

## MIT Open Access Articles

*STAG2 loss reshapes oncogenic enhancer-promoter looping in Ewing sarcoma*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

**Citation:** Adane, Biniam et al. "STAG2 loss rewires oncogenic and developmental programs to promote metastasis in Ewing sarcoma." 39, 6 (June 2016): P827-844.e10. © 2021 Elsevier Inc

**As Published:** <http://dx.doi.org/10.1016/j.ccell.2021.05.007>

**Publisher:** Elsevier BV

**Persistent URL:** <https://hdl.handle.net/1721.1/131197>

**Version:** Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

**Terms of use:** Creative Commons Attribution-NonCommercial-NoDerivs License



## **STAG2 loss reshapes oncogenic enhancer-promoter looping in Ewing sarcoma**

Brian Crompton<sup>1,2\*</sup>, Gabriela Alexe<sup>1,2,3\*</sup>, Elizabeth Hwang<sup>1</sup>, Denes Hnisz<sup>4</sup>, Caleb Lareau<sup>2,5,6</sup>, Abraham Weintraub<sup>4,7</sup>, Sarah Wang<sup>1</sup>, Amanda Balboni Iniguez<sup>1</sup>, Neekesh Dharia<sup>1,2</sup>, Diana Lu<sup>1</sup>, Benjamin Tanenbaum<sup>2</sup>, Michael Burger<sup>2</sup>, Francisca Vazquez<sup>2</sup>, Monica Schenone<sup>2</sup>, Steven Carr<sup>2</sup>, Martin Aryee<sup>2,6</sup>, Richard Young<sup>4^</sup>, Kimberly Stegmaier<sup>1,2^</sup>

<sup>1</sup>Dana-Farber/Boston Children's Cancer and Blood Disorders Center, Boston, MA, USA

<sup>2</sup>Broad Institute, Cambridge, MA, USA

<sup>3</sup>Bioinformatics Graduate Program, Boston University, Boston, MA, USA

<sup>4</sup>Whitehead Institute, Cambridge, MA, USA

<sup>5</sup>Department of Biostatistics, T.H. Chan School of Public Health, Harvard University, Boston, MA, USA

<sup>6</sup>Department of Pathology, Massachusetts General Hospital, Charlestown, MA, USA

<sup>7</sup>Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA.

\*Contributed Equally

^Jointly directed this work

Word Count: 5740

Figures: 7

Supplemental Figures: 6

Supplemental Tables: 10

Supplemental Material

### **Corresponding Authors:**

Kimberly Stegmaier

Dana-Farber Cancer Institute

450 Brookline Ave., D630

Boston, MA 02215

Phone: 617-632-4438

E-mail: kimberly\_stegmaier@dfci.harvard.edu

Richard Young

Whitehead Institute for Biomedical Research

9 Cambridge Center

Cambridge MA 02142

Phone: 617.258.5218

Email: young@wi.mit.edu

## **ABSTRACT**

Mutations in members of the cohesin complex, including *STAG2*, are associated with hematologic and solid malignancies but their role remains unclear. *STAG2* mutations are present in 15-20% of Ewing sarcoma tumors. We chose this disease as a model system for studying the effects of *STAG2* loss due to its simple genome. We found that in Ewing sarcoma cells with *STAG2* knockout, the cohesin complex is intact, and alternately incorporates *STAG1*, which becomes essential. Ewing sarcoma is characterized by rearrangements between *EWSR1* and ETS-family transcription factor genes, most commonly *FLI1*. *STAG2* knockout is paradoxically associated with a decrease in EWS/FLI target gene expression, resulting in an EWS/FLI “low” state described to promote metastasis. Loss of *STAG2* also decreased cohesin binding at EWS/FLI sites and altered DNA-DNA looping contacts between EWS/FLI enhancers and gene promoters. These studies demonstrate that *STAG2* mutations can alter transcriptional programs of oncogenic fusions, such as EWS/FLI.

