AWARD NUMBER: W81XWH-19-1-0180

TITLE: Transcription, R-loops, and RNA Splicing in Ewing Sarcoma

PRINCIPAL INVESTIGATOR: Liesl A. Lawrence

CONTRACTING ORGANIZATION: University of Texas Health Science Center at San Antonio

REPORT DATE: July 2020

TYPE OF REPORT: Annual

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

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
July 2020	Annual	7/1/2019 – 6/30/2020

4. TITLE AND SUBTITLE
Transcription, R-loops, and RNA Splicing in Ewing Sarcoma

5a. CONTRACT NUMBER
W81XWH-19-1-0180

5b. GRANT NUMBER
CA181177

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Liesl A. Lawrence (lawrencel@uthscsa.edu)

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

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

University of Texas Health Science Center
7703 Floyd Curl Dr
San Antonio TX 78229

8. PERFORMING ORGANIZATION REPORT NUMBER

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

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

10. SPONSOR/MONITOR'S ACRONYM(S)

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

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Ewing sarcoma (EwS) is an aggressive pediatric bone and soft tissue cancer driven primarily by the EWS-FLI1 fusion oncogene. EWS-FLI1 acts as a transcription factor and also interferes with normal regulation of transcription and transcription-associated RNA processing. We hypothesized that EWS-FLI1 driven hyper-activation of transcription causes a targetable dependence on RNA splicing in Ewing sarcoma. The specific aims of the project are to 1) determine the mechanistic relationship between EWS-FLI1-driven transcription, R-loops, and splicing vulnerabilities in EwS, and 2) establish splicing as a therapeutic target in EwS. We tested the effect of depletion of key splicing factors and found an increased sensitivity in EwS compared to control cells. This was partially rescued by depleting EWS-FLI1 or by overexpressing RNASEH1, which degrades R-loops. EwS cells were extremely sensitive to splicing inhibition, which showed synergy with multiple chemotherapeutic agents. Splicing inhibition also caused cell cycle arrest and induced apoptosis. This work provides novel insight into transcription regulation and its dysregulation by EWS-FLI1. In addition, our results point to RNA splicing as a potential new therapeutic target in Ewing sarcoma, which has the potential to benefit all Ewing sarcoma patients, especially those with chemo-resistant disease.

15. SUBJECT TERMS
Ewing sarcoma, splicing, transcription, R-loops

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
Unclassified	Unclassified	Unclassified	Unclassified	40	19b. TELEPHONE NUMBER *(include area code)*

TABLE OF CONTENTS

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

Ewing sarcoma (EwS) is an aggressive bone and soft tissue cancer that occurs in children, adolescents, and young adults[1]. It is most commonly driven by a chromosomal translocation resulting in the fusion protein EWS-FLI1. Chemotherapy has substantially improved overall survival, but the prognosis is still very poor for those with metastatic or recurrent disease, and toxicity is a major concern in pediatric patients. Consequently, effective, less toxic, targeted treatment strategies are much needed. Previous studies reported that EwS cells display changes in alternative splicing events, a phenomenon linked to the role of EWSR1 as an RNA-binding protein and proposed regulator of splicing[2-5]. Splicing occurs concurrently with transcription, and the regulation of the two processes is coordinated[6-9]. Our lab recently reported that EWS-FLI1 caused hyper-phosphorylation of RNA Polymerase II (RNAPII)[10]. Accordingly, EwS cells displayed high levels of transcription and R-loops and impaired transcription regulation in response to DNA damage. Notably, the set of genes that are alternatively spliced in EwS are highly expressed and have strong R-loop signal. In a genome-wide RNAi screen in the EwS cell line TC32, we found that RNA splicing was one of the top processes required for cell viability. We hypothesized that EWS-FLI1 driven hyper-activation of transcription causes a targetable dependence on RNA splicing in Ewing sarcoma. The specific aims of this project are to 1) determine the mechanistic relationship between EWS-FLI1-driven transcription, R-loops, and splicing vulnerabilities in EwS, and 2) establish splicing as a therapeutic target in EwS. During the first year of the project, great progress has been made towards accomplishing these aims. We tested the effect of depletion of key splicing factors and found an increased sensitivity in EwS compared to control cells. This was partially rescued by depleting EWS-FLI1 or by overexpressing RNASEH1, which degrades R-loops. EwS cells were extremely sensitive to the splicing inhibitor E7107, which caused cell cycle arrest, induced apoptosis, and showed strong synergy with vincristine. This work provides insight into the dysregulation of transcription and splicing by EWS-FLI1 and suggests RNA splicing as a novel therapeutic target in Ewing sarcoma.

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

Ewing sarcoma, splicing, transcription, R-loops

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.

1. Major task 1: Confirm the dependence of Ewing sarcoma cells on the splicing machinery.
    a. Subtask 1: Cell viability with siRNA knockdown and cDNA overexpression of splicing factor genes in Ewing sarcoma and control cell lines. (target completion 10/19; 100% completed)
    b. Subtask 2: RT-qPCR of splice isoform expression with siRNA knockdown and cDNA overexpression of 3 splicing factor genes. (target completion 01/20; 50% completed)
    c. Subtask 3: EU/flow cytometry transcription assay with siRNA knockdown and cDNA overexpression of 3 splicing factor genes. (target completion 01/20; 100% completed)
    d. Subtask 4: Western blot for RNAPII-pSer2 and RNAPII-pSer5 with siRNA knockdown and cDNA overexpression of 3 splicing factor genes. (target completion 01/20; 100% completed)
    e. Subtask 5: Dot blot for R-loops with siRNA knockdown and cDNA overexpression of 3 splicing factor genes. (target completion 01/20; 100% completed)
2. Major Task 2: Determine the causal relationship between transcription dysregulation, R- loops, and splicing in Ewing sarcoma.
    a. Subtask 1: RT-qPCR of splice isoform expression in the presence of DRB treatment, THZ1 treatment, and RNASEH1 overexpression. (target completion 07/20; 50% completed)
    b. Subtask 2: EU/flow cytometry in the presence of DRB treatment, THZ1 treatment, and RNASEH1 overexpression. (target completion 07/20; 50% completed)
    c. Subtask 3: Dot blot for R-loops in the presence of DRB treatment, THZ1 treatment, and RNASEH1 overexpression. (target completion 07/20; 50% completed)
    d. Subtask 4: Paired-end RNA-seq with and without RNASEH1 overexpression. (target completion 04/20; 0% completed)
    e. Subtask 5: GRO-seq in TC32 and MSC cells. (target completion 07/20; 50% completed)
    f. Subtask 6: DRIP-seq in TC32 and MSC cells. (target completion 07/20; 100% completed)
    g. Subtask 7: Bioinformatics analysis of RNA-seq, DRIP-seq, and GRO-seq data (target completion 10/20; 50% completed)
3. Major Task 3: Determine whether splicing inhibitors show EwS-specific toxicity, alone or in combination with standard therapeutic agents.
    a. Subtask 1: Cell viability dose response curves for E7107, E7107+vincristine, E7107+paclitaxel, and E7107+docetaxel in 9 cell lines. (target completion 01/20; 100% completed)
    b. Subtask 2: Apoptosis assay for E7107, E7107+vincristine, E7107+paclitaxel, and E7107+docetaxel in 9 cell lines. (target completion 01/20; 100% completed)
4. Major Task 4: Determine the biological effects of splicing inhibition in EwS cells.
    a. Subtask 1: RT-qPCR of splice isoform expression, with and without E7107, with and without RNASEH1 overexpression. (target completion 01/21; 0% completed)
    b. Subtask 2: EU/flow cytometry with and without E7107, with and without RNASEH1 overexpression. (target completion 01/21; 50% completed)
    c. Subtask 3: Dot blot for R-loops with and without E7107, with and without RNASEH1 overexpression. (target completion 01/21; 50% completed)
    d. Subtask 4: Flow cytometry cell cycle analysis with and without E7107, with and without RNASEH1 overexpression. (target completion 01/21; 50% completed)
    e. Subtask 5: DNA combing with and without E7107, with and without RNASEH1 overexpression, with and without HU treatment. (target completion 04/21; 0% completed)

f. Subtask 6: DR-GFP assay for homologous recombination, with and without E7107, with and without RNASEH1 overexpression. (target completion 04/21; 0% completed)

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.

The first major goal of the project was to confirm the dependence of Ewing sarcoma cells on the splicing machinery. Cell viability experiments were performed to assess the effect of splicing factor depletion on Ewing sarcoma (EwS) versus control cell lines. Cell viability assays were performed in 384-well plates; siRNA was delivered by reverse transfection using RNAiMax and cDNA by forward transfection using Lipofectamine 3000. Cell growth was measured over three days by the IncuCyte® ZOOM live cell analysis system followed by an endpoint Cell Titer Glo assay in the same plates to quantify viability by ATP production. 9 cell lines were evaluated including 6 EwS (TC32, ES1, ES7, A673, EW8, CHLA10) and 3 control (MSC (mesenchymal stem cells) IMR90 (fibroblast), and U2OS (osteosarcoma)); 3 representative cell lines are shown in figure 1. EwS cells showed a heightened sensitivity to splicing factor knockdown compared to control cells (Figure 1a,

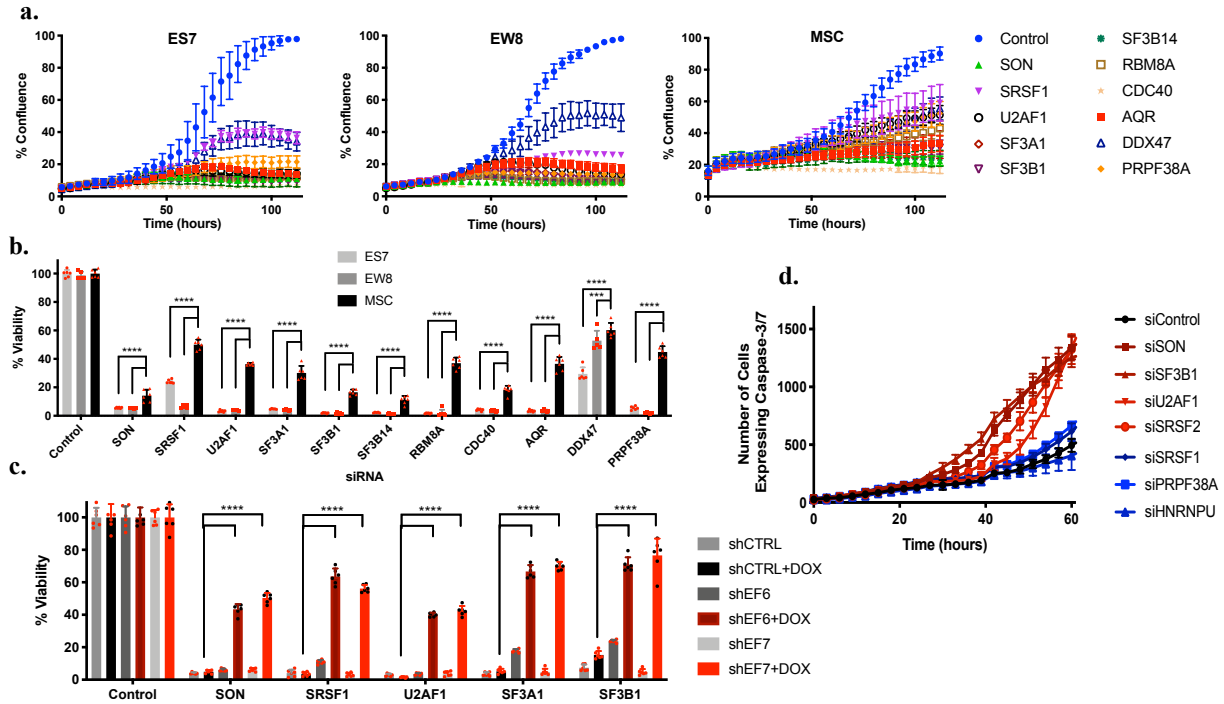

Figure 1. Growth (**a**) and Cell Titer-Glo viability (**b**) of EwS cells (ES7 and EW8) and MSC with siRNA knockdown of the indicated genes. **c**. Viability of TC32 cells with dox-inducible shRNA against EWS-FLI1 or control with knockdown of the indicated genes. **d**. Number of apoptotic TC32 cells (expressing caspase-3/7) over 60 hours following knockdown.

1b). Further, shRNA knockdown of EWS-FLI1 in the EwS cell line TC32 conferred resistance to loss of key splicing factors (Figure 1c). Depletion of some but not all splicing factor genes induced apoptosis in TC32 cells (Figure 1d).

The second major goal of the project was to determine the causal relationship between transcription dysregulation, R- loops, and splicing in Ewing sarcoma. We found that overexpression of RNASEH1, an enzyme that degrades R-loops, conferred resistance to loss of key splicing factors (Figure 2a). However, we were unable to obtain high transfection efficiency using transient transfection of RNASEH1. Therefore, we established stable inducible cell lines expressing RNASEH1 to use for these experiments. Consequently, these experiments were not completed in the original projected timeframe, but will be completed within the next reporting period.

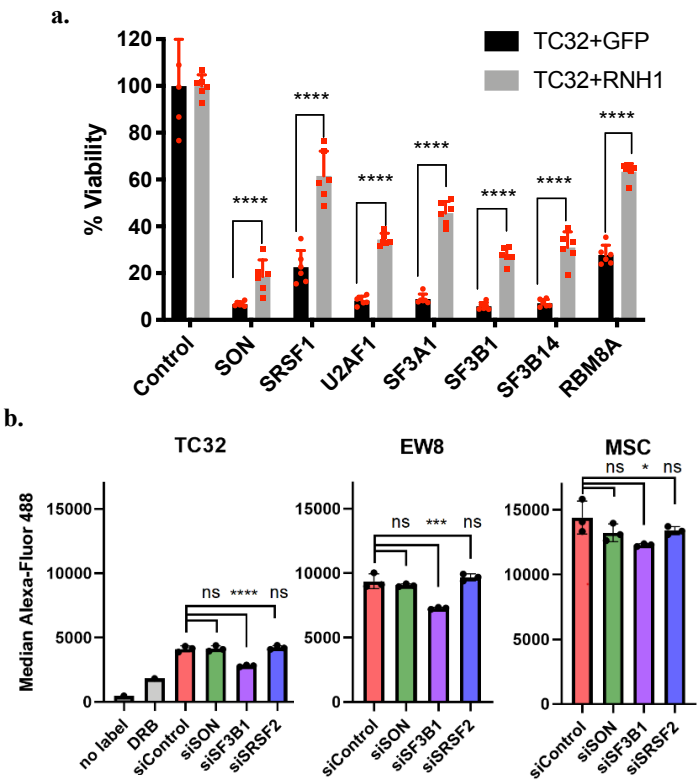

Figure 2. a. Viability of TC32 cells transfected with RNASEH1 or GFP with knockdown of the indicated genes. **b.** Quantification of ErU signal by flow cytometry with knockdown of the indicated genes.

