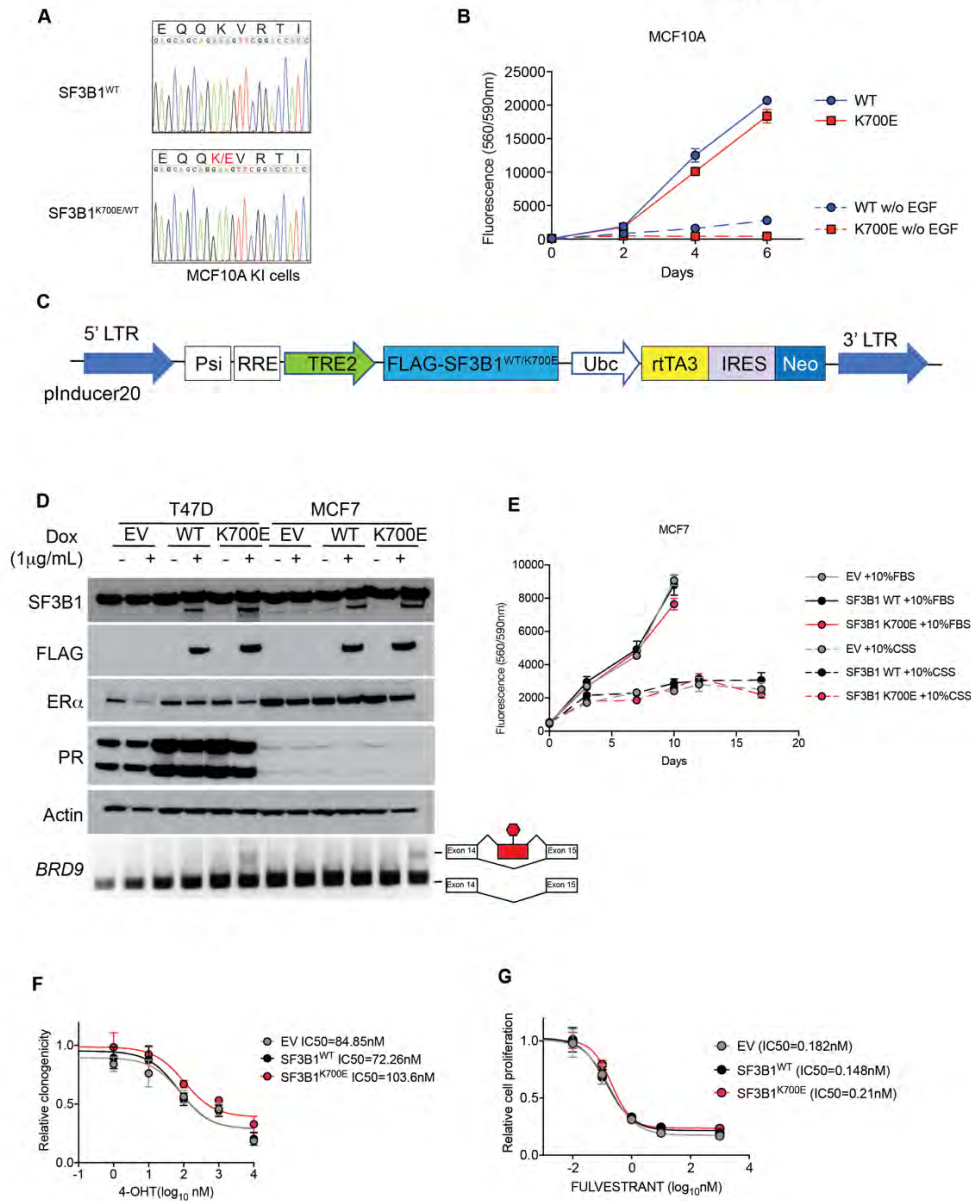
**Figure S3. Characterization of *MMTV-cre* transgenic mice. Related to Figure 2. (A)** IHC of smooth muscle actin (Sma), cytokeratin 14 (CK14), cytokeratin 8/18 (CK8/18), and Gata3 in mammary gland tumors (same as in Figure 2D) developing in *MMTV-cre R26-LSL-Pik3ca*<sup>H1047R</sup>, and *MMTV-cre Sf3b1*<sup>K700E/WT</sup> *R26-LSL-Pik3ca*<sup>H1047R</sup> mice. Bars: 50 $\mu$ M. **(B)** Hematoxylin and eosin (H&E) staining of sternums and salivary glands from *MMTV-cre R26-LSL-Pik3ca*<sup>H1047R</sup> and *MMTV-cre Sf3b1*<sup>K700E/WT</sup> *R26-LSL-Pik3ca*<sup>H1047R</sup> mice at the time of sacrifice due to mammary gland tumors, as well as *MMTV-cre* and *MMTV-cre Sf3b1*<sup>K700E/WT</sup> mice at same age. Bar: 250 $\mu$ M.