## **SIGNIFICANCE**

*STAG2* mutations occur in multiple cancers but their function has been elusive. In Ewing sarcoma, we found that loss of *STAG2* weakens chromatin interactions between EWS/FLI-bound enhancers and target genes. The resulting modified transcriptional program is associated with increased metastatic potential demonstrating that *STAG2* mutations can promote cancer cell plasticity.

## INTRODUCTION

Massively parallel sequencing efforts have revealed cancer-associated mutations in the cohesin complex, a multiprotein complex important in sister chromatid cohesion and gene regulation (1-11). Progress has been made in understanding the mechanistic role of cohesin mutations in the myeloid malignancies, whereas the role of *STAG2* mutations in solid tumors has remained more elusive (12,13). *STAG2* is a member of the cohesin complex, composed of SMC1a, SMC3, and RAD21, forming a ringed structure that can surround two strands of DNA. The fourth member of the complex is one of three members of the STAG protein family: STAG1, STAG2, or STAG3 (14,15). Studies have demonstrated that cohesin complexes containing either STAG1 or STAG2 are expressed concurrently in mitotic cells and bind to overlapping chromatin locations (16,17). The cohesin complex plays an important role in regulating sister chromatid alignment during cell division. Therefore, when loss of function mutations in *STAG2* were first identified in cancer, it was hypothesized that these events would cause cohesin dysfunction and improper chromosomal migration resulting in aneuploidy (1). However, subsequent studies demonstrated that *STAG2* mutations are not significantly associated with aneuploidy in hematopoietic malignancies, Ewing sarcoma, or bladder carcinoma (1,2,4,7,8).

The cohesin complex also plays a critical role in chromatin regulation of gene expression. Cohesin maintains chromatin accessibility at transcription factor binding sites during cell division and promote DNA-DNA contacts that form the bases of enhancer-promoter interactions and define the boundaries of topologically associated domains at CTCF

binding sites (18-20). Therefore, we hypothesized that the core ring of the cohesin complex remains intact in the absence of STAG2, but loss of STAG2 alters cohesion function resulting in changes in gene regulation. To test this hypothesis, we chose the pediatric solid tumor Ewing sarcoma as our model system, a malignancy defined by a simple genome and an oncogenic rearrangement between the *EWSR1* gene and an ETS-family transcription factor encoding gene, most commonly *FLI1* (6-8). *STAG2* mutations are present in 15-20% of tumors and lead to loss of expression of the gene (6-8). Studies have demonstrated that patients with *STAG2*-mutated Ewing sarcoma have a higher rate of metastatic disease and worse outcomes (7,8). In this study, we downregulated and knocked out *STAG2* in Ewing sarcoma cell lines expressing wild-type *STAG2* and examined the transcriptional and epigenetic effects of *STAG2* loss.

## **RESULTS**

### **STAG1 is Necessary for Cell Viability in Ewing Sarcoma Cells with STAG2 Loss**

To determine the effect of loss of STAG proteins on the composition of the cohesin complex, *STAG1* and *STAG2* were knocked out in the A673 Ewing sarcoma cell line by CRISPR-Cas9 gene editing. Co-immunoprecipitation of SMC1a demonstrated that SMC1a remains in complex with SMC3 and RAD21 with knockout of either *STAG* gene. *STAG1* incorporation into the complex increases with *STAG2* deletion whereas *STAG2* incorporation remained relatively stable with *STAG1* deletion (Supplementary Fig. 1A). These findings were confirmed in isogenic A673 cells clonally selected for either expression of wild-type *STAG2* or *STAG2* knockout (Fig. 1A and Supplementary Fig.

S1B-C). This data suggests that STAG2 is preferentially bound to cohesin when both STAG proteins are expressed. To test this hypothesis, we subjected protein lysates obtained from cells expressing STAG2 and cells with *STAG2* knockout to urea at increasing concentrations to examine the relative binding efficiency of STAG proteins to the cohesin complex. We found that STAG2 has a higher binding affinity for cohesin than STAG1, explaining why cohesin preferentially binds STAG2 when both STAG proteins are expressed (Fig. 1B).