To test effects on transcription, we used ethynyl ribouridine (ErU) incorporation/flow cytometry. Cells were treated with ErU to label nascent RNA, labeled by a click chemistry reaction, and analyzed by flow cytometry using the LSR Fortessa instrument. This revealed that knockdown of SON or SRSF2 had no significant effect on global transcription, although transcription decreased with knockdown of SF3B1 (Figure 2b). Additionally, we planned to use Global Run-On (GRO-seq) to examine active transcription genome-wide. Briefly, nuclei are extracted and active transcripts labeled with BrU, then pulled down and sequenced. GRO-seq experiments are in progress and expected to be completed within the next month.

We used DNA-RNA Immunoprecipitation (DRIP-seq) to examine R-loops genome-wide in the Ewing sarcoma cell line TC32 compared to MSC. Briefly, genomic DNA was extracted and fragmented using restriction enzyme digestion, then pulled down with the S9.6 antibody (which recognizes DNA-RNA hybrids) and sequenced. As expected, TC32 cells displayed increased levels of R-loops. We plan to

Table 1. Splicing Inhibitors.

Drug	Target
E7107	SF3B1
GEX1A	SF3B1
Pladienolide B	SF3B1
SPHINX31	SRPK1
T-025	CLK1, CLK2, CLK3, CLK4
AZ191	DYRK1a, DYRK1b
MS023	PRMT type I
GSK3326595	PRMT5
borrelidin	FBP21 (WBP4)

repeat the DRIP-seq in the context of splicing inhibition to determine its effect on R-loops and any differences in response between EwS and control cells.

The third major goal of the project was to determine whether splicing inhibitors show EwS-specific toxicity, alone or in combination with standard therapeutic agents. We tested a panel of 9 splicing inhibitors (Table 1) and found the most promising results with the SF3B1 inhibitor E7107, which was effective at low nanomolar concentrations and showed a good therapeutic window between EwS and control cell lines (Figure 3). We initially proposed to test E7107 in combination with etoposide, cyclophosphamide, and olaparib. Etoposide and cyclophosphamide are two of the chemotherapeutics in the standard treatment of EwS, and olaparib is a PARP1 inhibitor previously described to show toxicity to EwS in cell culture[1]. However, as these combinations showed no more than an additive effect, we additionally tested combinations of E7107 with vincristine, paclitaxel, and docetaxel. Vincristine is another standard EwS chemotherapeutic agent, and paclitaxel and docetaxel have shown synergy with splicing inhibition in other models. These three drugs showed synergy with E7107 in EwS cells, with vincristine showing the best differentiation between EwS and control cells (Figure 4a). 4 representative cell lines are shown of the 9 that were tested.

The fourth major goal of the project was to determine the biological effects of splicing inhibition in EwS cells. ErU incorporation assays revealed no significant impact of splicing inhibition on global transcription. We assayed cell cycle progression using flow cytometry (cells labeled with EdU and Fx Cycle Violet stain) and found that E7107 caused cells to accumulate in G1 phase (Figure 4c). We assayed apoptosis using an IncuCyte imaging assay in which live cells and cells expressing Caspase-3/7 are labeled with fluorescent markers, and found that E7107, vincristine, and the combination induced apoptosis in Ewing sarcoma cells but not in MSC (Figure 4b).

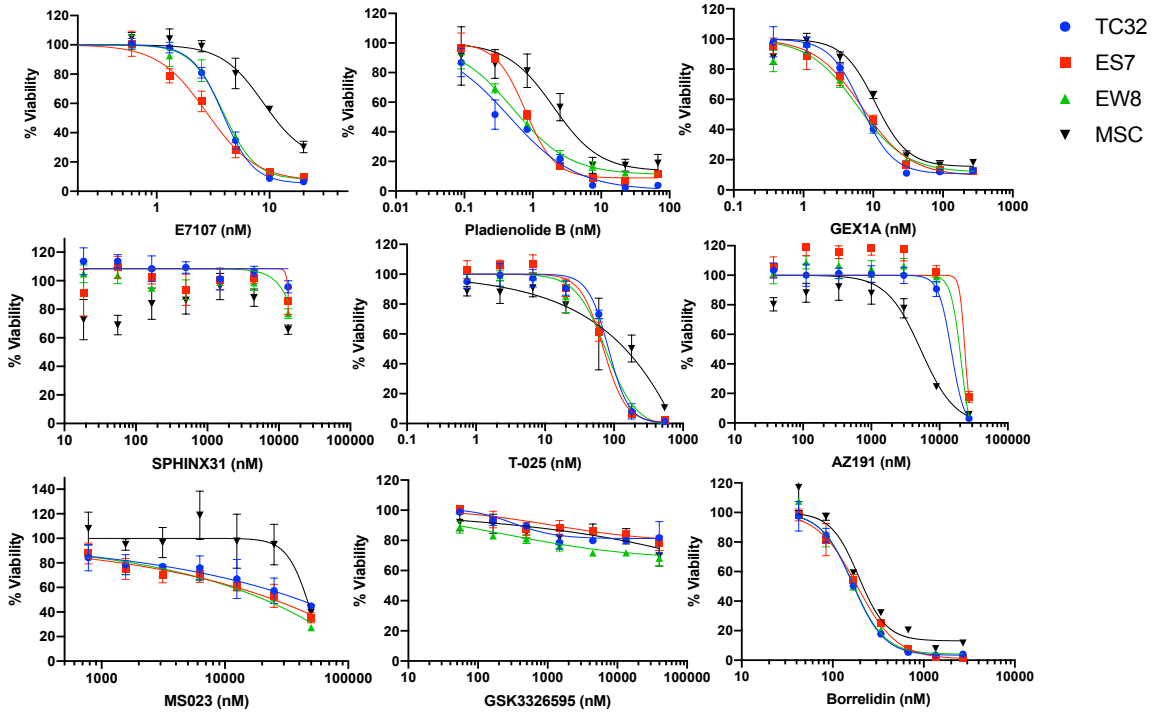

Figure 3. Cell viability of EwS or MSC cells following 3 days of treatment with splicing inhibitors (protein target indicated in parentheses).

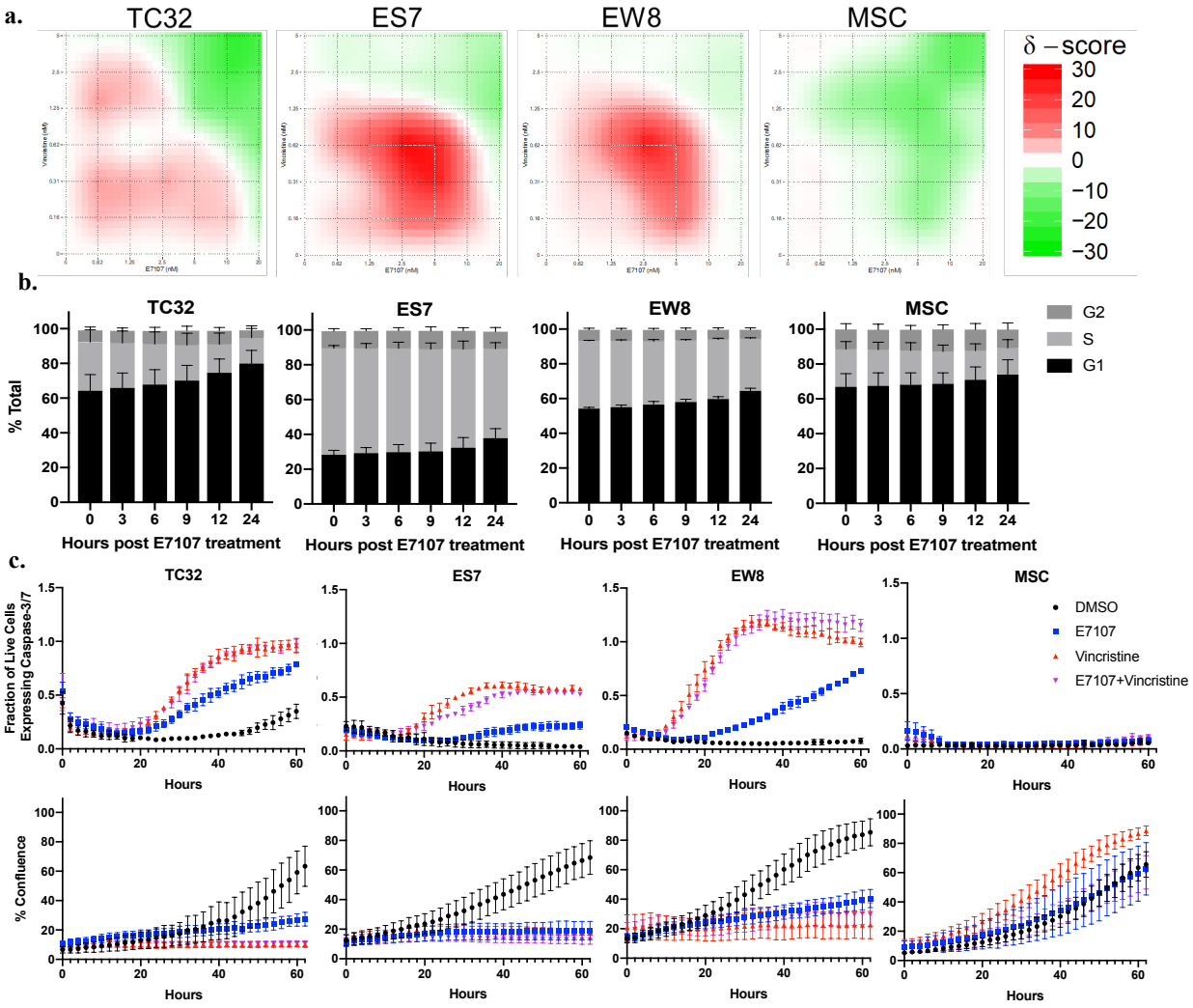

Figure 4. a. Heatmaps representing synergy between E7107 (x) and vincristine (y) analyzed using the Zero Interaction Potency model. Red indicates synergy and green indicates antagonism. **b.** Flow cytometry cell cycle analysis of cells treated with 2 nM E7107. **c.** Fraction of apoptotic cells (expressing caspase-3/7) over 60 hours of treatment (top) and corresponding growth curve (bottom).

Our results support the hypothesis that Ewing sarcoma is sensitive to splicing inhibition in a EWS-FLI1-dependent manner, providing a potential therapeutic window for targeting splicing in treatment. This sensitivity appears to be linked to the hyperphosphorylation of RNAPII that results from EWS-FLI1 expression and the resulting deregulated transcription and elevated R-loops.

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.

The main focus of this project is to train me as an independent investigator in the area of pediatric cancer research. Training activities included regular one-on-one and group meetings with my dissertation mentor, Dr. Alexander Bishop, to discuss methods, analysis, interpretation, and reporting of data. In addition, I held two meetings with my Dissertation Supervising Committee (12/9/19 and 5/14/20) to report and discuss progress on the project. I have also presented my work in a department presentation (4/27/20) and a short talk in the Greehey Children's Cancer Research Institute seminar series (2/7/20) to gain experience in presentation skills and obtain feedback from university faculty. Although COVID-19 travel restrictions prevented traveling to scientific meetings, I was able to present posters at the UT Health Cell Biology, Genetics, and Molecular Medicine student retreat (9/5/19) and Cell Systems and Anatomy department retreat (5/18/20) to discuss my work with other investigators and get helpful feedback. I co-authored a paper on Ewing sarcoma[11] using several of the methods I have become proficient in for this project, which will help me towards publishing the results of this project in the next year.

My training activities during the past year have also included participation in journal clubs and seminars. I have attended weekly Cancer Biology and Cell Biology/Genetics/Molecular Medicine journal clubs at UT Health San Antonio, in which students present and discuss recent original research articles with a substantial impact in biomedical research. I have also attended weekly seminars from a variety of seminar series including Cancer Development and Progression, Molecular Medicine, Pharmacology, and Precision of Science in Medicine, featuring both institutional faculty and top international scientists. This has helped me to keep up with current research in my own and related fields. Finally, I have attended a monthly Research Integrity workshop organized by the institutional training grant directors at UTHSA in which faculty and invited guests speak on topics such as conflicts of interest, research compliance, record keeping, reproducibility, peer review, and ownership of data.

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 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.

We are on track to complete all of the research described in the SOW in the next reporting period (7/1/20-6/30/21). We will perform all experiments requiring RNASEH1 overexpression to deplete R-loops using the stable inducible cell lines that we have established. We will complete GRO-seq experiments and bioinformatic analysis of GRO-seq and DRIP-seq to compare differences in transcription and R-loops. Additionally, we will further examine the biological effects of splicing inhibition in EwS cells using DNA combing and the DR-GFP assay for homologous recombination. We expect to submit a manuscript of our findings in a peer-reviewed scientific journal within the next year.

4. **IMPACT:** D*escribe 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).

Ewing sarcoma is a rare but very aggressive tumor of the bone and soft tissue that is most frequent in children, adolescents, and young adults. It is treated with chemotherapy along with radiation or surgery. Although most patients respond well to this aggressive treatment regimen, there is no second-line treatment when it fails, and there are severe side effects in children that can continue to affect them long after treatment has ended. Consequently, effective, less toxic, targeted treatment strategies are much needed. Our lab performed a genome-wide screen to identify genes required for Ewing sarcoma cells to survive. We found that Ewing sarcoma cells are particularly sensitive to loss of genes involved in RNA splicing. RNA splicing is additional processing that occurs after a gene is copied from DNA to RNA (transcription) before it can be used to make proteins. We recently published that the fusion oncogene that drives Ewing sarcoma, EWS-FLI1, increases the overall level of transcription activity in a cell[10]. This results in problems in transcription, with some of the newly made RNA sticking to the DNA, creating a structure called an R-loop. Worse, the whole transcription process is normally turned off/down in response to cellular damage, but this does not happen in Ewing sarcoma. This project aims to understand the relationship between altered regulation of transcription and splicing in Ewing sarcoma. A number of splicing inhibitors have been developed, one of which is currently in clinical trials for hematologic malignancies. We tested splicing inhibitors in Ewing sarcoma cells to determine whether this could be used as a new treatment strategy, either alone or in combination with standard chemotherapeutics. We found that Ewing sarcoma cells were extremely sensitive to splicing inhibition, which synergized with several standard chemotherapeutic agents to kill the cells at very low doses. Our data indicated that this sensitivity is linked to the hyperphosphorylation of RNAPII that results from EWS-FLI1 expression and the resulting deregulated transcription and elevated R-loops. Our results are expected to impact research in basic science by improving our understanding of regulation and coordination of transcription and splicing. In addition, our results point to RNA splicing as a potential new therapeutic target in Ewing sarcoma, which has the potential to benefit all Ewing sarcoma patients, especially those with chemo-resistant disease.

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.

We plan to add several experiments in order to better elucidate the protein-protein interactions between phospho-RNAPII, splicing factors and EWS-FLI1. Our initial model was that the hyperactivation of transcription simply produces such a large volume of transcripts that it overwhelms the RNA processing machinery, and this and the resulting R-loop accumulation leads to a sensitivity to any further perturbation of splicing. However, recent publications investigating interactions of phospho-RNAPII have indicated that phosphorylation removes RNAPII from phase separated condensates associated with Mediator and initiation and moves it to a new droplet including splicing factors and associated with elongation[12,13]. This process would likely be impacted by EWS-FLI1 and its disruption of RNAPII regulation. This raises the possibility that the splicing defect is in part due to sequestration of splicing factors with phosphorylated RNAPII as well as R-loop associated stress. To test this, we will first perform chromatin and nuclear co-immunoprecipitation in TC32 (EwS) and MSC (control) cell lines with and without E7107 treatment. We expect to see increased association between pRNAPII and splicing factors in EwS cells. We will further characterize this with a reporter assay[13] in which the RNAPII C-terminal domain is fused to Lac binding domain and CFP and transfected into U2OS cells carrying Lac operon repeats, and SF3B1 or SRSF2 recruitment is detected by IF in the presence or absence of EWS-FLI1. These experiments will promote the goals of the project by providing further insight into the relationship between transcription dysregulation, R- loops, and splicing in Ewing sarcoma.