**Figure S4**



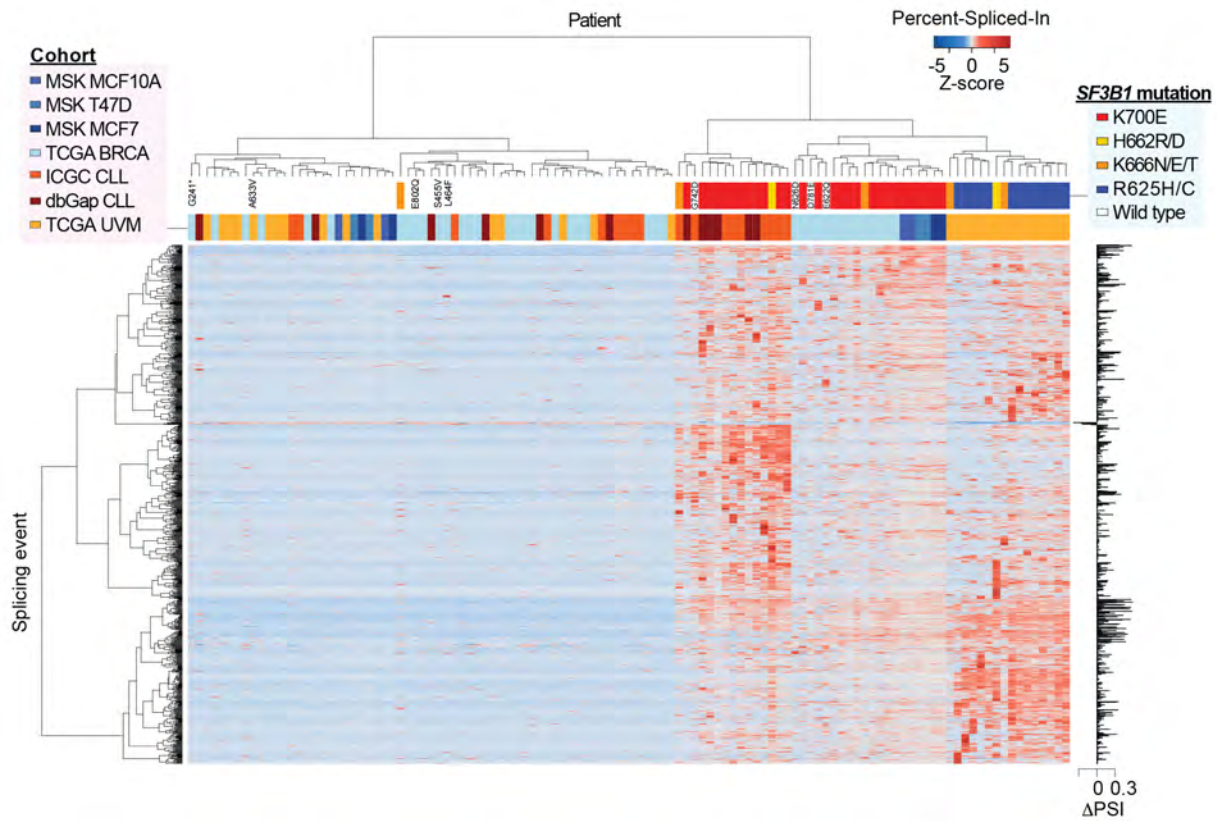
**Figure S4. Hematologic analysis of *MMTV-cre* control, *Sf3b1*<sup>K700E/WT</sup>, *Pik3ca*<sup>H1047R</sup>, and double mutant (DM) mice. Related to Figure 2. (A)** Red blood cell (RBC) counts, mean corpuscular volume (MCV), platelet counts (PLT), and white blood cell counts of each genotype of *MMTV-cre* mice at the age of 6-10 months. **(B)** Immunophenotyping of peripheral blood mononuclear cells in samples from (A) by FACS. Representative gating on DAPI<sup>-</sup> cells is shown at top and the percentage of B220<sup>+</sup>, CD3<sup>+</sup>, Gr-1<sup>+</sup> CD11b<sup>+</sup> and CD11b<sup>+</sup> c-Kit<sup>+</sup> peripheral blood (PB) cells is shown at bottom. **(C)** Analysis of hematopoietic stem and progenitor cells in bone marrow samples from on *MMTV-cre* control (“WT”), *Sf3b1*<sup>K700E/WT</sup> (“K700E”), *Pik3ca*<sup>H1047R</sup> (“H1047R”), and double mutant (DM) mice by FACS. Representative gating on DAPI<sup>-</sup> cells is shown at top and the percentage of LSK population and myeloid progenitors is shown at bottom. Abbreviations: “WT” = *MMTV-cre* control; “K700E” = *MMTV-cre Sf3b1*<sup>K700E/WT</sup>; “H1047R” = *MMTV-cre R26-LSL-Pik3ca*<sup>H1047R</sup>; “DM” = *MMTV-cre Sf3b1*<sup>K700E/WT</sup> *R26-LSL-Pik3ca*<sup>H1047R</sup>. P values were derived by Dunn’s multiple comparisons test.

**Figure S5**



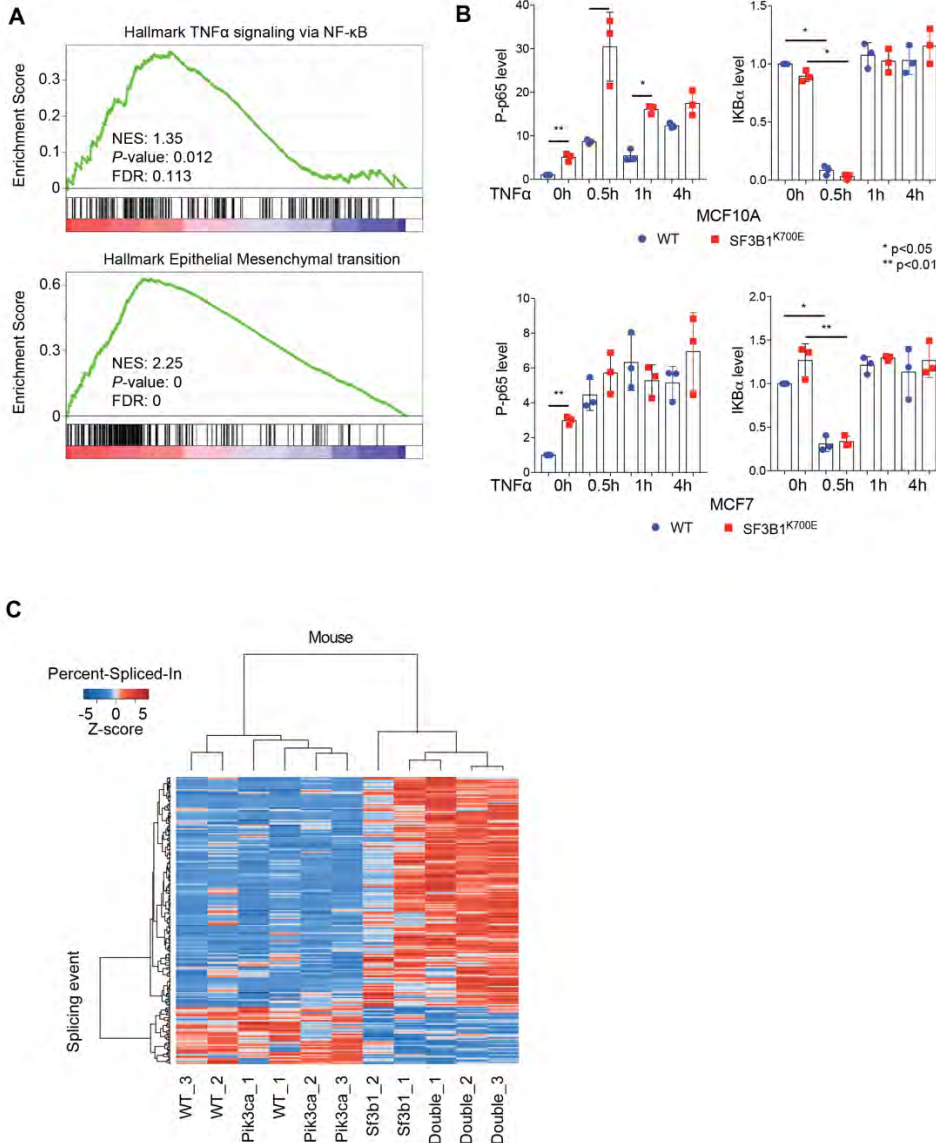