Given that *STAG2* knockout does not disrupt the integrity of the cohesin complex and that loss-of-function mutations in *STAG2* are found in an aggressive subset of Ewing sarcomas, we did not expect STAG2 loss to adversely affect cell viability. Indeed, *STAG2* knockout did not affect viability in Ewing cells in short-term cultures (Supplementary Fig. S1D). Next, we hypothesized that cells would not tolerate the loss of both STAG2 and STAG1. In Ewing sarcoma cell lines transduced with a genome-scale CRISPR screening library, *STAG1* was the top-scoring dependency of *STAG2* null Ewing sarcoma cell lines when compared to both Ewing and non-Ewing cancer cell lines with wild-type *STAG2* (Fig. 1C and Supplementary Fig. S2A). We also found that *STAG1* was the top-scoring dependency when comparing the A673 cells clonally selected for *STAG2* knockout compared to A673 cells expressing wild-type *STAG2* (Fig. 1D). These findings were validated by treating cell lines with individual *STAG1*-targeting CRISPR guides (Fig 1E-F and Supplementary Fig. S2B-C) This data confirms recent reports that Ewing sarcoma cells are dependent on the presence of one or more STAG proteins for cell viability and

that STAG1 may be an appealing therapeutic target for tumors with loss of STAG2 expression (21,22).

### **STAG2 Loss Alters Expression of EWS/FLI-regulated Genes**

Our previous studies, and that of others, suggested that STAG2 loss may be associated with metastatic disease and poor outcome in Ewing sarcoma, and we observed that *STAG2* null cells demonstrate transcriptional changes enriched for signatures of migration and metastasis (7,8). To further define the effects of STAG2 loss on gene expression in Ewing sarcoma, we performed transcriptome profiling of A673 and TC71 Ewing cells clonally selected for *STAG2* knockout (Supplementary Fig. S3A-D). Gene expression changes induced by STAG2 loss in the A673 cell line resulted in corresponding changes in protein expression and a similar pattern of gene expression changes in the TC71 cell line (Fig. 2A). We confirmed that *STAG2* knockout induced gene expression changes enriched for signatures of metastasis (Fig 2B-C). Unbiased gene-set enrichment analysis (GSEA) revealed that these expression changes were highly associated with programs driven by EWS/FLI activity (Fig. 2D-F). Strikingly, genes that are upregulated by EWS/FLI were highly correlated with genes downregulated by *STAG2* loss in cell lines and in human tumors (Fig. 2G-I, Supplementary Fig. S3E-H, and Supplemental Tables S1-3).

### **STAG2 Loss Does Not Prevent the Establishment of EWS/FLI-bound Enhancers**

In Ewing sarcoma, EWS/FLI upregulates the transcription of target genes by establishing novel EWS/FLI-bound enhancer sites (23). Downregulation of EWS/FLI expression or

prevention of EWS/FLI chromatin binding would be expected to interfere with EWS/FLI-regulated transcription. However, we found that EWS/FLI expression is not suppressed by *STAG2* knockout suggesting an alternative mechanism must account for this transcriptional reprogramming (Fig. 3A-C). To determine whether *STAG2* knockout alters EWS/FLI binding to chromatin, we performed chromatin immunoprecipitation with sequencing (ChIP-Seq). EWS/FLI binding sites were largely maintained but with a small decrement in genome-wide binding intensity (Fig. 3D). EWS/FLI sites with a detectable decrease in binding were significantly associated with changes in the expression of nearby genes, but the log<sub>2</sub> fold-change in median gene expression was only -0.0064 (Fig 3E). EWS/FLI sites with a detectable increase in binding were not significantly associated with changes in gene expression (Fig 3E). Therefore, this decrement in EWS/FLI binding is insufficient to account for the transcriptional alterations identified by RNA-Seq in Ewing cells with *STAG2* knockout.