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.

Since we were unable to obtain high transfection efficiency using transient transfection of RNASEH1, we established stable inducible clonal cell lines expressing RNASEH1 to use for all experiments requiring R-loop depletion. Consequently, these experiments have not yet been completed but, as we have successfully established these cell lines, we expect they will be completed within the next reporting period.

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.

Nothing to report.

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

Nothing to report.

Significant changes in use or care of vertebrate animals

Nothing to report.

Significant changes in use of biohazards and/or select agents

Nothing to report.

6. **PRODUCTS:** *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).*

Miller H, Gorthi A, Bassani N, Lawrence L, Iskra B, Bishop A. Reconstruction of Ewing Sarcoma Developmental Context from Mass-Scale Transcriptomics Reveals Characteristics of EWSR1-FLI1 Permissibility. Cancers. 2020 April 11; 12(4), 948. Status: published; acknowledgement of federal support: yes

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).*

Nothing to report.

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.*

Poster presentations:
Lawrence L, Gorthi A, Chen Y, Bishop A. Targeting the Dysregulation of Transcription and Splicing in Ewing Sarcoma. Cell Systems and Anatomy Departmental Retreat, UT Health San Antonio. San Antonio, TX. May 18, 2020. *Poster Competition – Honorable Mention*

Lawrence L. Transcription, R-loops, and RNA Splicing in Ewing Sarcoma. Cell Biology, Genetics, and Molecular Medicine Retreat, UT Health San Antonio; September 5, 2019; San Antonio, TX.

Short talks:
Lawrence L. Dysregulation of Transcription and Splicing in Ewing Sarcoma. Greehey Children's Cancer Research Institute seminar series, UT Health San Antonio; February 7, 2020; San Antonio, TX.

- **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.

Nothing to report.

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

Nothing to report.

- **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.

Nothing to report.

- **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. Examples include:

- *data or databases;*
- *physical collections;*
- *audio or video products;*
- *software;*
- *models;*
- *educational aids or curricula;*
- *instruments or equipment;*
- *research material (e.g., Germplasm; cell lines, DNA probes, animal models);*
- *clinical interventions;*
- *new business creation; and*
- *other.*

Nothing to report.

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

What individuals have worked on the project?

Provide the following information for: (1) PDs/PIs; and (2) each person who has worked at least one person month per year on the project during the reporting period, regardless of the source of compensation (a person month equals approximately 160 hours of effort). If information is unchanged from a previous submission, provide the name only and indicate "no change".

Example:

Name:	*Mary Smith*
Project Role:	*Graduate Student*
Researcher Identifier (e.g. ORCID ID):	*1234567*
Nearest person month worked:	*5*
Contribution to Project:	*Ms. Smith has performed work in the area of combined error-control and constrained coding.*
Funding Support:	*The Ford Foundation (Complete only if the funding support is provided from other than this award.)*

Name:	Liesl Lawrence
Project Role:	PI (Graduate Student)
Researcher Identifier (e.g. ORCID ID):	https://orcid.org/0000-0001-8809-6278
Nearest person month worked:	12
Contribution to Project:	Ms. Lawrence the laboratory experiments and data analysis.
Funding Support:	N/A

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.

Liesl Lawrence (PI): Nothing to report

Alexander Bishop (Mentor):
NIH/NCI R01 CA241554 (*PI: Bishop*) 05/2020 - 04/2025
Dysregulated transcription processes in Ewing sarcoma
To determine the consequences of dysregulated transcription regulation in Ewing sarcoma.

SU2C-CRUK Pediatric Cancer New Discoveries Challenge (*PI: Bishop*) 10/2020 - 09/2022
Targeting R-loop stability in Ewing sarcoma
To determine the therapeutic utility of targeting R-loop stabilizing mechanisms in Ewing sarcoma.

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.*

H3 Biomedicine (Cambridge, MA) supplied E7107, a splicing inhibitor that has clinical relevance.

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.*

References

1 Grunewald, T. G. P. *et al.* Ewing sarcoma. *Nature reviews. Disease primers* **4**, 5, doi:10.1038/s41572-018-0003-x (2018).

2 Chansky, H. A., Hu, M., Hickstein, D. D. & Yang, L. Oncogenic TLS/ERG and EWS/Fli-1 fusion proteins inhibit RNA splicing mediated by YB-1 protein. *Cancer research* **61**, 3586-3590 (2001).

3 Paronetto, M. P., Minana, B. & Valcarcel, J. The Ewing sarcoma protein regulates DNA damage-induced alternative splicing. *Molecular cell* **43**, 353-368, doi:10.1016/j.molcel.2011.05.035 (2011).

4 Selvanathan, S. P. *et al.* Oncogenic fusion protein EWS-FLI1 is a network hub that regulates alternative splicing. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E1307-1316, doi:10.1073/pnas.1500536112 (2015).

5 Yang, L., Chansky, H. A. & Hickstein, D. D. EWS.Fli-1 fusion protein interacts with hyperphosphorylated RNA polymerase II and interferes with serine-arginine protein-mediated RNA splicing. *J Biol Chem* **275**, 37612-37618, doi:10.1074/jbc.M005739200 (2000).

6 Bentley, D. L. Coupling mRNA processing with transcription in time and space. *Nature reviews. Genetics* **15**, 163-175, doi:10.1038/nrg3662 (2014).

7 Lee, K. M. & Tarn, W. Y. Coupling pre-mRNA processing to transcription on the RNA factory assembly line. *RNA biology* **10**, 380-390, doi:10.4161/rna.23697 (2013).

8 Naftelberg, S., Schor, I. E., Ast, G. & Kornblihtt, A. R. Regulation of alternative splicing through coupling with transcription and chromatin structure. *Annual review of biochemistry* **84**, 165-198, doi:10.1146/annurev-biochem-060614-034242 (2015).

9 Ip, J. Y. *et al.* Global impact of RNA polymerase II elongation inhibition on alternative splicing regulation. *Genome research* **21**, 390-401, doi:10.1101/gr.111070.110 (2011).

10 Gorthi, A. *et al.* EWS-FLI1 increases transcription to cause R-loops and block BRCA1 repair in Ewing sarcoma. *Nature* **555**, 387-391, doi:10.1038/nature25748 (2018).

11 Miller, H. E. *et al.* Reconstruction of Ewing Sarcoma Developmental Context from Mass-Scale Transcriptomics Reveals Characteristics of EWSR1-FLI1 Permissibility. *Cancers (Basel)* **12**, doi:10.3390/cancers12040948 (2020).

12 Guo, Y. E. *et al.* Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature* **572**, 543-548, doi:10.1038/s41586-019-1464-0 (2019).

13 Boija, A. *et al.* Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell* **175**, 1842-1855.e1816, doi:10.1016/j.cell.2018.10.042 (2018).

W81XWH-19-1-0180: Transcription, R-Loops, and RNA Splicing in Ewing Sarcoma

PI: Liesl Lawrence, University of Texas Health Science Center at San Antonio, TX. **Budget:** $220,350

Topic Area: Cancer in Children, Adolescents, and Young Adults

Mechanism: FY18 Peer Reviewed Cancer Research Program; Horizon Award

Research Area(s): 0209 (Transcription), 0805 (Targeted Therapies)
Award Status: 1 July 2019 – 30 June 2021

Study Goals:

Ewing sarcoma is an aggressive pediatric bone and soft tissue cancer driven primarily by the EWS-FLI1 fusion oncogene. EWS-FLI1 acts as a transcription factor and also interferes with normal regulation of transcription and transcription-associated RNA processing. The goal of this project is to evaluate the hypothesis that EWS-FLI1 driven hyper-activation of transcription causes a targetable dependence on RNA splicing in Ewing sarcoma.

Specific Aims:

1) Determine the mechanistic relationship between EWS-FLI1-driven transcription, R-loops, and splicing vulnerabilities in Ewing Sarcoma.
   a) Confirm the dependence of Ewing sarcoma cells on the splicing machinery.
   b) Elucidate the cause and effect relationships between transcription dysregulation, R-loops, and splicing in Ewing sarcoma cells.
2) Evaluate the potential of splicing as a therapeutic target in Ewing sarcoma.
   a) Determine whether splicing inhibitors show Ewing sarcoma-specific toxicity, alone or in combination with standard therapeutic agents.
   b) Determine the biological effects of splicing inhibition in Ewing sarcoma cells.

Key Accomplishments and Outcomes:

Publications:

Miller H, Gorthi A, Bassani N, Lawrence L, Iskra B, Bishop A. Reconstruction of Ewing Sarcoma Developmental Context from Mass-Scale Transcriptomics Reveals Characteristics of EWSR1-FLI1 Permissibility. Cancers. 2020 April 11; 12(4), 948.

Patents: none to date

Funding Obtained: none to date

cancers

MDPI

Article

Reconstruction of Ewing Sarcoma Developmental Context from Mass-Scale Transcriptomics Reveals Characteristics of EWSR1-FLI1 Permissibility

Henry E. Miller [1,2], Aparna Gorthi [1,2], Nicklas Bassani [2], Liesl A. Lawrence [1,2], Brian S. Iskra [1,2] and Alexander J. R. Bishop [1,2,*]

1 Department of Cell Systems and Anatomy, University of Texas Health at San Antonio, San Antonio, TX 78229, USA; millerh1@livemail.uthscsa.edu (H.E.M.)
2 Greehey Children's Cancer Research Institute, University of Texas Health at San Antonio, San Antonio, TX 78229, USA
* Correspondence: bishopa@uthscsa.edu; Tel.: +1-210-562-9060

Received: 29 March 2020; Accepted: 10 April 2020; Published: 11 April 2020

Abstract: Ewing sarcoma is an aggressive pediatric cancer of enigmatic cellular origins typically resulting from a single translocation event t (11; 22) (q24; q12). The resulting fusion gene, *EWSR1-FLI1*, is toxic or unstable in most primary tissues. Consequently, attempts to model Ewing sarcomagenesis have proven unsuccessful thus far, highlighting the need to identify the cellular features which permit stable EWSR1-FLI1 expression. By re-analyzing publicly available RNA-Sequencing data with manifold learning techniques, we uncovered a group of Ewing-like tissues belonging to a developmental trajectory between pluripotent, neuroectodermal, and mesodermal cell states. Furthermore, we demonstrated that EWSR1-FLI1 expression levels control the activation of these developmental trajectories within Ewing sarcoma cells. Subsequent analysis and experimental validation demonstrated that the capability to resolve R-loops and mitigate replication stress are probable prerequisites for stable EWSR1-FLI1 expression in primary tissues. Taken together, our results demonstrate how EWSR1-FLI1 hijacks developmental gene programs and advances our understanding of Ewing sarcomagenesis.

Keywords: Ewing sarcoma; EWSR1-FLI1; transcriptomics; manifold learning; single cell biology; R-loops; replication stress; sarcomagenesis; developmental trajectories; cell identity

1. Introduction

Ewing sarcoma is an aggressive cancer of the bone and soft tissues resulting from the fusion of Ewing sarcoma RNA Binding Protein 1 (*EWSR1*) to Friend leukemia integration 1 transcription factor (*FLI1*) (85%) or E-twenty-six-related (ETS-related) gene (*ERG*) (15%) [1]. Ewing sarcoma's clinical histopathology is heterogenous with some tumors displaying more mesenchymal or more neuroectodermal characteristics [2]. Given the relative paucity of mutations in these tumors [3], Ewing sarcoma's heterogeneity is likely determined by the cell from which it arises. However, ectopic EWSR1-FLI1 expression is toxic or unstable in most primary cell types and, to date, no successful model of Ewing sarcomagenesis has been established [4–7]. These facts highlight the urgent need to understand the contextual requirements for stable EWSR1-FLI1 expression within normal tissue types.

Owing to the recent efforts of Lachmann et al. in reprocessing hundreds of thousands of publicly available transcriptomics datasets, it is now computationally feasible to mine gene expression data at mass-scale from a vast range of tissue types, diseases, and drug conditions [8]. Manifold learning provides a set of tools for analyzing high dimensionality data, revealing both global and local

relationships [9]. In this study, we utilized manifold learning approaches to analyze 40,903 bulk transcriptomes of normal tissues and Ewing sarcoma. From this analysis, we elucidated the transcriptomic relationship between Ewing sarcoma and the context of normal developmental tissues from which it likely arises, uncovering cellular features which likely provide permissibility for stable EWSR1-FLI1 expression.

2. Results

2.1. Ewing Sarcoma Developmental Context Reconstructed from Mass-Scale Transcriptomics

To elucidate the transcriptomic context in which Ewing sarcoma arises, we processed 40,903 bulk transcriptomes of normal tissues and Ewing sarcoma samples obtained from the ARCHS[4] data repository [8]. First, we implemented a Uniform Manifold Approximation and Projection (UMAP) embedding to reveal the global similarities and differences between the samples in our data set [10]. In coordination with Louvain clustering [11], this approach accurately reconstructed known biological groups, such as hepatic, adipose and immune tissues (Figures 1A and S1A).

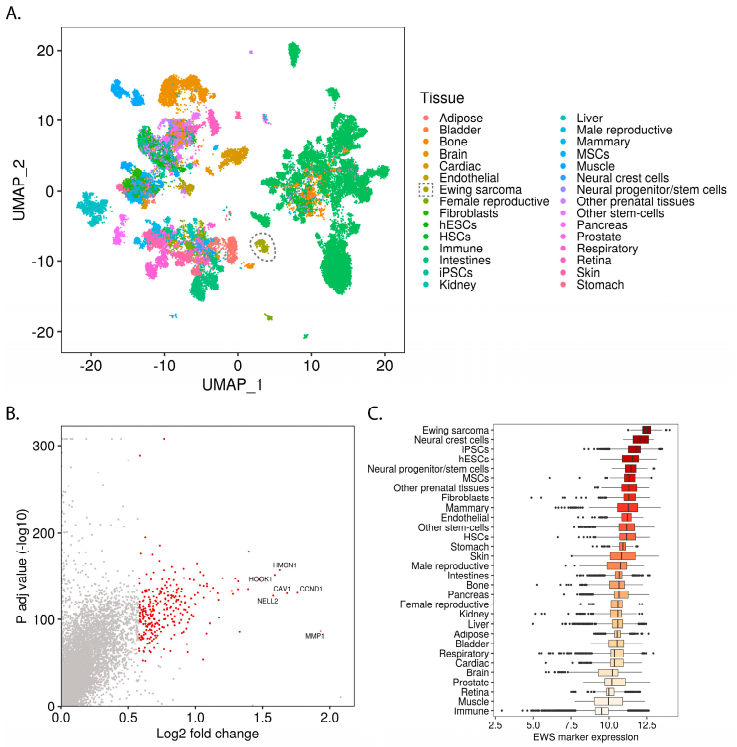

Figure 1. Mass-scale mining of bulk transcriptomes reveals the relationship of Ewing sarcoma to normal tissue types: (**A**) UMAP embedding of bulk transcriptomic profiles from normal tissues and Ewing sarcoma samples (Ewing samples circled); (**B**) One-way volcano plot showing the top Ewing sarcoma marker genes from Wilcoxon rank-sum testing with Bonferroni correction; (**C**) Box-plot comparing the Ewing sarcoma marker gene expression levels for different samples (median of variance stabilizing transform (VST) transformed and geometric mean normalized read counts for all Ewing sarcoma marker genes within each sample), grouped by tissue and ordered by median expression.