**Figure S5. Generation and validation of *SF3B1* wild-type and mutant human breast cancer cell lines. Related to Figure 3. (A)** Sanger sequencing of *SF3B1* cDNA around sequences coding for K700 from MCF10A cells with or without knockin (KI) of *SF3B1* K700E mutation. **(B)** Proliferation curve of wild-type MCF10A cells and isogenic MCF10A cells with *SF3B1*<sup>K700E/WT</sup> knockin in the presence or absence of EGF. **(C)** Schematic showing the construct of doxycycline-inducible vectors to express FLAG-tagged *SF3B1* WT or mutant (K700E) protein in breast cancer cell lines. **(D)** Western blot of *SF3B1*, FLAG-*SF3B1*, ER $\alpha$ , PR and  $\beta$ -actin from MCF7 and T47D cells expressing FLAG-*SF3B1*<sup>WT</sup> or FLAG-*SF3B1*<sup>K700E</sup> proteins induced by doxycycline (Dox) for 48 hours. RT-PCR of *BRD9* splicing events is shown on bottom with a schematic showing the inclusion of a poison exon between exon 14 and 15. **(E)** Proliferation curve of MCF7 cells expressing empty vector (EV), wild-type *SF3B1* (*SF3B1*<sup>WT</sup>) and mutant *SF3B1* (*SF3B1*<sup>K700E</sup>) in culture media supplemented with 10% FBS or charcoal stripped serum (CSS). **(F)** and **(G)** IC<sub>50</sub> of MCF7 cells from (E) following 5 days of exposure to 4-Hydroxytamoxifen **(F)** and fulvestrant **(G)**.