EWS/FLI was recently found to act as a pioneer factor in Ewing sarcoma, generating novel enhancer regions at chromatin binding sites (23). To determine whether *STAG2* knockout had an effect on the ability of EWS/FLI to establish novel enhancer regions at chromatin binding sites, we performed ChIP-Seq for H3K27Ac, a histone mark associated with enhancer activity (Supplementary Fig. S4A) (24). Consistent with prior reports, we found that H3K27Ac occupancy was highest at EWS/FLI binding sites containing GGAA repeats where EWS/FLI acts as an enhancer (Supplementary Fig. S4B). However, *STAG2* knockout did not affect enhancer markings at EWS/FLI binding sites (Fig. 3F)

(23). This data demonstrates that STAG2 loss has a minimal effect on the establishment of EWS/FLI-bound, Ewing-specific enhancer regions.

### **STAG2 Loss Impairs Cohesin Binding**

Although it has been established that cohesin is frequently recruited to enhancer regions to maintain chromatin accessibility and facilitate enhancer promoter interactions (20), it was previously unknown if cohesin binds to EWS/FLI enhancer sites. To examine this, we performed ChIP-Seq for SMC1a and found that over 22% (1041 of 4709) of EWS/FLI-bound sites overlap with SMC1a-bound sites ( $p = 0.001$ ; Fig. 4A). We then examined whether sites co-bound by EWS/FLI and SMC1a were present in other cell types or only in Ewing sarcoma cells. In fact, only 5.6% of these genomic locations (58 of 1041) were found to be bound by cohesin in other cancer and non-cancer cell types, demonstrating that these EWS/FLI-associated cohesin binding sites are unique to Ewing sarcoma (Fig. 4A). Furthermore, SMC1a binding intensity and chromatin accessibility, as measured by ATAC-Seq, were higher at EWS/FLI-bound sites containing GGAA repeats compared to EWS/FLI-bound sites with a single GGAA sequence (Fig. 4B-C) (23). These data confirm that cohesin is recruited to EWS/FLI enhancers that are specific to the oncogenic process in Ewing sarcoma.

Interestingly, our SMC1a ChIP-Seq data demonstrates that SMC1a binding intensity and chromatin accessibility decrease genome-wide and at EWS/FLI binding sites in Ewing cells with *STAG2* knockout (Fig. 4D-F and Supplementary Fig. S5A-B). Therefore, we hypothesized that loss of cohesin binding is contributing to the gene expression changes

observed in STAG2 loss. If true, we would expect that the genes most dependent on the presence of STAG2 for expression would be regulated by nearby enhancers that have open chromatin and are bound by high levels of EWS/FLI, cohesin and enhancer markings. We found that genes downregulated by *STAG2* knockout were located nearest to EWS/FLI peaks with higher SMC1a, H3K27Ac, and chromatin accessibility signal when STAG2 is expressed, compared to genes that are upregulated or stable when STAG2 is lost (Fig. 5A-C). This suggests that downregulated genes are reliant on both cohesin binding and EWS/FLI enhancer activity for expression. Accordingly, we also found that the EWS/FLI binding sites that were concurrently marked by high levels of SMC1a, H3K27Ac, and chromatin accessibility signal, were located near genes downregulated with *STAG2* knockout (Fig. 5D-F).

### **Loss of STAG2 preferentially weakens DNA-DNA contact between enhancers and gene promoters**

To determine whether loss of cohesin binding was diminishing the contact between EWS/FLI-bound enhancer sites and the promoter regions of their effector genes, we next performed genome-wide chromatin conformation profiling by SMC1a HiChIP in Ewing cells with and without *STAG2* knockout (25). Previous studies have demonstrated that a large proportion of chromatin contacts occur at CTCF-bound insulator regions (18,19). We thus also performed CTCF ChIP-Seq to differentiate chromatin loops defining insulated neighborhoods from loops occurring at non-CTCF binding sites, such as those involving enhancer interactions with promoters (Supplementary Fig. S5C). We identified 22,680 high-confidence chromatin loops in cells expressing wild-type *STAG2*