To uncover the normal tissues which display Ewing-like transcriptomic profiles, we calculated a set of Ewing sarcoma marker genes (Figure 1B) and then determined which normal tissues express these genes at high levels (Figures 1C and S1C).

Ewing sarcoma marker genes were calculated by comparing the gene expression of Ewing sarcoma samples to that of the normal tissue samples. The resulting gene set is highly enriched for direct EWSR1-FLI1 targets (Figure S1D) and for genes previously associated with Ewing sarcoma (custom "Ewing sarcoma gene set", see Methods) (Figure S1E), indicating the validity of our approach.

To uncover the most Ewing-like normal tissues, we grouped samples by tissue type (Figure 1C) and cluster membership (Figure S1B) and ranked them by median Ewing sarcoma marker gene expression level (Figures 1C, S1B, S2 and S3). This analysis demonstrated the similarity of Ewing sarcoma with pluripotent stem-cells (iPSCs and hESCs), neural crest cells, neural progenitors, and multipotent mesenchymal stromal cells (MSCs) among others (Figure 1C). Given that these tissues are involved in known developmental transitions (e.g., gastrulation), we hypothesized that an analysis of Ewing sarcoma within this context would reveal developmental programs involved in Ewing sarcomagenesis. Furthermore, we suspected that this analysis would also elucidate characteristics of normal developmental tissue types which permit the stable expression of EWSR1-FLI1.

Of note, while cluster membership tended to follow tissue annotation, there were notable exceptions relevant to several Ewing-like clusters (Figure S3). This result indicated that Ewing-like transcriptional states might arise within subtypes of otherwise non-Ewing-like tissue categories. Consequently, we defined "Ewing-like normal tissues" based on which samples belonged to the group of top Ewing-like clusters (Figure S2; clusters 17, 16, 24, 2, 19, 4, and 28) rather than the top Ewing-like tissue categories.

While UMAP is highly effective at grouping samples based on transcriptomic similarities, the Potential of Heat-diffusion for Affinity-based Trajectory Embedding (PHATE) algorithm has demonstrated superiority in revealing subtle transitions between cell states [12]. We applied the PHATE algorithm to reveal the location of Ewing sarcoma in the transitions among Ewing-like normal tissues (Figure 2A). When labeled for germ layer lineage, the PHATE embedding reveals Ewing sarcoma's location within this developmental context (Figure 2B; Video S1). Interestingly, even when Ewing sarcoma samples were removed from the analysis, the shape of the PHATE embedding was not noticeably altered (Figure S4). Given that PHATE organized Ewing sarcoma cells along a normal developmental transition from pluripotent/neuroectodermal to mesodermal tissues, the possibility was raised that EWSR1-FLI1 controls lineage-specific gene programs that define cellular identity within these progenitor populations.

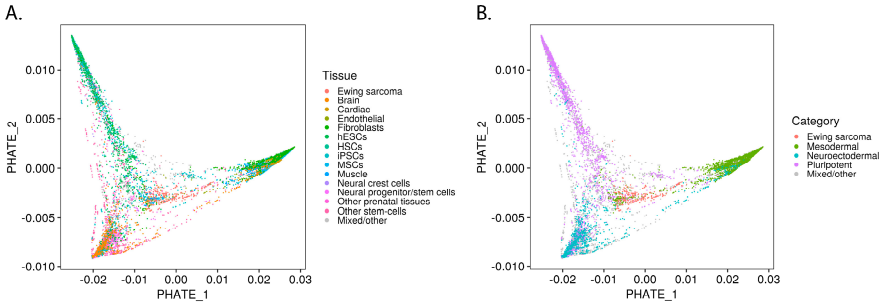

Figure 2. Ewing sarcoma developmental context reconstructed from bulk transcriptomes of Ewing-like tissues: (**A**) PHATE embedding of top Ewing-like samples; (**B**) PHATE embedding labeled by developmental lineage.

2.2. EWSR1-FLI1 *Expression Levels Determine Ewing Sarcoma's PHATE within a Developmental Context*

It was previously reported that ectopic EWSR1-FLI1 expression induces neuroectodermal marker expression in MSCs [13] and that EWSR1-FLI1 knockdown in Ewing sarcoma cells promotes mesenchymal marker expression [14]. These findings led us to hypothesize that EWSR1-FLI1 activity controls the position of Ewing sarcoma in the pluripotent/neuroectodermal to mesodermal

developmental trajectory along PHATE_1 (Figure 2B; Video S1). Studies which involved depleting EWSR1-FLI1 were analyzed for corresponding changes in PHATE_1 position (Figure S5). It was found in each case that EWSR1-FLI1 knock-down drove cells higher on PHATE_1 towards the mesodermal branch and away from the pluripotency/neuroectodermal lineage branches. This phenomenon is exemplified by a comparison of samples with and without EWSR1-FLI1 shRNA knockdown (transcriptomic data obtained from Howarth et al. [15]; GSE60949) (Figure 3B). To confirm this finding, we first calculated the Pearson correlation of gene expression and PHATE_1 position across Ewing samples, yielding a "PHATE_1 correlation score" (signed R^2) for every gene. This revealed the genes which drive samples higher on PHATE_1 and vice versa (Figure 3C). After ranking genes by their PHATE_1 correlation score, we were able to determine what pathways were correlated with higher and lower PHATE_1 positions using gene set enrichment analysis (GSEA) [16] (Figure 3D). From this analysis we found that markers of low EWSR1-FLI1 expression were strongly correlated with increasing PHATE_1 scores and vice versa. In agreement with the previous analysis, this result also indicates that the transition from low to high EWSR1-FLI1 expression correlates with the transition from mesodermal to pluripotent/neuroectodermal cell states in normal tissues. This result was further confirmed by GSEA of other pathways correlated with Ewing sarcoma's position in PHATE_1, using gene sets from the Molecular Signatures Database (MSigDB) Chemical and Genetic Perturbations (C2:CGP) collection [17]. As expected, the correlation of gene expression with PHATE_1 in Ewing cells was significantly enriched for mesenchymal-like cancer pathways (in the case of positive correlations), such as "Verhaak Glioblastoma Mesenchymal", and pluripotent-like pathways (in the case of negative correlations), such as "Wong Embryonic Stem Cell Core" (Figure S7A). These results further confirmed our observation that EWSR1-FLI1 expression pushes cells along an innate developmental trajectory between mesodermal and pluripotent/neuroectodermal cell states. In addition to EWSR1-FLI1 knock-down, there were several other interventions which significantly pushed Ewing sarcoma along this developmental trajectory (Figure S6).

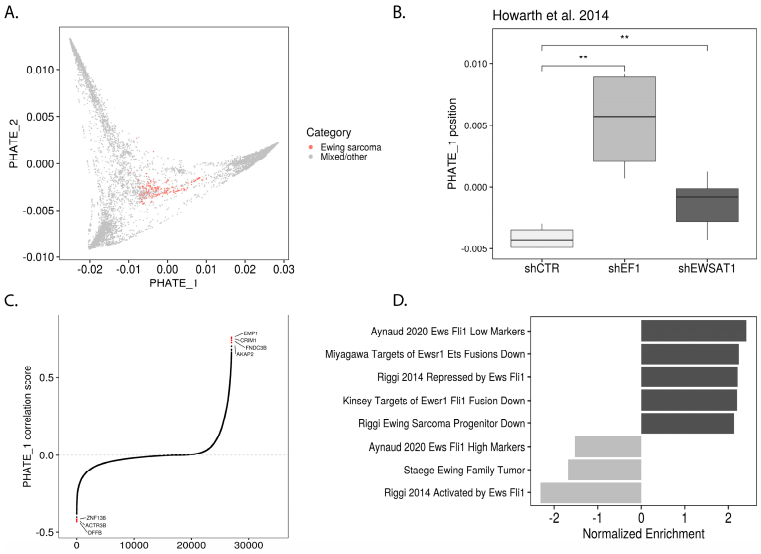

Figure 3. Ewing sarcoma's position in underlying developmental trajectory controlled by EWSR1-FLI1 expression levels: (**A**) PHATE embedding with Ewing sarcoma samples highlighted; (**B**) Box-plot showing difference in location along PHATE_1 between A673 cells exposed to control shRNA or shRNA targeting EWSR1-FLI1 (shEF1) and Ewing sarcoma associated transcript 1 (EWSAT1) [15] (one-tail *t* test, ** $p \leq 0.01$); (**C**) Genes in Ewing sarcoma samples ranked by PHATE_1 correlation score (signed R^2); (**D**) Bar-plot showing enrichment of Ewing sarcoma gene sets within PHATE_1 correlation scores as determined by GSEA.

It was previously reported that lysine-specific histone demethylase 1 (LSD1) inhibition disrupts the Ewing sarcoma transcriptome [18]. In agreement with this finding, we found that LSD1-inhibiting interventions like SP2509 treatment and LSD1 knock-down pushed Ewing sarcoma higher on PHATE_1 (Figure S6B–D). The response to LSD1 inhibition was observed in vitro, but, as LSD1 inhibitors are currently being tested clinically for Ewing sarcoma, it remains to be evaluated whether the same response would occur in vivo. Furthermore, recent literature indicates that EWSR1-FLI1 antagonizes TEA domain transcription factor 1 (TEAD1) transcriptional programs [19]. We found that inhibition of TEAD1 pushes Ewing sarcoma lower on PHATE_1, indicating that this antagonism is likely bi-directional (Figure S6A).

To test whether Ewing sarcoma's PHATE_1 gene correlations were distinct from those of the underlying developmental context, these analyses were repeated in the absence of any Ewing samples and the results were compared (Figure S7). Quite surprisingly, a significant overlap in C2:CGP and Ewing sarcoma gene set enrichment was observed between the gene correlations along PHATE_1 calculated from Ewing sarcoma samples and those calculated from the Ewing-like normal tissues (Figure S7C,D). The conservation of Ewing sarcoma pathway enrichment in the transition between normal tissue states provides further confirmation that EWSR1-FLI1 controls the movement of cells along this innate developmental trajectory. Furthermore, the enrichment of Ewing sarcoma gene sets in the transitions among primary tissue types indicates that Ewing sarcoma gene sets are largely markers of cellular identity rather than bona fide markers of Ewing sarcoma.

2.3. PHATE_1 Gene Scores Identify Mesenchymal-Like Cellular Subpopulation in Ewing Sarcoma Single Cell Transcriptomes

Recent reports indicate that EWSR1-FLI1 expression levels play a role in defining tumor heterogeneity, particularly in defining proliferative and migratory subpopulations [14,20]. In the above results, we found that EWSR1-FLI1 pushes Ewing sarcoma cells along a developmental trajectory between pluripotent/neuroectodermal and mesodermal cell states. Consequently, we hypothesized that developmental gene expression profiles would also be evident at the single cell level and correlate with markers of EWSR1-FLI1 expression. We generated single cell transcriptomes of Ewing sarcoma cell lines and merged them with recently published single cell profiles of patient-derived xenografts (PDXs) (Figures 4A and S8A,B). Cell cycle states and Louvain clusters were assigned to each cell (Figure 4B,C). Importantly, the proportion of cells in G1 compared to G2 or S phase depended on whether cells were derived from a PDX or cell line (Figure S8C,D).

To uncover cellular subpopulations shaped by developmental gene expression programs, we took the PHATE_1 correlation scores for each gene and used them to weight the gene expression profile of each single cell (see Methods). By calculating the median of these weighted gene expression levels, we assigned a "PHATE_1 score" to each cell. Similar to the bulk analysis, a higher PHATE_1 score indicates a cell has a more mesodermal transcriptome and a lower PHATE_1 score indicates a more pluripotent/neuroectodermal transcriptome. To uncover cellular subpopulations which display evidence of higher or lower PHATE_1 score, we ranked each cluster by the median PHATE_1 scores of the cells which belong to it (Figure 4D).

The top PHATE_1-high cluster (Cluster_8) was further characterized by the identification of positive and negative marker genes via differential gene expression analysis of this cluster compared to all others (Figure 4E; Table S3). Pathway enrichment was conducted using the positive and negative selection markers and gene sets from multiple MSigDB collections (custom "integrated" gene set collection, see Methods) (Figure 4E; Table S3). The enrichment of high PHATE_1 scores (mesodermal branch) in this cluster was mirrored by the pathway enrichment for mesenchymal-like gene sets such as "Hallmark Epithelial Mesenchymal Transition", loss of cell cycle activation indicated by "Iglesias E2f Targets Up (Down)", and for the EWSR1-FLI1-low gene set "Kinsey Targets of Ewsr1 Fli1 Fusion Down". These results recapitulated the bulk analysis, finding that EWSR1-FLI1 expression levels control the activation of developmental gene programs and lineage commitment. Furthermore, this

analysis demonstrated the relevance of these developmental programs in defining Ewing sarcoma tumor heterogeneity.