Figure S6



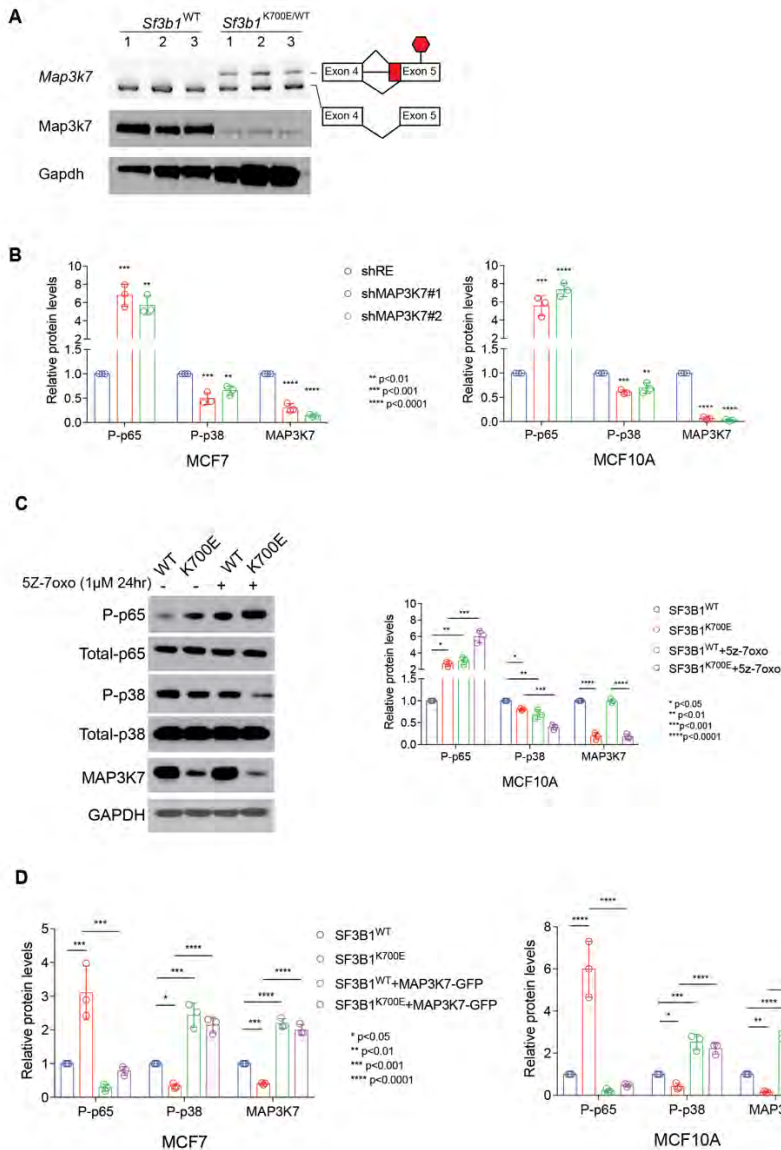
**Figure S6. Differential aberrant 3' splicing events across *SF3B1* mutant cancers. Related to Figure 3.** Hierarchical clustering and heatmap analysis of differential 3' splice sites (3'ss) between *SF3B1* mutant and wild-type (WT) samples. Rows and columns represent cryptic 3'ss events and samples, respectively. 618 cryptic 3'ss identified from the pan-cancer dataset are shown. Z-score in the matrix represents normalized Percent-Spliced-In (PSI). BRCA: breast cancer; CLL; chronic lymphocytic leukemia; UVM: uveal melanoma.