(Supplemental Table S4). Of these high-confidence loops, 2,362 loops were anchored by EWS/FLI on one end and a gene promoter at the other end (Supplemental Table S5). The genes associated with EWS/FLI bound loops were highly enriched for EWS/FLI gene signatures (Supplemental Table S6), and this data represents, to our knowledge, the first genome-wide survey of direct interactions between EWS/FLI and target genes. Of note, the resolution of the HiChIP anchors is limited to 5kb, obscuring EWS/FLI-gene contacts occurring in very close proximity, such as the putative interaction between the *NROB1* promoter and a nearby EWS/FLI binding site (26). To confirm that loops anchored by EWS/FLI on at least one end were unique in Ewing sarcoma, we compared our HiChIP data to recently published SMC1a HiChIP data generated from GM12878 lymphoblastoid cells (25). We identified 2,591 loops in A673 that were significantly less prevalent or undetected in GM12878, including 721 loops (27.8%) anchored by at least one EWS-FLI peak. Conversely, of the 5,387 loops stronger in GM12878, only 125 (2.3%) were bound by EWS-FLI in A673 cells. Together, these results support a model that the binding of EWS/FLI to chromatin creates novel enhancer-promoter interactions in Ewing sarcoma.

To further characterize the effects of the *STAG2* mutation in Ewing sarcoma, we performed differential looping analyses of HiChIP data between the *STAG2* knockout and wild-type clones. We identified 1061 high-confidence chromatin loops with a significant decrease in the number of measured contacts in *STAG2* KO cells and 1104 loops with a significant increase in the number of contacts (Supplemental Fig. S6 and Supplemental Table S7). Altered loops were significantly enriched for loops anchored by EWS/FLI binding sites on one or more edges ( $P < 0.0001$ ; Fig. 6A) but not for insulated regions

anchored on both edges by CTCF ( $P = 0.19$ ). This finding is consistent with a recent report that enhancer-promoter chromatin interactions are more dependent on STAG2-bound cohesin, while CTCF-CTCF contacts are dependent on the presence of either STAG1- or STAG2-bound cohesin (27). Furthermore, loops with decreased contacts in *STAG2* knockout cells had a larger median length between loop edges than loops with increased contacts (Fig. 6B). This was also true when restricting the analysis only to loops anchored by EWS/FLI, but not for loops anchored by CTCF (Fig. 6C-D). Finally, we found that the genes located at the anchor sites of altered loops were significantly enriched for genes downregulated (but not upregulated) in the *STAG2* knockout condition (Fig. 6E). Furthermore, when this analysis was restricted to genes located at anchor sites of altered loops that were also anchored on the other edge by an EWS/FLI peak, there was significant enrichment for genes downregulated by *STAG2* loss (Fig. 6F). These genes were also enriched in signatures of genes upregulated by EWS/FLI (Supplemental Table S8). Therefore, we propose a model by which *STAG2* loss perturbs EWS/FLI enhancer-promoter interactions, weakening long-range interactions, leading to a decrease in EWS/FLI-directed gene expression.

Our model and the data generated herein provide a future opportunity to study novel genes contributing to the aggressive phenotype of *STAG2*-mutated Ewing sarcoma, including genes that have been overlooked in studies lacking chromatin looping data to confirm EWS/FLI enhancer-promoter interactions. For example, *PHLDA3* (pleckstrin homology like domain family A member 3), a known tumor suppressor gene in esophageal and pancreatic cancer (28,29), is downregulated in signatures of metastasis

(30,31), but was not previously associated with Ewing sarcoma. Two recently published gene expression data sets suggests that loss of EWS/FLI expression leads to downregulation of *PHLDA3* although it fails to meet the standard cut-off for identifying EWS/FLI regulated genes when using expression data exclusively (Fig. 7A) (23,32). Similarly, while *SIRP $\alpha$* 's cancer-associated role in macrophages has been studied (33), the role of its expression in Ewing sarcoma cells remains unknown even though its expression is altered in published gene sets profiling models of EWS/FLI modulation (Fig. 7A) (23,26,32,34,35). Our HiChIP data confirms the presence of chromatin contacts between EWS/FLI sites and the promoters of *PHLDA3* and *SIRP $\alpha$*  genes, and we demonstrate a decrease in EWS/FLI promoter-enhancer interactions at these genes and downregulation of their expression with *STAG2* KO (Fig 7B-E). These analyses suggest that the incorporation of chromatin interaction data is necessary for our ability to precisely define the direct gene regulatory activity of EWS/FLI.