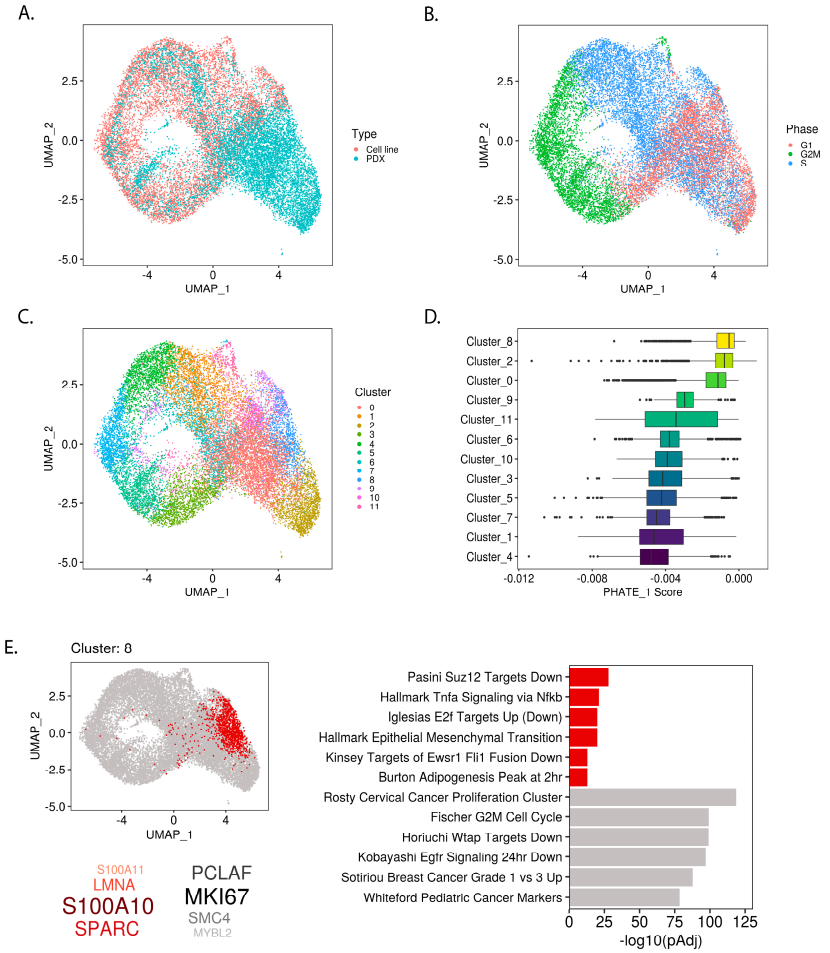

Figure 4. Mesodermal-like subpopulation within single-cell Ewing sarcoma transcriptomes revealed by PHATE_1 expression scores: (**A**) UMAP embedding of single-cell Ewing sarcoma transcriptomes in cell lines and patient-derived xenografts (PDXs) after alignment; (**B**) UMAP showing cell-cycle phase of each cell imputed from gene expression data and (**C**) Louvain clustering assignments; (**D**) Box-plot showing the PHATE_1 expression scores for each sample in each cluster, organized by median score; (**E**) Cluster marker analysis of cluster 8 shows location within UMAP embedding, the top positive marker genes (red) and negative marker genes (grey), and top enriched pathways (with the "integrated" gene set collection) in positive and negative marker genes (red and grey respectively).

2.4. Permissible PHATE_1-Low Tissues Show Important Transcriptional Similarities with EWSR1-FLI1 *Transcriptome*

EWSR1-FLI1 expression is toxic to most cells into which it is introduced. However, several primary cell types are capable of stable ectopic EWSR1-FLI1 expression: iPSCs, hESCs (only tested with p53 knock-down), neural crest cells, and neural crest-derived MSCs (NC-MSCs) [6,21–23]. Interestingly, each of these tissue types (except for NC-MSCs which were not available for analysis) belongs to the PHATE_1-low developmental context (Figure 2A) and display high expression of Ewing sarcoma marker genes (Figure 1C). These findings led us to hypothesize that PHATE_1-low tissues can

stably express EWSR1-FLI1 because of their basal similarity with the EWSR1-FLI1 transcriptome. To characterize the shared biological features that could account for permissibility, we performed a comparative pathway enrichment for genes of the EWSR1-FLI1 transcriptome and markers of PHATE_1-low tissues (Figure 5). The "EWSR1-FLI1 Transcriptome" was defined as the genes which are activated by EWSR1-FLI1 or increase after ectopic EWSR1-FLI1 expression (see Methods for additional details). The set of "PHATE_1-low Markers" were defined as the genes with a PHATE_1 correlation score <−0.2 in normal tissues. For both groups, pathway enrichment (gene set overrepresentation) was calculated using a collection of gene sets from MSigDB ("integrated" collection, see Methods For additional details). A highly significant overlap was found (Figure 5A).

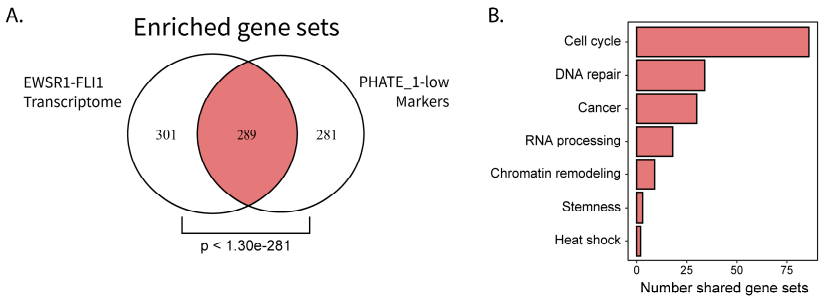

Figure 5. Comparative pathway enrichment reveals potential characteristics of EWSR1-FLI1 permissibility: (**A**) Venn diagram comparing pathway enrichment results for EWSR1-FLI1 transcriptome and PHATE_1-low gene markers (hypergeometric test p value displayed); (**B**) Bar plot showing number of shared gene sets assigned to integrative biological categories.

The 289 shared gene sets were assigned to general biological categories (see Methods for additional details) and the number of gene sets in each category was compared (Figure 5B). Interestingly, the top results belonged to biological processes related primarily with cell cycle, DNA damage response, and RNA processing.

2.5. The Capability to Resolve R-Loops and Replication Stress are Probable Requirements for Stable EWSR1-FLI1 *Expression*

Ewing sarcoma cells display high rates of replication and transcription [24], which is unsurprising given their cancer status. However, recent reports indicate that pluripotent tissues also display these characteristics [25–27]. Furthermore, high rates of transcription combined with high rates of proliferation can create transcriptional stress and genome instability via the production of R-loops (genomic structures formed by the hybridization of RNA and DNA) [28]. Consequently, we hypothesized that factors which resolve R-loops and replication stress are required to support the high replication and transcription rates resulting from EWSR1-FLI1 expression. If correct, we predict the following should also be true: (1) PHATE_1-low, but not PHATE_1-high, tissues display high levels of R-loop accumulation, (2) that Ewing sarcoma cells depend upon factors which resolve R-loops and replication-stress, and (3) that this dependency is shared by PHATE_1-low tissue types but not PHATE_1-high tissue types.

If our first assertion is correct, then it follows that PHATE_1-low (pluripotent/neuroectodermal) tissues would show significantly higher accumulation of R-loops compared to PHATE_1-high (mesodermal) tissues. We mined publicly available DNA-RNA immunoprecipitation sequencing (DRIP-Seq) data from a recently published gene expression omnibus (GEO) accession (GSE145964; Table S4) to test this assertion. The data was re-processed and the number of peaks (R-loop sites) in each cell type was compared (Figure 6A). This analysis revealed that there are significantly more R-loops in PHATE_1-low cell types (iPSCs, NSCs, and hESCs) compared to a PHATE_1-high cell type

(MSCs). Furthermore, we noted that a large number of DRIP-Seq peaks are found exclusively in iPSCs compared MSCs (Figure 6B).

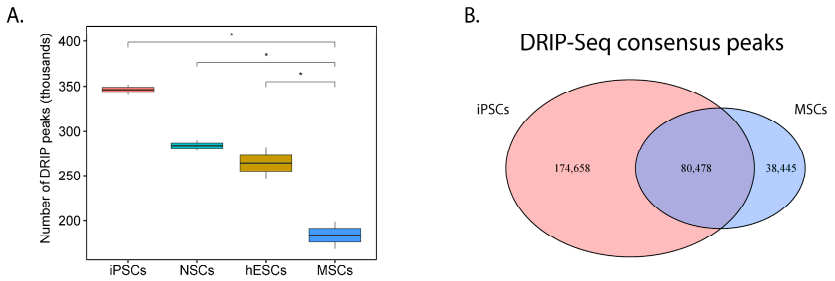

Figure 6. R-loop accumulation is a shared feature of both Ewing sarcoma and its PHATE_1-low developmental context: (**A**) Box-plot comparing number of detected DRIP-Seq peaks (R-loop sites) between cell types found in Ewing sarcoma's developmental context (iPSCs, induced pluripotent stem cells; NSCs, neural stem cells; hESCs, human embryonic stem cells; MSCs, multipotent mesenchymal stromal cells) (one-tail *t* test, * $p \leq 0.05$); (**B**) Venn diagram comparing DRIP-Seq peaks in iPSCs compared to MSCs.

Second, we predicted that Ewing sarcoma would depend upon R-loop and replication-stress-mitigating factors to maintain its high levels of transcription and proliferation. Previously, we and others noted that Ewing sarcoma cells activate the Ataxia telangiectasia and Rad3 related (ATR) replication stress signaling pathway and that this depended upon the presence of R-loops [24,29]. However, we had not identified any downstream effectors that might be key to processing R-loops and maintaining genome stability in these tumors. By examining the gene correlations along PHATE_1 in Ewing sarcoma cells (Table S2), we discovered several genes strongly correlated with an EWSR1-FLI1-high cell state which are relevant to R-loop resolution and replication stress: 1) the Fanconi Anemia pathway (FANCI, FANCD2, and FANCA), and 2) Flap endonuclease 1 (FEN1).

Fanconi Anemia genes play a critical role in resolving both interstrand crosslinks and R-loops, supporting replication fork integrity and preserving genome stability [30]. Compared to IMR90, a fibroblast and PHATE_1-high cell type, we observed dramatic increase in the post-translationally modified (activated/ mono-ubiquitinated) forms of FANCD2 and FANCI protein in Ewing sarcoma cells (black triangles, Figure 7A) and a marked increase in cell death with knockdown of multiple Fanconi anemia genes (Figure 7B). These results suggest a dependence on this pathway for survival in Ewing sarcoma but not in IMR90 cells.

FEN1 is an essential enzyme that removes 5′ overhangs during DNA repair and replication [31]. FEN1 has also been implicated in processing R-loops at telomeres to limit telomere fragility [31,32]. This led us to hypothesize FEN1 could be important for resolving R-loops in Ewing sarcoma and, consequently, maintaining high proliferation rates. We experimentally validated that FEN1 was upregulated in Ewing sarcoma cell lines compared to IMR90 (Figure 7C). Furthermore, the chemical inhibition of FEN1 resulted in severe toxicity in Ewing sarcoma, but not in IMR90 (Figure 7D). More importantly, unlike in IMR90, FEN1 inhibition led to further accumulation of R-loops in Ewing sarcoma cells (Figure S9).

Finally, it was asserted that PHATE_1-low tissues, but not PHATE_1-high tissues, would demonstrate a dependency upon the factors which resolve R-loops and replication stress (the Fanconi Anemia and flap endonuclease genes). We demonstrated above that PHATE_1-high cell types do not accumulate high R-loop levels (Figures 6 and S9). We also demonstrated that IMR90, a PHATE_1-high tissue type, does not express high levels of the Fanconi Anemia and flap endonuclease genes (Figure 7A,C) or depend upon them for survival (Figure 7B,D). Therefore, it remained only to determine whether PHATE_1-low tissues demonstrate a dependency upon these factors similarly to Ewing sarcoma. We calculated normalized read counts of Fanconi Anemia and flap endonuclease genes for

each sample. The results were grouped by tissue type, and tissues were ordered by median expression of all four genes (Figure 8). As expected, we found that PHATE_1-low tissue types (neural crest cells, neural progenitor/stem cells, iPSCs, and hESCs) showed higher expression of Fanconi Anemia and flap endonuclease genes compared to all other tissues except for Ewing sarcoma and HSCs.

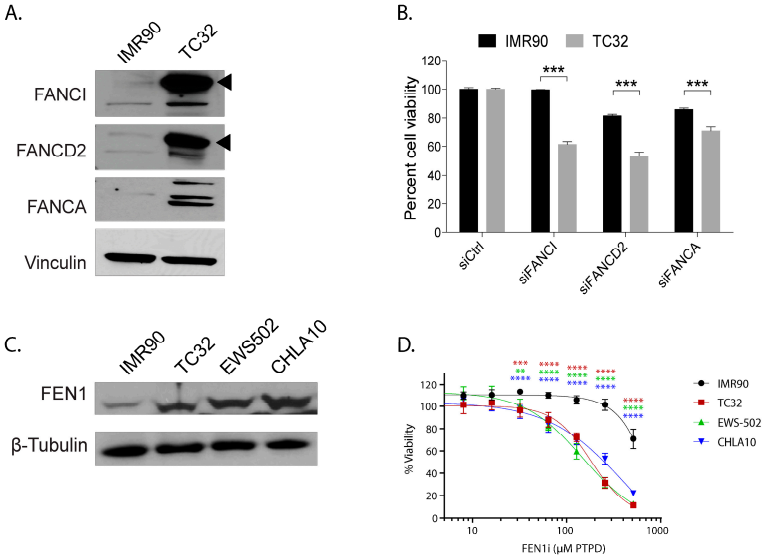

Figure 7. The response to replication stress is a defining feature for Ewing sarcoma compared to the PHATE_1-high developmental context. (**A**) Western blot showing expression of Fanconi Anemia proteins FANCA, FANCD2, and FANCI in TC32 (Ewing sarcoma cell line) compared to IMR90 (fibroblast cell line); (**B**) Bar-plot showing the impact of FANC genes knockdown on cell viability in TC32 compared to IMR90; (**C**) Western blot showing the expression of FEN1 in Ewing sarcoma cell lines TC32, EWS502, and CHLA10 compared to IMR90; (**D**) Line plot showing the viability of Ewing sarcoma cells compared to IMR90 with increasing doses of FEN1 inhibitor. (** $p \leq 0.01$; *** $p \leq 0.001$; **** $p \leq 0.0001$).

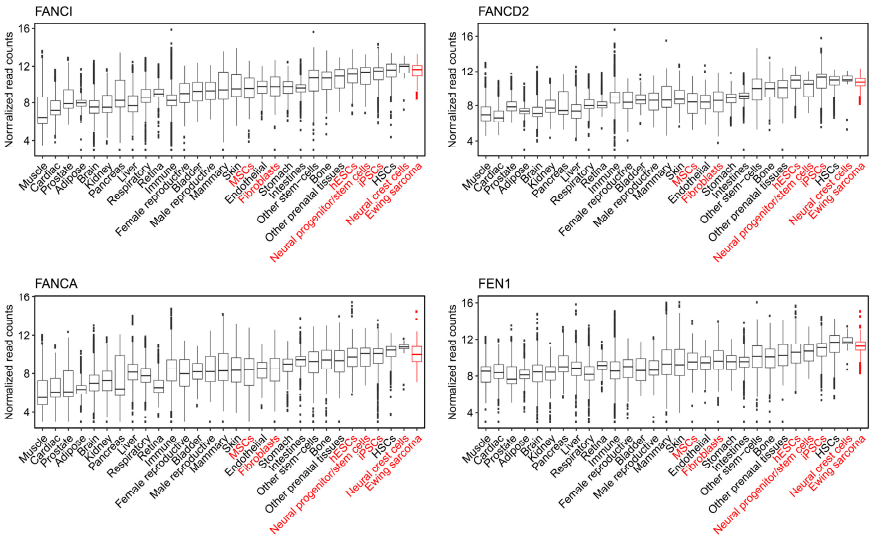

Figure 8. Normalized expression of Fanconi Anemia and flap endonuclease genes across normal tissue types and Ewing sarcoma.

Taken together, these results support the assertion that R-loop and replication stress resolving factors are probably necessary for stable EWSR1-FLI1 expression in primary tissue types.

3. Discussion

Elucidating the transcriptomic context in which Ewing sarcoma arises involves both the question of which normal cell states are permissive for stable fusion gene expression and of how tumor cells hijack developmental pathways to maintain proliferation, resist treatment, and metastasize. To address these questions, we undertook an unbiased mass-scale re-analysis of publicly available RNA-Sequencing data sets including 40,643 samples from a wide variety of normal tissue types and 260 Ewing sarcoma samples. By analyzing the marker genes which distinguish Ewing sarcoma from other samples, we identified normal tissues with Ewing-like gene expression patterns (Figure 1C). Interestingly, the tissues which showed Ewing-like transcriptomes belong to developmental transitions such as gastrulation (Figure 2B; Video S1). This led us to hypothesize that EWSR1-FLI1 reactivates developmental gene programs in normal tissue types as part of Ewing sarcomagenesis.