## Figure S7



**Figure S7. Activation of NF- $\kappa$ B signaling in *SF3B1* mutant breast tumors. Related to Figure 4. (A)** Gene set enrichment analysis identifies enrichment of hallmark gene sets of TNF $\alpha$  signaling via NF- $\kappa$ B (top) and epithelial mesenchymal transition (bottom) in *SF3B1* mutant versus wild-type human breast tumors from TCGA. FDR: false discovery rate; NES: normalized enrichment score. **(B)** Quantitation of Western blots in Figure 4B. Densitometry of phosphorylated p65 was normalized to the level of total p65, and I $\kappa$ B $\alpha$  was normalized to GAPDH. Results were presented as levels relative to the untreated wild-type cells (WT 0h). The mean from three independent experiments is shown with standard deviation. \* p<0.05, \*\* p<0.01, and p values were derived from Student's t test. **(C)** Differential aberrant 3'splicing events across mouse mammary epithelial cell samples. Hierarchical clustering and heatmap analysis of differential 3'ss among mammary epithelial cell samples from *MMTV-cre* control (WT), *MMTV-cre Sf3b1<sup>K700E/WT</sup>* (*SF3B1*), *MMTV-cre R26-LSL-Pik3ca<sup>H1047R</sup>* (*PIK3CA*) and *MMTV-cre Sf3b1<sup>K700E/WT</sup> R26-LSL-Pik3ca<sup>H1047R</sup>* (Double) mice. Rows and columns represent cryptic 3'ss events and samples, respectively. 219 cryptic 3'ss identified are shown. Z-score in the matrix represents normalized Percent-Spliced-In (PSI).

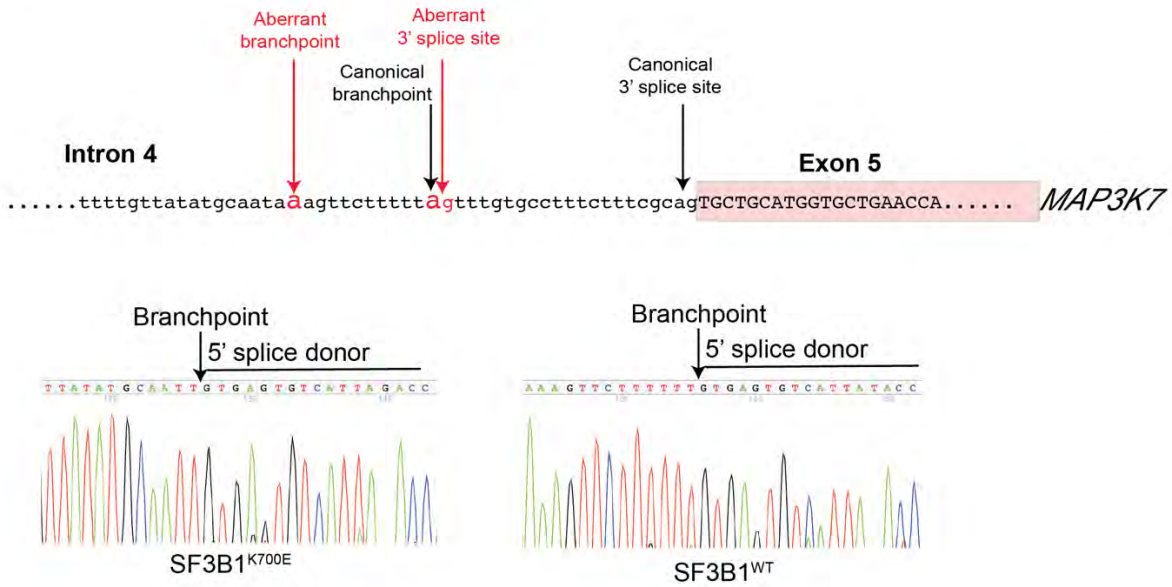
## Figure S8



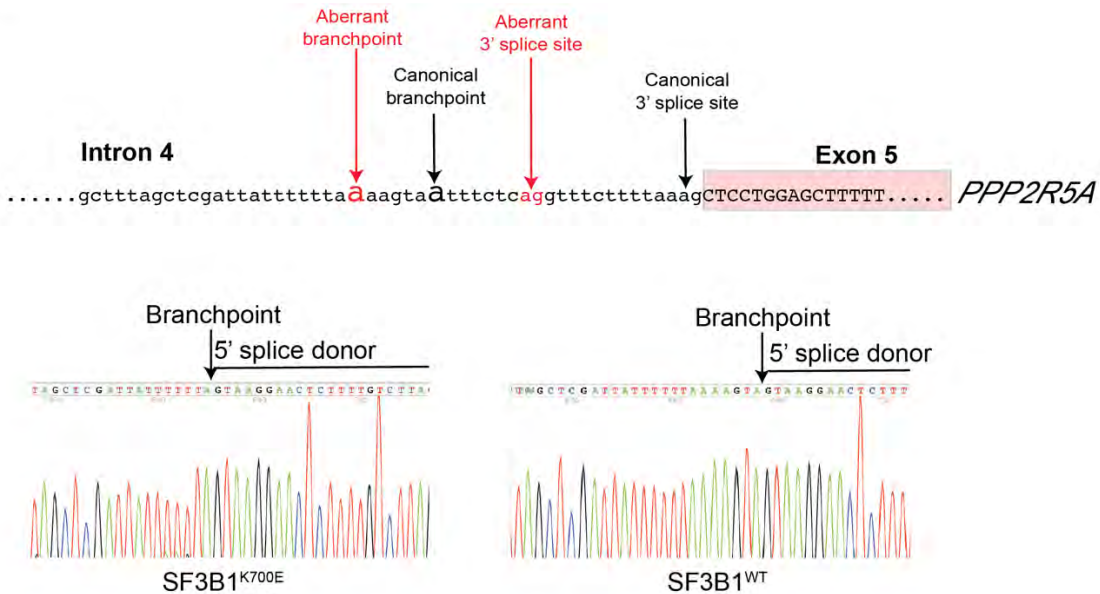
**Figure S8. Aberrant splicing and downregulation of MAP3K7 by mutant SF3B1 promotes NF- $\kappa$ B activation in breast cancer. Related to Figure 5. (A)** RT-PCR (top) and Western blot analyses of Map3K7 on mammary gland tumors collected from *MMTV-cre R26-LSL-Pik3ca*<sup>H1047R</sup>, and *MMTV-cre Sf3b1*<sup>K700E/WT</sup> *R26-LSL-Pik3ca*<sup>H1047R</sup> mice. **(B)** Quantitation of MAP3K7 and phospho-p65 and phospho-p38 levels in MCF10A and MCF7 cells transduced with anti-MAP3K7 shRNA (Western blots in Figure 5G). Phosphorylated p65 or p38 levels were normalized to total p65 or p38 levels, respectively, and MAP3K7 level was normalized to GAPDH. Results are represented as levels relative to cells transduced with shRNA targeting Renilla luciferase (shRE). P values were derived by Dunnett's test. **(C)** MAP3K7 and phospho- and total p65 and p38 levels in isogenic MCF10A cells with or without SF3B1<sup>K700E</sup> knockin treated with 5Z-7-Oxozeaenol (5Z-7oxo) versus DMSO (-). Quantitation shown at right. P values were derived by Tukey's multiple comparison's test. **(D)** Quantitation of MAP3K7 and phospho-p65 and phospho-p38 levels in MCF10A and MCF7 cells transduced with or without MAP3K7-GFP cDNA (Western blots in Figure 5H). Results are represented as levels relative to wild-type control. P values were derived by Tukey's multiple comparison's test.