To further elucidate the transcriptomic landscape within the population of Ewing sarcoma and Ewing-like tissues, we implemented the PHATE data diffusion algorithm [12]. While UMAP offers a global and local perspective on the similarities and differences between various samples in a data set, PHATE is capable of accurately reconstructing the transition states between samples based on their gene expression profiles [10,12]. We demonstrated that PHATE can successfully recapitulate the known biological transitions within the bulk transcriptomes of Ewing sarcoma-like tissues (Figures 2 and S4; Video S1). This approach was particularly effective in revealing the transition states between pluripotent, neuroectodermal, and mesodermal lineages, revealing the location of Ewing sarcoma within this developmental context (Figure 2B; Video S1). Though Ewing sarcoma is believed to arise from mesenchymal origins, it also displays markers of pluripotency and neuroectodermal lineages [2]. This dichotomy was recapitulated by PHATE as it portrayed Ewing sarcoma as stretched between the mesodermal branch and the juncture between the pluripotent and neuroectodermal branches (Figure 2B; Video S1).

Multiple studies have demonstrated that ectopic EWSR1-FLI1 expression leads to the expression of neuroectodermal markers in MSCs and that EWSR1-FLI1 depletion in Ewing sarcoma cells leads to increased mesenchymal gene expression patterns [13,14]. Consistent with these findings, our analysis showed that samples with lower EWSR1-FLI1 expression had higher PHATE_1 positions (more mesodermal), and samples with higher EWSR1-FLI1 expression had lower PHATE_1 positions (more pluripotent/neuroectodermal) (Figures 3 and S5).

Several transcriptomic studies have defined Ewing sarcoma-specific gene sets, many of which have been formalized in the Molecular Signatures Database (MSigDB) Chemical and Genetic Perturbations (C2 : CGP) collection [17]. These gene sets are typically generated through transcriptomic experiments in which a non-Ewing cell is engineered to express EWSR1-FLI1 or a Ewing cell is subjected to knock-down of EWSR1-FLI1 and the change in gene expression is determined. As expected, these gene sets are highly enriched in the gene expression correlations of Ewing sarcoma samples along PHATE_1 (Figure 3D). However, the same is true for the underlying developmental context. Even when no Ewing sarcoma samples are considered, the gene correlations within the developmental context along PHATE_1 are also highly enriched for "Ewing sarcoma gene sets" (Figure S7D,F). These results indicate the degree to which EWSR1-FLI1 hijacks normal developmental trajectories and also indicates that many of the classical Ewing sarcoma gene sets may be, in fact, markers of cell lineage rather than bona fide markers of Ewing sarcoma.

Previous reports demonstrated that inhibition of LSD1 interferes with the EWSR1-FLI1 transcriptomic program [18]. Recapitulating this observation, we demonstrated SP2509 treatment acts similarly to EWSR1-FLI1 knockdown by shifting multiple Ewing sarcoma cell lines significantly higher on PHATE_1 (Figure S6B,C). Furthermore, previous studies have described an antagonistic relationship between TEAD1 and EWSR1-FLI1 in which EWSR1-FLI1 prevents the transcription of TEAD1 target

genes, preventing ECM sensing [19]. We demonstrate that TEAD1 siRNA shifts Ewing sarcoma cells significantly lower on PHATE_1, toward the pluripotent/neuroectodermal branch (Figure S6A). This finding implies that the antagonism between EWSR1-FLI1 and TEAD1 is bi-directional and that Ewing sarcoma cells can be induced to undergo further transformation in basal conditions through TEAD1 knock down.

Recent studies have demonstrated that EWSR1-FLI1 activity plays a central role in defining Ewing tumor heterogeneity [14,20]. Combining single cell transcriptomes from cell line and PDX samples, we identified a subpopulation which displayed the highest PHATE_1 score (mesodermal). As expected, our cluster marker analysis confirmed that this subpopulation is enriched for mesodermal lineage markers and markers of low EWSR1-FLI1 expression (Figure 4E). Interestingly, this analysis also revealed the degree to which this subpopulation is defined by markers of invasiveness and metastasis. The top positive markers in these cells are S100 Calcium Binding Protein A10 (*S100A10*), secreted Protein Acidic And Cysteine Rich (*SPARC*), and lamin A/C (*LMNA*), all three of which have been firmly linked with metastatic progression in multiple cancers [33–35]. Furthermore, pathway enrichment of the markers in this cluster reveals evidence of epithelial to mesenchymal transition and low-EWSR1-FLI1 expression alongside increased expression of multiple metastasis-related pathways (e.g., "WU_CELL_MIGRATION", $p < 5.94 \times 10^{-9}$), implicating these cells as a likely metastatic subpopulation (Figure 4E; Table S3). Recent literature has indicated that EWSR1-FLI1 levels fluctuate within tumors and that this heterogeneity controls the ability of cells to proliferate or metastasize [20]. Our findings reveal that cell proliferation and metastasis are not unique to Ewing sarcoma but are capabilities derived from its hijacking of the transcriptional programs found within its developmental context.

Though many consider bone-marrow MSCs to be the probable cell of origin for Ewing sarcoma, neither they, nor fibroblasts (another PHATE_1-high tissue), are capable of stable EWSR1-FLI1 expression [4,6,23,36]. Conversely, PHATE_1-low tissues iPSCs, hESCs (only tested with p53 knock-down), and neural crest cells each tolerate stable EWSR1-FLI1 expression and, in the case of hESCs, show dependency upon the fusion gene [6,21–23]. Given the transcriptional similarity between PHATE_1-low tissue types and EWSR1-FLI1-high Ewing sarcoma samples, we hypothesized that these primary tissues might possess qualities that make them innately permissive for EWSR1-FLI1 expression. Subsequent analysis revealed that PHATE_1-low tissues share a high degree of similarity with the EWSR1-FLI1 transcriptome, largely pertaining to proliferation, DNA repair, and RNA processing pathways (Figure 5).

Previous reports demonstrated that both Ewing sarcoma and pluripotent tissues display high levels of proliferation and transcription [24–26]. A known consequence of transcriptional activity is the accumulation of R-loops, three-stranded genomic structures resulting from the hybridization of RNA to DNA [37]. The failure to resolve R-loops leads to replication stress and genomic instability, a condition exacerbated by high replication rates [37]. Ewing sarcoma displays a high abundance of R-loops [24], but it maintains a relatively stable genome [3]. Therefore, we hypothesized that Ewing sarcoma relies upon factors which resolve R-loops and mitigate replication stress. Given the permissivity for EWSR1-FLI1 in PHATE_1-low tissues, we further hypothesized that they would display high R-loop levels and share Ewing sarcoma's predicted dependency upon R-loop and replication-stress-mitigating factors.

It was experimentally revealed that the Fanconi Anemia and flap endonuclease genes, which have roles in resolving R-loops and mitigating replication stress, are necessary for cell viability in Ewing sarcoma (Figure 7). Furthermore, reanalysis of public DRIP sequencing data revealed that PHATE_1-low tissue types display significantly higher R-loop accumulation compared to PHATE_1-high tissues (Figure 6) and higher expression of Fanconi Anemia and flap endonuclease genes compared to all other tissue types except for Ewing sarcoma and HSCs (Figure 8). Importantly, these experiments also demonstrated that PHATE_1-high tissues which do not permit stable EWSR1-FLI1 expression have (1) lower levels of R-loops compared to PHATE_1-low tissues and Ewing sarcoma (Figures 6 and S8) and (2) low dependence on Fanconi Anemia and flap endonuclease genes (Figures 7 and 8). Taken together,

these results indicate that the factors which resolve R-loop accumulation and replication stress, such as the Fanconi Anemia and flap endonuclease proteins, are likely prerequisites for stable EWSR1-FLI1 expression in primary tissues.

There exists a natural antagonism between replication and transcription [27,37]. Replication forks can stall, reverse, or collapse due to collisions with transcriptional machinery [38]. R-loops form naturally during transcription but, if unresolved, contribute to transcription-replication conflicts [39]. Without the ability to mitigate this antagonism, cells cannot maintain hyperproliferation and hypertranscription without suffering genome instability [27]. We demonstrated herein that Ewing sarcoma and PHATE_1-low cell types display the ability to resolve R-loops and replication stress. Consequently, we propose that this mitigating capability is required to maintain the high rates of proliferation and transcription which these tissues generally display (Ewing sarcoma and pluripotent stem cells show both hypertranscription and hyperproliferation) [24–27]. However, we cannot exclude the possibility that R-loop levels, R-loop resolving factors and factors involved in responding to replication stress are independent correlates of altered proliferative state. It remains now for subsequent experiments to determine whether these mitigating factors can (1) improve the permissibility of other normal tissues for modeling Ewing sarcomagenesis and (2) guide the search for which cell types and cell states Ewing sarcoma could arise from. Furthermore, according to the depmap database, *FEN1* is among the top dependencies for Ewing sarcoma [40]. Taken together with our findings, this supports the concept that Ewing sarcoma cells rely upon factors that mitigate replication-transcription conflict for survival. Consequently, we propose that inhibitors which target these factors may represent robust novel therapeutics for the treatment of Ewing sarcoma patients.

4. Materials and Methods

4.1. Ewing Sarcoma Cell Lines

Ewing cells TC32, CHLA10, CHLA9, and TC71 were purchased from Children's Oncology Group (COG) and the EWS502 Ewing sarcoma cell line was a kind gift from Dr. Stephen Lessnick (Nationwide Children's Hospital, Columbus, OH, USA). IMR90, a primary fibroblast cell line was used as control and purchased from ATCC. TC32, TC71, and EWS502 were grown in RPMI supplemented with 10% Fetal Bovine Serum (Atlanta Biologicals, Flowery Branch, GA, USA), CHLA10 and CHLA9 in IMDM supplemented with 20% fetal bovine serum and IMR90 in DMEM supplemented with 10% fetal bovine serum. Cells were maintained in 37 °C in a humidified atmosphere with 5% CO_2.

4.2. Single Cell RNA-Sequencing (scRNA-Seq) of Ewing Sarcoma Cell Lines

To obtained a single cell suspension, 70–80% confluent cells were washed twice with HBSS (Corning, Corning, NY, USA), recollected using 0.25% Trypsin, 2.21 mM EDTA, 1X sodium bicarbonate solution (Corning) and centrifuged for 5 minutes at 300 rcf at 4 °C. Cell pellets were then resuspended in cold DPBS, without calcium and without magnesium (Corning) supplemented with 0.04% bovine serum albumin (BSA) to minimize cell sticking. Viability was measured using an automated cell counter (Nexcelom, Lawrence, MA, USA), centrifuged as described above and resuspended at the desired final concentration in cold DPBS, without calcium and without magnesium - 0.04% BSA. Finally, single cell suspensions were strain through a 70 µm cell sieve and processed for sequencing.

Library prep was performed with the 10x Genomics v3.1 3′ kit (Pleasanton, CA, USA). All samples were sequenced on the Illumina HiSeq 3000 (San Diego, CA, USA) with a 28 + 8 + 100 configuration. CHLA9 and CHLA10 libraries were re-sequenced on the Illumina NovaSeq (San Diego, CA, USA) with a 28 + 8 + 91 configuration.

4.3. Cell Viability Assay

Cells were seeded at 30% confluence in 384 well plates and treated with FEN1 inhibitor (PTPD, Glixx Laboratories Inc., Hopkinton, MA, USA) the following day. Cytotoxicity was evaluated after

72 h of treatment using Celltiter-Glo (Promega, Madison, WI, USA). For transfection experiments, siRNA against FANCI, FANCD2, FANCA and non-targeting Control were purchased from SantaCruz Biotechnology Inc. (Dallas, TX, USA). Cells were incubated with siRNA and Lipofectamine RNAiMax by reverse transfection in 96 well plates, following manufacturer's instructions. Cell viability was evaluated after 72 h of transfection using Celltiter-Glo.

4.4. Western Blot Analysis

Whole cell lysates were prepared using RIPA buffer, separated on 3–8% gradient gels using the NuPage system (Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA) and transferred onto nitrocellulose membrane. Blots were incubated with 1:1000 dilution of the antibodies overnight at 4 °C and developed using enhanced chemiluminescence (Super ECL or West Femto, Thermo Fisher Scientific). The following antibodies were used: FEN1 (sc-13051, Santa Cruz), FANCA (sc-28215, Santa Cruz, Dallas, TX, USA), FANCD2 (NB100-182, Novus, Wrentham, MA, USA), FANCI (A301-254A, Bethyl Labs, Montgomery, TX, USA), β-Tubulin (cs2128, Cell Signaling, Danvers, MA, USA), Vinculin (cs13901, Cell Signaling), goat anti-mouse IgG-HRP and goat anti-rabbit IgG-HRP (Santa Cruz Biotech Inc.). Full images of blots are available in the Supplemental Materials.

4.5. RNA:DNA Hybrid Intensity Analysis

Extent of genomic R-loops in the cells was measured using dot blots. Briefly, DNA harvested and purified by phenol-chloroform-ethanol method was digested using a cocktail of restriction enzymes (HindIII, EcoRI, BsrGI, XbaI and SspI, NEB). Each sample (0.5 μg of purified digested genomic DNA) was loaded in quadruplicate on to a pre-wet H^+ nylon membrane (Amersham Hybond, GE Healthcare Life Sciences, Chicago, IL, USA) and allowed to incubate for 20 min. Membranes were then washed twice with dH_2O, rinsed in 2× SSC buffer and then left to air-dry at room temperature. For single-stranded DNA, there was an additional denaturation step (incubation in 0.5N NaOH, 1.5M HCl for 10 min), followed by a 10 min incubation in neutralization buffer (1M NaCl, 0.5M Tris-HCl pH 7). Membranes were blocked with 5% non-fat dry milk in TBS-T and incubated with either S9.6 antibody (ENH001, 1:1000, Kerafast, Boston, MA, USA) or ssDNA antibody (MAB3034 1:5000 Millipore, Burlington, MA, USA) overnight. Blots were analyzed with enhanced chemiluminescence (Super ECL) using Image Studio (LI-COR, Lincoln, NE, USA) to measure signal intensity of each dot. The corresponding ssDNA signal was used to normalize R-loop intensity. Cells were treated with a FEN1 inhibitor (RF00974SC, Maybridge Chemicals, Loughborough, UK).