# Figure S9

A

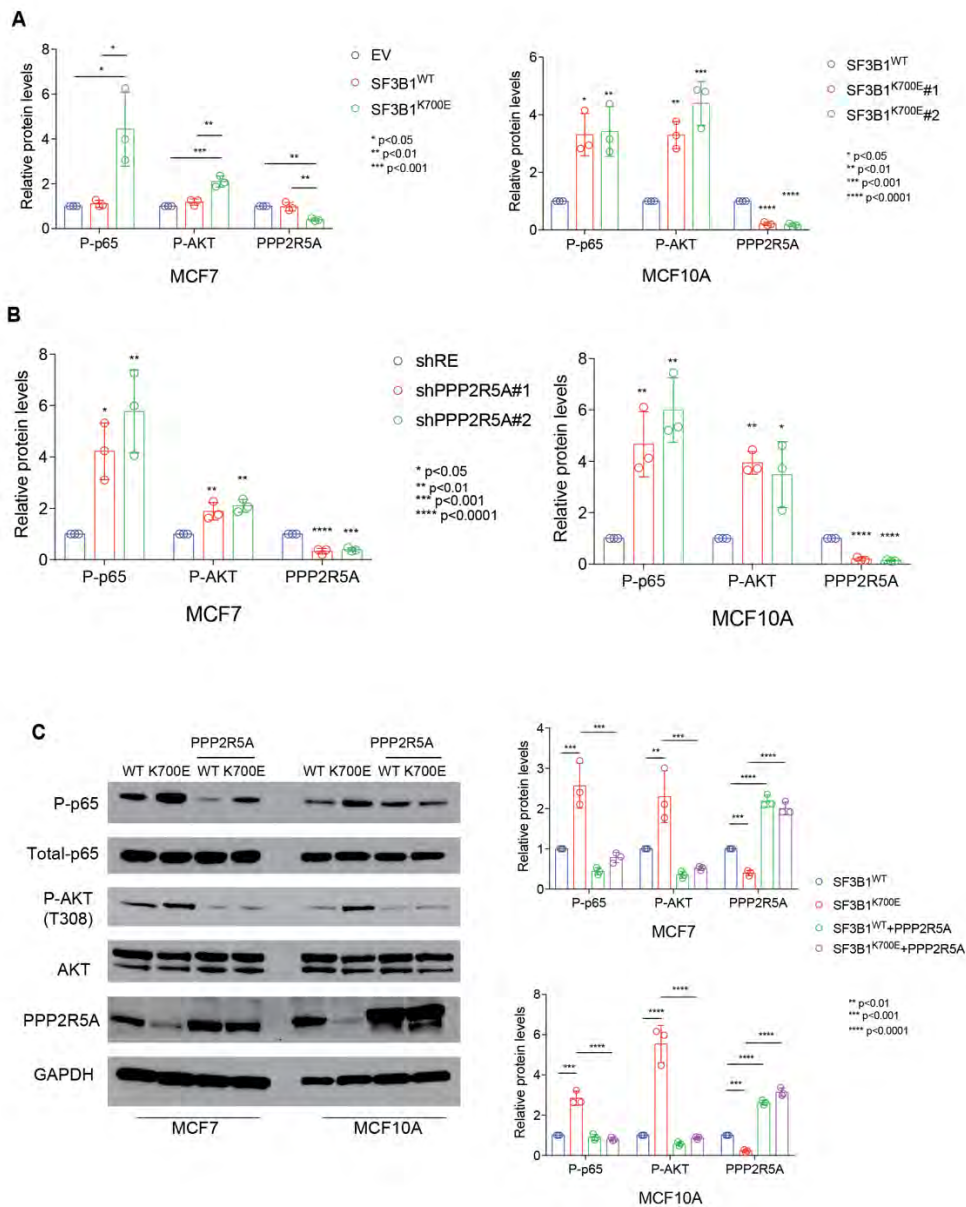


B



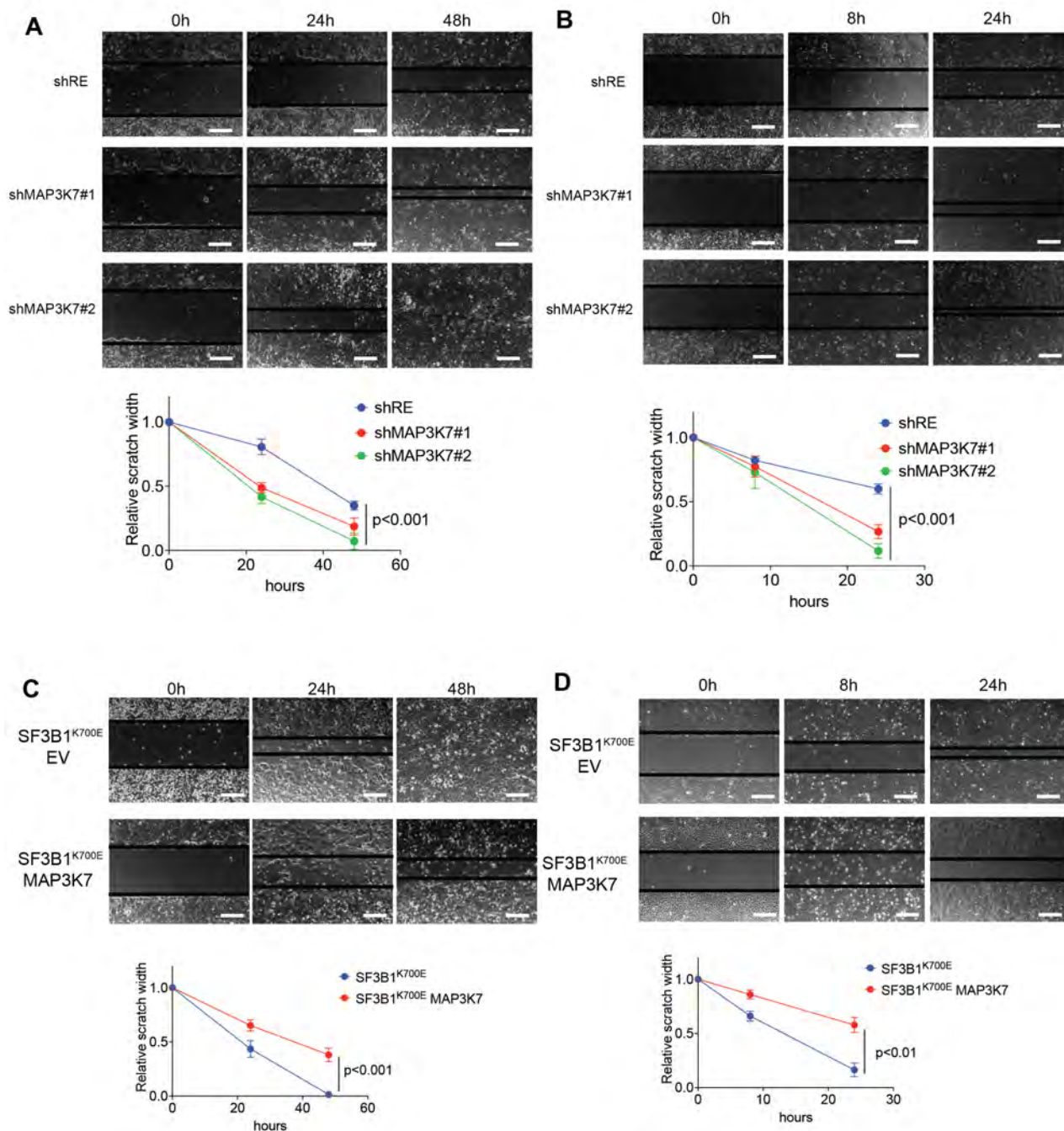
**Figure S9. Mapping of canonical and aberrant branchpoints of *MAP3K7* (A) and *PPP2R5A* (B) differential splicing events. Related to Figures 5-6.** Branchpoints were mapped by lariat RT-PCR as described in Methods. Canonical branchpoint was mapped in parental MCF10A cells and aberrant branchpoint was mapped in MCF10A cells with heterozygous knockin of *SF3B1* K700E mutation. Sanger sequencing of lariat RT-PCR product depicting the joining of branchpoint and 5' splice donor site is shown below a schematic of sequences around the intron-exon boundary.

**Figure S10**



**Figure S10. Aberrant splicing and downregulation of PPP2R5A by mutant SF3B1 promotes AKT activation and inflammatory signaling in breast cancer. Related to Figure 6. (A)** Quantitation of PPP2R5A and phospho- p65 and AKT (at T308) levels in MCF7 or MCF10A cells expressing mutant SF3B1 (Western blots in Figure 6D). **(B)** Quantitation of PPP2R5A and phospho- p65 and AKT (at T308) levels in MCF10A and MCF7 cells transduced with anti-PPP2R5A shRNA (Western blots in Figure 6F). **(C)** Quantitation of PPP2R5A and phospho- and total p65 and AKT (at T308) levels in MCF10A and MCF7 cells transduced with or without PPP2R5A cDNA (Western blots in Figure 6F). Phosphorylated protein level was normalized to total level of the protein, and PPP2R5A level was normalized to GAPDH. Results are represented as levels relative to cells transduced with shRNA targeting Renilla luciferase (shRE) in (B) or wild-type control in (C). P values were derived by Dunnett's test in (A) and (B) or Tukey's test in (C).