4.6. Gene Set Collections Generated and Utilized

Gene sets used within this study for gene set enrichment analysis (GSEA) and gene set enrichment (over-representation analysis) were obtained from the Molecular Signatures Database (MSigDB) v7 accessed via the *msigdbr* function of the *msigdbr* package v.7.0.1 or generated from literature sources. A custom gene set category, termed "integrated", was taken as the combination of gene sets from the "H" (Hallmark), "C2 : CGP" (chemical and genetic perturbations), "C2 : CP" (Canonical pathways; includes *BioCarta*, *KEGG*, *PID*, and *Reactome*), and "C5 : GO" (Gene Ontology) collections. Additionally, the custom collection of "Ewing sarcoma gene sets" was curated in a three-step process: (1) Five new gene sets were defined from recent literature: "Riggi 2014 activated by EWSR1-FLI1" and "Riggi 2014 repressed by EWSR1-FLI1" were defined from Riggi et al. 2014 as the genes which are bound by EWSR1-FLI1 and transcriptionally activated or repressed based on knockdown experiments in Ewing sarcoma cell lines [41]. "Aynaud 2020 EWSR1-FLI1-high markers", and "Aynaud 2020 EWSR1-FLI1-low markers" were defined from Aynaud et al. 2020 from independent component analysis of single cell RNA-Sequencing data in a Ewing sarcoma cell line with an inducible knockdown of EWSR1-FLI1, representing gene expression profiles specific to Ewing sarcoma cells with high or low expression of EWSR1-FLI1 respectively [14]. "Aynaud 2020 activated by EWSR1-FLI1" was taken as the intersection of "Aynaud 2020 EWSR1-FLI1-high markers" with the set of genes that show increasing H3K27ac

signal in the days following the cessation of full EWSR1-FLI1 knockdown [14]. (2) Pre-existing gene sets related to Ewing sarcoma were gathered from the C2 : CGP collection of MSigDB using the regex term "EWING|_EWS|EWSR1". (3) Six gene sets of the resulting collection were censored. "Rorie targets of EWSR1-FLI1 fusion down" and "Rorie targets of EWSR1-FLI1 fusion up" involved the expression of EWSR1-FLI1 in neuroblastoma cells and were, therefore, not relevant to a study with a normal tissue background [42]. "Torchia targets of EWSR1-FLI1 fusion up", "Torchia targets of EWSR1-FLI1 fusion top 20 up", "Torchia targets of EWSR1-FLI1 fusion down", and "Torchia targets of EWSR1-FLI1 fusion top 20 down" were conducted in a mouse model of leukemia, which was deemed unsuitable because of the need for a normal tissue background and the divergent genetic background of mice from humans [43]. Furthermore, the set of "EWSR1-FLI1 targets" from Figure S1D was taken as the genes from "Ewing sarcoma gene sets" belonging to gene sets which involved an EWSR1-FLI1 or FLI1 ChIP-Seq binding study. Finally, meta gene sets were calculated as part of the analysis in Figure 5 by using regular expressions relevant to the category in question to classify gene sets from the "integrated" collection. The "EWSR1-FLI1 transcriptome" was defined as the "Ewing sarcoma gene sets" which comes from studies that observe changes in gene expression after ectopic expression of EWSR1-FLI1 in non-Ewing tissues or after knock-down of EWSR1-FLI1 in Ewing sarcoma. These gene sets were found through a regex search using the term "_UP|ACTIVATED|HIGH" in the Ewing sarcoma gene sets collection.

4.7. Pre-Processing of Publicly Available Bulk RNA-Seq Data

Standardized bulk RNA-Seq data was downloaded from the ARCHS[4] data repository v8 (February 2020). Sample metadata was categorized by tissue type and tumor status with a custom regex dictionary in coordination with cell line information downloaded from Cellosaurus v33 (December 2019). Samples from single cell experiments were removed. Tumor samples that did not belong to the 'ewing sarcoma' category, didn't fit a unique tissue type classification, or which contained fewer than 10 million reads aligned to the human transcriptome were also filtered out. The result was 40,643 samples from normal tissues and 260 Ewing sarcoma (EWS) samples (Table S1). Furthermore, genes with zero counts in 10% or more of samples were discarded.

Using the vst function from DESeq2 v1.26.0 counts were geometric-mean normalized and a variance-stabilizing transform was applied, generating a homoscedastic dataset suitable for dimensionality reduction. The resulting normalized gene count matrix had the dimensions 28,621 genes X 40,903 samples. VST-transformed gene variance was calculated for every gene using the rowVars function from the matrixStats v0.55.0 package. The top 10 thousand variable genes were selected for principle component analysis (PCA) calculated with the prcomp function from the stats v3.6.2 package.

The top 100 principle components (PCs) were subsequently used to construct a nearest neighbor graph using the FindNeighbors function of the Seurat v3.1.3 package. Louvain clustering was also performed using the FindClusters function of Seurat and the umap function of the uwot v0.1.5 package was used to calculate a UMAP embedding (Table S1). Ewing sarcoma marker genes were obtained by using the FindMarkers function of Seurat using the Wilcoxon rank sum model. Only genes with an average log2 fold change >0.58 and p adjusted value <1E-50 were considered as Ewing sarcoma marker genes (Table S1). Ewing sarcoma marker scores for each sample were calculated as the median of VST-transformed counts for the Ewing sarcoma marker genes within that sample (Table S1). Additionally, the expression of Fanconi Anemia and flap endonuclease genes across tissue types was taken as the VST-transformed read counts of each gene within each sample (Table S4). Additionally, comparison of Ewing marker expression between tissue categories and Louvain clusters (Figure S2) was performed using the pairwise.t.test function of the stats R package with 'holm' multiple testing correction applied. The enrichment of tissue type within each cluster was calculated as the number of samples of that tissue type in the cluster divided by the total number of samples of that tissue type and then scaled but not centered for each column of the heatmap (Figure S3).

4.8. PHATE Analysis of Top EWS-Like Clusters

The top Louvain clusters by Ewing sarcoma marker gene expression were selected for further analysis via the PHATE embedding algorithm [12]. PHATE embeddings were calculated using the phate function of the phateR v1.0.0 package on the matrix of VST-transformed counts. PHATE was calculated for 3 dimensions at a lower nearest-neighbor setting to improve the ability to visualize and remove samples which did not participate in the EWS-containing trajectory (plots generated by accompanying scripts). A final PHATE embedding was optimized from the remaining samples by iterating the nearest neighbor parameter until the developmental context was fully visible (Table S2).

PHATE-correlated genes were calculated using the cor function of the stats R package (Table S2). For this analysis, data were "median-by-ratio" normalized using MedianNorm and GetNormalizedMat functions of the EBSeq v1.26.0 package. Gene Set Enrichment Analysis (GSEA) was calculated using the fgsea function of the fgsea v1.12.0 package by using the signed R^2 of gene correlations along PHATE_1 as the gene ranking metric (Table S2). The impact of various interventions on PHATE_1 position in Ewing sarcoma samples was calculated using a *t* test implemented via the stat_compare_means function of the ggpubr package. The *t* test was one-tailed in the case of EWS-FLI1 knock-down experiments, and two-tailed in the case of other interventions.

4.9. Processing of Ewing Sarcoma Cell Line and Publicly Available scRNA-Seq Data

For Ewing sarcoma cell line scRNA-Seq data, reads were demultiplexed using bcl2fastq (HiSeq 3000-generated reads) and the demux command of the internal 10x genomics pre-processing pipeline (NovaSeq-generated reads). Reads were subsequently standardized to the output style of bcl2fastq. Read counts were produced using the cellranger v3.1 pipeline with the count command and the built-in hg38 reference.

For each dataset, quality control cell filtering was performed based on three metrics calculated by Seurat: (1) Percent mitochondrial reads, (2) number of UMI counts, and (3) number of unique genes identified. Violin plots were used to find thresholds and statistical outliers were automatically removed (plots generated via accompanying scrips). Following quality control, read counts were log normalized and scaled using the NormalizeData and ScaleData functions of Seurat respectively. Then, principle component analysis (PCA) was calculated using the RunPCA function of Seurat. A nearest-neighbor graph and Louvain clustering were calculated using the FindNeighbors and FindClusters functions of Seurat respectively. A UMAP embedding was calculated with the RunUMAP function of Seurat. Cell cycle scoring was performed using the CellCycleScoring function of Seurat. Marker genes were calculated using the FindAllMarker function of Seurat (Table S3). Pathway enrichment of marker genes for each cluster was calculated using the enricher function of clusterProfiler with annotations taken from the "integrated" gene set collection described above (Table S3). Word clouds were produced using the wordcloud function of the wordcloud package using integer ranking of marker false-discovery rate within each cluster as the "frequency" metric.

Data set integration was performed using canonical correlation analysis (CCA) in the manner described previously by the authors of Seurat (Figure S8A,B) [44]. First, each sample was separately normalized and the top 5000 highly variable genes were determined using the NormalizeData and FindVariableFeatures functions of Seurat respectively. Then, the FindIntegrationAnchors and IntegrateData functions of Seurat were called to generate an integrated count matrix for downstream analysis via the pipeline described above. PHATE scores were calculated at the single-cell level by multiplying the scaled single-cell gene counts by the corresponding PHATE_1 signed R^2 gene correlation value obtained from bulk RNA samples. This process weighted each single cell transcriptome by the contribution of the developmental context's PHATE_1 signature. To obtain a PHATE score for each single cell, the median of all non-zero weighted gene counts was calculated. The resulting cell metadata with PCA, UMAP, and PHATE analysis results is made available here (Table S3).

4.10. Processing of ssDRIP-Seq Data from Developmental Context Cell Types

ssDRIP-Seq (strand-specific DNA : RNA immuno-precipitation sequencing) data was obtained from a recently published public repository (GSE145964, Table S4). First, raw reads were downloaded in SRA format using the prefetch command from SRA-toolkit v2.10.1. Then the parallel-fastq-dump package/command v0.6.3 was used to split the SRA files into fastq files which were subsequently trimmed and filtered using the fastp package/command v0.20.0. Then the reads were aligned to GRCh38 using the bwa mem command of the bwa package v0.7.17. Alignment files were split by first-in-pair strand, converted to BAM format, sorted, and indexed using samtools v1.9. Strand-specific broad peaks were called using macs2 v2.2.6 with the settings "—nomodel—extsize 147—broad". BAM reads were assigned to peaks to construct a peak-count matrix for the forward-strand reads and perform PCA using the DiffBind v2.14.0 package. Peaks from batch 2 were removed due to the technical effects discovered by PCA (plots generated via accompanying scripts). Remaining peaks were filtered out if they had a q value >0.001 and were subsequently compared to determine if the number of called peaks was different between cell types using a one-tailed *t* test with an alternative hypothesis of "greater" implemented via the stat_compare_means function of the ggpubr package. Finally, the peaks which passed q value filtering were overlapped to create a consensus peak set for each cell type using the findOverlapsOfPeaks function of the ChIPpeakAnno package v3.20.0. Then the consensus peak sets for iPSCs and MSCs were overlapped and plotted.

4.11. Data and Software Availability

All sequencing datasets generated in this study have been deposited in GEO (GSE146221). Publicly available data sets used in this study are listed in Table S1 (RNA-Seq) and Table S4 (DRIP-Seq). All scripts and data used to generate the bioinformatics analyses in this paper have been made publicly available on GitHub (millerh1/Ewing-sarcoma-paper-Miller-2020). Analyses were performed using the R computing language v3.6.2 [45].

5. Conclusions

Recent years have yielded powerful insights into the molecular mechanisms by which EWSR1-FLI1 controls gene expression and cellular behavior. However, it is still unclear what cellular characteristics are required for stable EWSR1-FLI1 expression in primary tissues. Consequently, successful spontaneous models of Ewing sarcoma do not yet exist. To address this gap in knowledge, we mined publicly available transcriptomic profiles of Ewing sarcoma and normal tissue types. This allowed us to identify a population of Ewing-like normal tissues belonging to developmental processes such as gastrulation. Additional analysis and experimental validation revealed stable expression of EWSR1-FLI1 likely requires factors which resolve R-loops and replication stress. Taken together, our findings suggest that Ewing sarcomagenesis likely occurs in tissues which already tolerate high proliferation and transcriptional activity because of their capacity for resolving R-loops and mitigating replication stress.

Supplementary Materials: The following are available online at http://www.mdpi.com/2072-6694/12/4/948/s1, Figure S1: Reconstruction of Ewing sarcoma developmental context from bulk transcriptomics, Figure S2: Heatmap showing comparison of Ewing marker expression levels between different tissue types and clusters, Figure S3: Heatmap showing the scaled enrichment of each tissue type within each cluster, Figure S4: PHATE embedding displaying germ lineage annotations with Ewing sarcoma samples included and recalculated with Ewing sarcoma samples removed, Figure S5: EWS-FLI1 controls the position of Ewing sarcoma within PHATE_1, Figure S6: Multiple interventions make a significant impact on position of Ewing sarcoma samples within PHATE_1, Figure S7: Ewing sarcoma cells traverse underlying developmental trajectories in response to EWS-FLI1 expression and other stimuli, Figure S8: Integrative single cell RNA sequencing analysis of Ewing sarcoma cell lines and PDXs, Figure S9: Bar-plot showing effect of FEN1 inhibitor (RF00974SC) treatment on R-loop accumulation in a Ewing sarcoma cell line (TC32) and a fibroblast cell line (IMR90), Table S1: Abbreviations, Ewing sarcoma markers, and detailed list of bulk RNA Seq samples, Table S2: PHATE analysis data (PHATE positions, gene correlations, and GSEA results), Table S3: Single cell analysis data (cell metadata with UMAP embedding, cluster markers, cluster gene set enrichment results), Table S4: R-loop and replication stress analysis data (Fanconi Anemia and flap endonuclease gene expression values and DRIP-Seq sample info), Video S1: 3D PHATE embedding, full western blot images included in "cancers-772370 - WB figures" PDF file.

Author Contributions: Conceptualization, H.E.M., A.G., B.S.I. and A.J.R.B.; Data curation, H.E.M.; Formal analysis, H.E.M.; Funding acquisition, A.G., L.A.L., B.S.I. and A.J.R.B.; Investigation, H.E.M., A.G., N.B. and L.A.L.; Methodology, H.E.M., A.G., B.S.I. and A.J.R.B.; Project administration, A.J.R.B.; Resources, A.J.R.B.; Software, H.E.M.; Supervision, A.J.R.B.; Visualization, H.E.M., A.G. and L.A.L.; Writing—original draft, H.E.M.; Writing—review & editing, A.G., L.A.L., B.S.I. and A.J.R.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by NIH, grant numbers 1R01CA152063 and 1R01CA241554 and CPRIT, grant number RP150445, to A.J.R.B.; CPRIT training grant (RP101491), NCATS/NIH Translational Science Training Scholarship (TL1TR002647), NCI postdoctoral training grant, grant (T32CA148724) and AACR-AstraZeneca START grant (18-40-12-GORT) to A.G.; CDMRP Peer Reviewed Cancer Research Program Horizon Award, grant number W81XWH-19-1-0180, and NIH, grant number TL1TR002647, to L.A.L.; NIH, grant number T32GM113896/STXMSTP, and NIH/NIA, grant number F30 AG057213, to B.S.I; NGS sequencing data were generated in the GCCRI Genome Sequencing Facility which is supported by UT Health San Antonio, NIH-NCI P30 CA054174 (Mays Cancer Center at UT Health San Antonio), NIH Shared Instrument grant 1S10OD021805-01 (S10 grant), and CPRIT Core Facility Award (RP160732).

Acknowledgments: We are grateful for the assistance from Dawn Garcia, Korri Weldon and Zhao Lai from GCCRI Genome Sequencing Facility. We thank 10x Genomics for generating the sequencing data for the Ewing sarcoma cell lines, CHLA9 and CHLA10. Finally, we would like to thank those who critically reviewed this manuscript: Kevin Kanda, Pramiti Mukhopadhyay, Adam Kosti, and Stephanie Fedorchak.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Delattre, O.; Zucman, J.; Plougastel, B.; Desmaze, C.; Melot, T.; Peter, M.; Kovar, H.; Joubert, I.; De Jong, P.; Rouleau, G.; et al. Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours. *Nature* **1992**, *359*, 162–165. [CrossRef] [PubMed]
2. Ferreira, B.I.; Alonso, J.; Carrillo, J.; Acquadro, F.; Largo, C.; Suela, J.; Teixeira, M.R.; Cerveira, N.; Molares, A.; Goméz-López, G.; et al. Array CGH and gene-expression profiling reveals distinct genomic instability patterns associated with DNA repair and cell-cycle checkpoint pathways in Ewing's sarcoma. *Oncogene* **2008**, *27*, 2084–2090. [CrossRef] [PubMed]
3. Brohl, A.S.; Solomon, D.A.; Chang, W.; Wang, J.; Song, Y.; Sindiri, S.; Patidar, R.; Hurd, L.; Chen, L.; Shern, J.F.; et al. The Genomic Landscape of the Ewing Sarcoma Family of Tumors Reveals Recurrent STAG2 Mutation. *PLoS Genet.* **2014**, *10*, e1004475. [CrossRef]
4. Lessnick, S.L.; Dacwag, C.S.; Golub, T.R. The Ewing's sarcoma oncoprotein EWS/FLI induces a p53-dependent growth arrest in primary human fibroblasts. *Cancer Cell* **2002**, *1*, 393–401. [CrossRef]
5. Sohn, E.J.; Li, H.; Reidy, K.; Beers, L.F.; Christensen, B.L.; Lee, S.B. EWS/FLI1 oncogene activates caspase 3 transcription and triggers apoptosis in vivo. *Cancer Res.* **2010**, *70*, 1154–1163. [CrossRef] [PubMed]
6. Torres-Ruiz, R.; Martinez-Lage, M.; Martin, M.C.; Garcia, A.; Bueno, C.; Castaño, J.; Ramirez, J.C.; Menendez, P.; Cigudosa, J.C.; Rodriguez-Perales, S. Efficient Recreation of t(11;22) EWSR1-FLI1+ in Human Stem Cells Using CRISPR/Cas9. *Stem Cell Reports* **2017**, *8*, 1408–1420. [CrossRef] [PubMed]
7. Minas, T.Z.; Surdez, D.; Javaheri, T.; Tanaka, M.; Howarth, M.; Kang, H.J.; Han, J.; Han, Z.Y.; Sax, B.; Kream, B.E.; et al. Combined experience of six independent laboratories attempting to create an Ewing sarcoma mouse model. *Oncotarget* **2017**, *8*, 34141–34163. [CrossRef]
8. Lachmann, A.; Torre, D.; Keenan, A.B.; Jagodnik, K.M.; Lee, H.J.; Wang, L.; Silverstein, M.C.; Ma'ayan, A. Massive mining of publicly available RNA-seq data from human and mouse. *Nat. Commun.* **2018**, *9*, 1366. [CrossRef]
9. Moon, K.R.; Stanley, J.S.; Burkhardt, D.; van Dijk, D.v.; Wolf, G.; Krishnaswamy, S. Manifold learning-based methods for analyzing single-cell RNA-sequencing data. *Curr. Opin. Syst. Biol.* **2018**, *7*, 36–46. [CrossRef]
10. McInnes, L.; Healy, J.; Saul, N.; Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw.* **2018**, *3*, 861. [CrossRef]
11. Blondel, V.D.; Guillaume, J.L.; Lambiotte, R.; Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, *2008*, P10008. [CrossRef]

12. Moon, K.R.; van Dijk, D.; Wang, Z.; Gigante, S.; Burkhardt, D.B.; Chen, W.S.; Yim, K.; van den Elzen, A.; Hirn, M.J.; Coifman, R.R.; et al. Visualizing structure and transitions in high-dimensional biological data. *Nat. Biotechnol.* **2019**, *37*, 1482–1492. [CrossRef] [PubMed]
13. Riggi, N.; Suvà, M.L.; Suvà, D.; Cironi, L.; Provero, P.; Tercier, S.; Joseph, J.M.; Stehle, J.C.; Baumer, K.; Kindler, V.; et al. EWS-FLI-1 expression triggers a ewing's sarcoma initiation program in primary human mesenchymal stem cells. *Cancer Res.* **2008**, *68*, 2176–2185. [CrossRef] [PubMed]
14. Aynaud, M.M.; Mirabeau, O.; Gruel, N.; Grossetête, S.; Boeva, V.; Durand, S.; Surdez, D.; Saulnier, O.; Zaïdi, S.; Gribkova, S.; et al. Transcriptional Programs Define Intratumoral Heterogeneity of Ewing Sarcoma at Single-Cell Resolution. *Cell Rep.* **2020**, *30*, 1767–1779. [CrossRef] [PubMed]
15. Howarth, M.M.; Simpson, D.; Ngok, S.P.; Nieves, B.; Chen, R.; Siprashvili, Z.; Vaka, D.; Breese, M.R.; Crompton, B.D.; Alexe, G.; et al. Long noncoding RNA EWSAT1-mediated gene repression facilitates Ewing sarcoma oncogenesis. *J. Clin. Invest.* **2014**, *124*, 5275–5290. [CrossRef]
16. Subramanian, A.; Tamayo, P.; Mootha, V.K.; Mukherjee, S.; Ebert, B.L.; Gillette, M.A.; Paulovich, A.; Pomeroy, S.L.; Golub, T.R.; Lander, E.S.; et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15545–15550. [CrossRef]
17. Liberzon, A.; Subramanian, A.; Pinchback, R.; Thorvaldsdóttir, H.; Tamayo, P.; Mesirov, J.P. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **2011**, *27*, 1739–1740. [CrossRef]
18. Sankar, S.; Theisen, E.R.; Bearss, J.; Mulvihill, T.; Hoffman, L.M.; Sorna, V.; Beckerle, M.C.; Sharma, S.; Lessnick, S.L. Reversible LSD1 inhibition interferes with global EWS/ETS transcriptional activity and impedes Ewing sarcoma tumor growth. *Clin. Cancer Res.* **2014**, *20*, 4584–4597. [CrossRef]
19. Katschnig, A.M.; Kauer, M.O.; Schwentner, R.; Tomazou, E.M.; Mutz, C.N.; Linder, M.; Sibilia, M.; Alonso, J.; Aryee, D.N.T.; Kovar, H. EWS-FLI1 perturbs MRTFB/YAP-1/TEAD target gene regulation inhibiting cytoskeletal autoregulatory feedback in Ewing sarcoma. *Oncogene* **2017**, *36*, 5995–6005. [CrossRef]
20. Franzetti, G.A.; Laud-Duval, K.; Van Der Ent, W.; Brisac, A.; Irondelle, M.; Aubert, S.; Dirksen, U.; Bouvier, C.; De Pinieux, G.; Snaar-Jagalska, E.; et al. Cell-to-cell heterogeneity of EWSR1-FLI1 activity determines proliferation/migration choices in Ewing sarcoma cells. *Oncogene* **2017**, *36*, 3505–3514. [CrossRef]
21. von Levetzow, C.; Jiang, X.; Gwye, Y.; von Levetzow, G.; Hung, L.; Cooper, A.; Hsu, J.H.R.; Lawlor, E.R. Modeling initiation of ewing sarcoma in human neural crest cells. *PLoS ONE* **2011**, *6*, e19305. [CrossRef] [PubMed]
22. Gordon, D.J.; Motwani, M.; Pellman, D. Modeling the initiation of Ewing sarcoma tumorigenesis in differentiating human embryonic stem cells. *Oncogene* **2016**, *35*, 3092–30102. [CrossRef] [PubMed]
23. Kovar, H.; Amatruda, J.; Brunet, E.; Burdach, S.; Cidre-Aranaz, F.; De Alava, E.; Dirksen, U.; Van Der Ent, W.; Grohar, P.; Grünewald, T.G.P.; et al. The second European interdisciplinary Ewing sarcoma research summit—A joint effort to deconstructing the multiple layers of a complex disease. *Oncotarget* **2016**, *7*, 8613–8624. [CrossRef] [PubMed]
24. Gorthi, A.; Romero, J.C.; Loranc, E.; Cao, L.; Lawrence, L.A.; Goodale, E.; Iniguez, A.B.; Bernard, X.; Masamsetti, V.P.; Roston, S.; et al. EWS-FLI1 increases transcription to cause R-Loops and block BRCA1 repair in Ewing sarcoma. *Nature* **2018**, *555*, 387–391. [CrossRef]
25. Percharde, M.; Bulut-Karslioglu, A.; Ramalho-Santos, M. Hypertranscription in Development, Stem Cells, and Regeneration. *Dev. Cell* **2017**, *40*, 9–21. [CrossRef]
26. Hu, X.; Eastman, A.E.; Guo, S. Cell cycle dynamics in the reprogramming of cellular identity. *FEBS Lett.* **2019**, *593*, 2840–2852. [CrossRef]
27. Babos, K.N.; Galloway, K.E.; Kisler, K.; Zitting, M.; Li, Y.; Shi, Y.; Quintino, B.; Chow, R.H.; Zlokovic, B.V.; Ichida, J.K. Mitigating Antagonism between Transcription and Proliferation Allows Near-Deterministic Cellular Reprogramming. *Cell Stem Cell* **2019**, *25*, 486–500. [CrossRef]
28. Zatreanu, D.; Han, Z.; Mitter, R.; Tumini, E.; Williams, H.; Gregersen, L.; Dirac-Svejstrup, A.B.; Roma, S.; Stewart, A.; Aguilera, A.; et al. Elongation Factor TFIIS Prevents Transcription Stress and R-Loop Accumulation to Maintain Genome Stability. *Mol. Cell* **2019**, *76*, 57–69. [CrossRef]
29. Nieto-Soler, M.; Morgado-Palacin, I.; Lafarga, V.; Lecona, E.; Murga, M.; Callen, E.; Azorin, D.; Alonso, J.; Lopez-Contreras, A.J.; Nussenzweig, A.; et al. Efficacy of ATR inhibitors as single agents in Ewing sarcoma. *Oncotarget* **2016**, *7*, 58759–58767. [CrossRef]

30. García-Rubio, M.L.; Pérez-Calero, C.; Barroso, S.I.; Tumini, E.; Herrera-Moyano, E.; Rosado, I.V.; Aguilera, A. The Fanconi Anemia Pathway Protects Genome Integrity from R-loops. *PLoS Genet.* **2015**, *11*, e1005674. [CrossRef]
31. Teasley, D.C.; Parajuli, S.; Nguyen, M.; Moore, H.R.; Alspach, E.; Lock, Y.J.; Honaker, Y.; Saharia, A.; Piwnica-Worms, H.; Stewart, S.A. Flap endonuclease 1 limits telomere fragility on the leading strand. *J. Biol. Chem.* **2015**, *290*, 15133–15145. [CrossRef] [PubMed]
32. Kazak, L.; Reyes, A.; He, J.; Wood, S.R.; Brea-Calvo, G.; Holen, T.T.; Holt, I.J. A Cryptic Targeting Signal Creates a Mitochondrial FEN1 Isoform with Tailed R-Loop Binding Properties. *PLoS ONE* **2013**, *8*, e62340. [CrossRef] [PubMed]
33. Tantyo, N.A.; Karyadi, A.S.; Rasman, S.Z.; Salim, M.R.G.; Devina, A.; Sumarpo, A. The prognostic value of S100A10 expression in cancer. *Oncol. Lett.* **2019**, *17*, 1417–1424. [CrossRef] [PubMed]
34. Arnold, S.A.; Brekken, R.A. SPARC: A matricellular regulator of tumorigenesis. *J. Cell Commun. Signal.* **2009**, *3*, 255–273. [CrossRef]
35. González-Cruz, R.D.; Dahl, K.N.; Darling, E.M. The emerging role of Lamin C as an important LMNA isoform in mechanophenotype. *Front. Cell Dev. Biol.* **2018**, *6*, 151. [CrossRef]
36. Riggi, N.; Suvà, M.L.; De Vito, C.; Provero, P.; Stehle, J.C.; Baumer, K.; Cironi, L.; Janiszewska, M.; Petricevic, T.; Suvà, D.; et al. EWS-FLI-1 modulates miRNA145 and SOX2 expression to initiate mesenchymal stem cell reprogramming toward Ewing sarcoma cancer stem cells. *Genes Dev.* **2010**, *24*, 916–932. [CrossRef]
37. Crossley, M.P.; Bocek, M.; Cimprich, K.A. R-Loops as Cellular Regulators and Genomic Threats. *Mol. Cell* **2019**, *73*, 398–411. [CrossRef]
38. Gadaleta, M.C.; Noguchi, E. Regulation of DNA replication through natural impediments in the eukaryotic genome. *Genes (Basel).* **2017**, *8*, 98. [CrossRef]
39. Parajuli, S.; Teasley, D.C.; Murali, B.; Jackson, J.; Vindigni, A.; Stewart, S.A. Human ribonuclease H1 resolves R-loops and thereby enables progression of the DNA replication fork. *J. Biol. Chem.* **2017**, *292*, 15216–15224. [CrossRef]
40. DepMap. The Cancer Dependency Map Consortium. Available online: https://depmap.org/portal/depmap/ (accessed on 8 April 2020).
41. Riggi, N.; Knoechel, B.; Gillespie, S.M.; Rheinbay, E.; Boulay, G.; Suvà, M.L.; Rossetti, N.E.; Boonseng, W.E.; Oksuz, O.; Cook, E.B.; et al. EWS-FLI1Utilizes Divergent Chromatin Remodeling Mechanisms to Directly Activate or Repress Enhancer Elements in Ewing Sarcoma. *Cancer Cell* **2014**, *26*, 668–681. [CrossRef]
42. Rorie, C.J.; Thomas, V.D.; Chen, P.; Pierce, H.H.; O'Bryan, J.P.; Weissman, B.E. The Ews/Fli-1 Fusion Gene Switches the Differentiation Program of Neuroblastomas to Ewing Sarcoma/Peripheral Primitive Neuroectodermal Tumors. *Cancer Res.* **2004**, *64*, 1266–1277. [CrossRef] [PubMed]
43. Torchia, E.C.; Boyd, K.; Rehg, J.E.; Qu, C.; Baker, S.J. EWS/FLI-1 Induces Rapid Onset of Myeloid/Erythroid Leukemia in Mice. *Mol. Cell. Biol.* **2007**, *27*, 7918–7934. [CrossRef] [PubMed]
44. Stuart, T.; Butler, A.; Hoffman, P.; Hafemeister, C.; Papalexi, E.; Mauck, W.M.; Hao, Y.; Stoeckius, M.; Smibert, P.; Satija, R. Comprehensive Integration of Single-Cell Data. *Cell* **2019**, *177*, 1888–1902. [CrossRef] [PubMed]
45. R, 3.6.2; R: A Language and Environment for Statistical Computing; R Core Team. Available online: https://www.r-project.org/ (accessed on 8 April 2020).