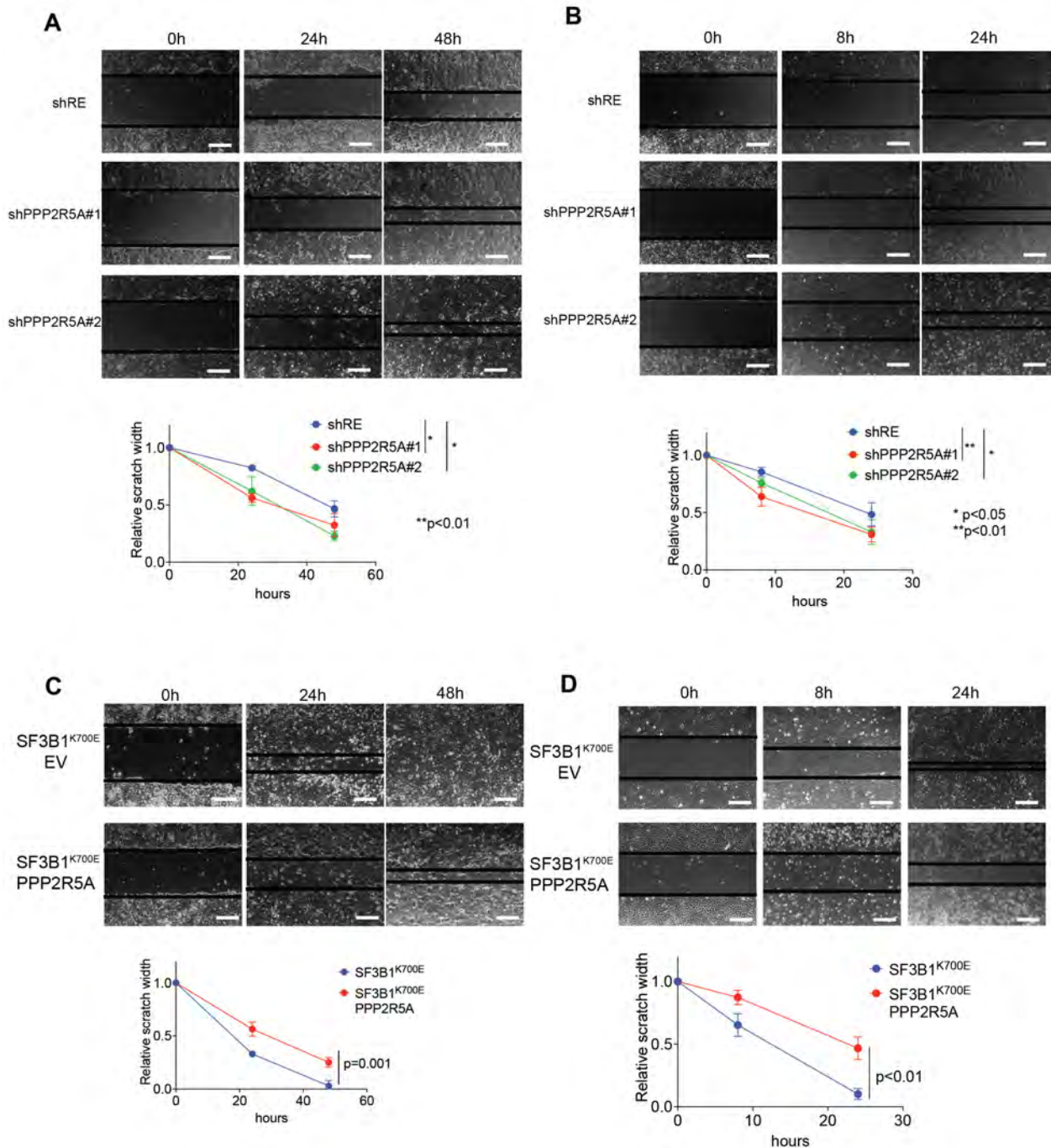
## Figure S11



**Figure S11. MAP3K7 inhibits cell migration in breast cancer cells. Related to Figure 7.**

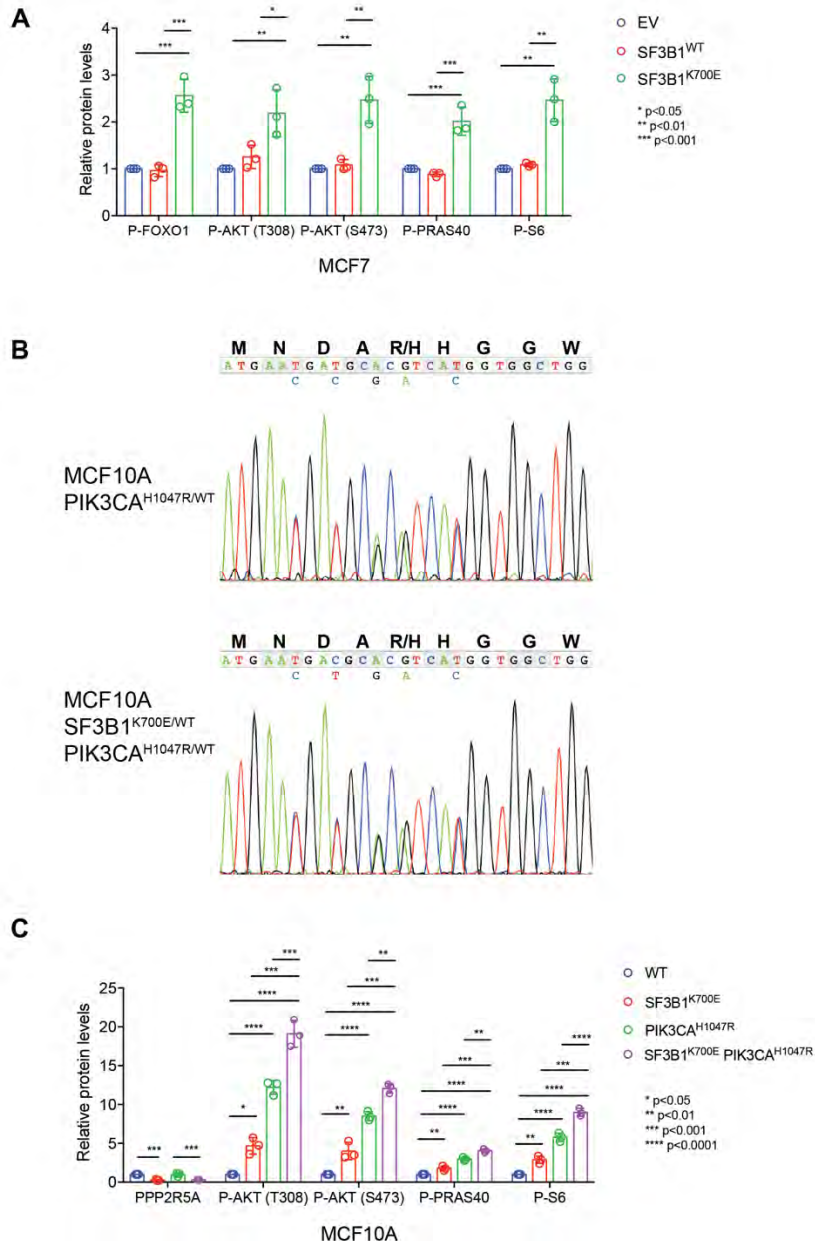
Representative photos (top) at various time following wound creation in cell culture of MCF7 (**A**) and MCF10A cells (**B**) transduced with anti-*MAP3K7* shRNA. Scale bars represent 200 $\mu$ M. Representative photos (top) at various time following wound creation in cell culture of MCF7 (**C**) and MCF10A cells (**D**) expressing SF3B1 K700E mutant transduced with or without *MAP3K7-GFP* cDNA (Western blots in Figure 5H). Relative scratch width over time shown below the images (mean value  $\pm$  standard deviation shown below; p values by ANOVA).

## Figure S12



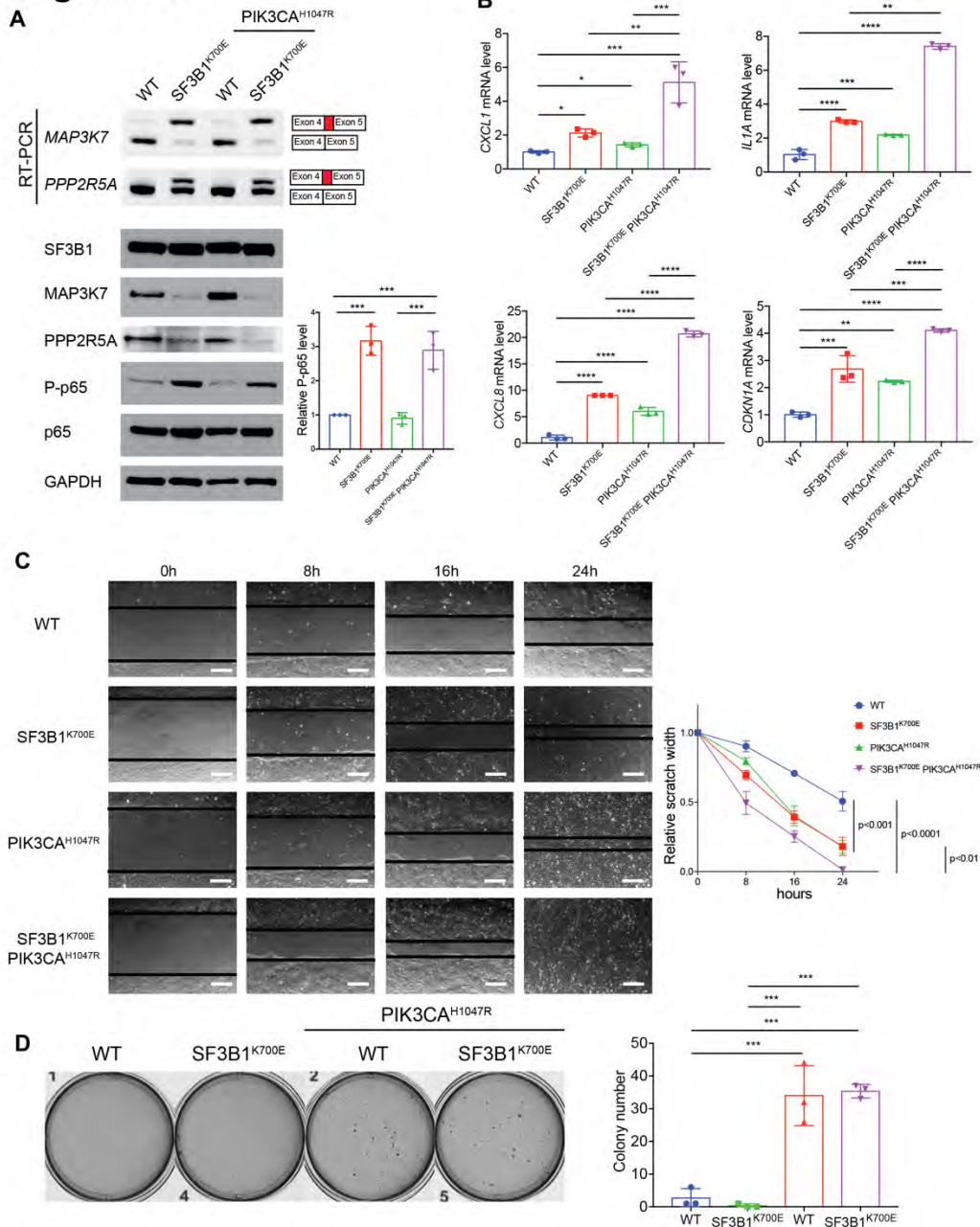
**Figure S12. Knockdown of PPP2R5A promotes migration of breast cancer cells. Related to Figure 7.** Representative photos (top) at various time following wound creation in cell culture of MCF7 (A) and MCF10A cells (B) transduced with anti-PPP2R5A shRNA. Scale bars represent 200 $\mu$ M. Representative photos (top) at various time following wound creation in cell culture of MCF7 (C) and MCF10A cells (D) expressing SF3B1 K700E mutant transduced with or without PPP2R5A cDNA (Western blots in Figure S10C). Relative scratch width over time shown below the images (mean value  $\pm$  standard deviation shown below; p values by ANOVA).

## Figure S13



**Figure S13. *SF3B1*<sup>K700E</sup> mutation sensitizes breast cancer cells to AKT inhibitors. Related to Figure 8. (A)** Quantitation of \phospho- AKT and its downstream target in MCF7 expressing empty vector (EV), *SF3B1* WT, or *SF3B1* K700E cDNA (Western blots in Figure 8A). **(B)** Validation of *PIK3CA*<sup>H1047R</sup> knockin in MCF10A cells. Sanger sequencing shows the heterozygous knockin of *PIK3CA*<sup>H1047R</sup> mutation in parental or *SF3B1*<sup>K700E</sup> mutant MCF10A cells. Four additional silent mutations were knocked in on the same allele to prevent further editing by CRSIPR-Cas9. **(C)** Quantitation of Western blots of MCF10A WT cells as well as *SF3B1*<sup>K700E</sup>/WT knockin cells, *PIK3CA*<sup>H1047R</sup>/WT knockin cells and *SF3B1*<sup>K700E</sup> *PIK3CA*<sup>H1047R</sup> double knockin cells cultured in the absence of EGF (Western blots in Figure 8C). Phosphorylated protein level was normalized to total level of the protein, and PPP2R5A level was normalized to GAPDH. Results are represented as levels relative to wild-type control. P values were derived by Tukey's multiple comparisons test.

## Figure S14



### Figure S14. Activation of NF- $\kappa$ B signaling in *SF3B1*/*PIK3CA* double-mutant MCF10A cells.

**Related to Figure 8. (A)** RT-PCR and Western blot of MCF10A WT cells as well as *SF3B1*<sup>K700E/WT</sup> knockin cells, *PIK3CA*<sup>H1047R/WT</sup> knockin cells and *SF3B1*<sup>K700E/WT</sup> *PIK3CA*<sup>H1047R/WT</sup> double knockin cells cultured in the absence of EGF. Quantification of phospho-p65 (normalized to total p65) is shown. **(B)**

**(B)** Quantitative RT-PCR of pro-inflammatory genes in the cells above. **(C)** Representative photos (top) at 0, 8, 16 and 24 hours following wound creation in cell culture of MCF10A cells in (A). Scale bars represent 200 $\mu$ M. Relative scratch width over time shown at right (mean value + standard deviation shown below; p values by ANOVA). **(D)** Representative photos of colonies formed from MCF10A cells in (A) on soft gar on 6-well plates 16 days after seeding. Colony numbers of three experiments are shown at right. \* p<0.05, \*\* p<0.01, \*\*\* p<0.001, \*\*\*\* p<0.0001, and p values in (A), (B) and (D) were derived by Tukey's multiple comparisons test.

**(D)** Representative photos of colonies formed from MCF10A cells in (A) on soft gar on 6-well plates 16 days after seeding. Colony numbers of three experiments are shown at right. \* p<0.05, \*\* p<0.01, \*\*\* p<0.001, \*\*\*\* p<0.0001, and p values in (A), (B) and (D) were derived by Tukey's multiple comparisons test.