## **DISCUSSION**

The discovery of mutations in multiple members of the cohesin complex, as well as other chromatin/transcriptional regulators, was one of the major findings in massively parallel sequencing efforts in cancer. Furthermore, chromatin remodeling has emerged as a recurrent mechanism of drug resistance in cancer (36,37). A complete knowledge of how epigenetic regulators interact with each other and with transcription factors, however, is lacking, slowing progress in understanding how to manipulate these molecules to favor cancer cell death. A more complete picture of normal and pathologic epigenetic processes will expedite our ability to effectively target transcriptionally driven cancers.

One component of the epigenetic machinery is the cohesin complex. Emerging data demonstrates that cohesin plays a significant role in transcriptional regulation by securing DNA-DNA contacts between enhancers and promoters (20). By broadly surveying chromatin markings coupled with chromatin interactions in Ewing sarcoma cells, we have now generated a genome-wide connectome in this disease, building on previous work demonstrating that EWS/FLI directly regulates gene expression through enhancer-promoter interactions (23). This chromatin interactome data can be used to verify genes that are directly regulated by EWS/FLI and to identify novel direct targets of EWS/FLI, such as *PHLDA3*, which was previously missed using EWS/FLI ChIP-seq data alone, likely due to the imperfections in connecting transcription factor binding sites with the targets that they directly regulate. As technologies to map promoter-enhancer interactions continue to improve, including resolving interactions at less than 5kb, so will our ability to more comprehensively and precisely map the direct effectors of oncogenic transcription factors.

Here, we demonstrate in Ewing sarcoma that STAG2 loss leads to a decrease in cohesin binding intensity across the genome and this disproportionately affects the expression of genes dependent on enhancer elements located farthest from the gene promoter regions. Genes regulated by the oncogenic EWS/FLI fusion protein are among the most enriched for alterations in gene expression. While one obvious prediction would be that loss of STAG2 reinforces the EWS/FLI oncogenic program, surprisingly, *STAG2* deletion represses a subset of EWS/FLI regulated genes. In this way, STAG2 loss appears to

downregulate a subset of genes directly upregulated by EWS/FLI binding, effectively dampening the activity of the EWS/FLI and promoting an EWS/FLI “low” state. Recent data demonstrates that the titration of EWS/FLI transcriptional activity promotes cellular plasticity, which may be necessary for Ewing sarcoma cells facing changes in environmental pressures, such as cells undergoing metastasis (38). Indeed, work by others suggests that high levels of EWS/FLI may in fact repress metastatic potential in Ewing sarcoma tumors and that an EWS/FLI “low” state may enable metastases (38). In line with these published studies and our current results, we previously showed that *STAG2* mutations were enriched in patients with metastatic tumors and that tumors with no *STAG2* protein were enriched for gene signatures of metastasis based on RNA sequencing. Now, we have also validated that cells isogenic for loss of *STAG2* are also enriched for gene signatures of metastasis in Ewing sarcoma cells and for an EWS/FLI “low” state. Taken together, our data supports a new model whereby a second oncogenic event attenuates the activity of the initiating oncogene to minimize effects that are unfavorable to the cancer cell. In this case, EWS/FLI is thought to promote the initiation and proliferation of Ewing sarcoma tumors but impair their metastasis. Thus, one oncogenic effect of *STAG2* mutations may be to mitigate the subset of EWS/FLI genes that prevents metastasis. Further investigation will be necessary to validate the specific genes responsible for this activity.

Importantly, it is likely that the role of *STAG2* loss in regulating oncogenic gene expression is more complex than currently appreciated. In our data, for example, EWS/FLI target genes were not the only genes with altered expression. We noted that numerous

chromatin loops not associated with EWS/FLI were also lost and gained, presumably contributing to other aspects of the transcriptional reprogramming observed. It is likely that some of these altered loops play a role in regulating the activity of other transcription factors in Ewing sarcoma cells, and the alteration of these interactions may further contribute to cellular plasticity favoring an oncogenic state. Moreover, because the cohesin complex plays several important roles in the cell, other functions not explored in this work, such as its role in DNA damage and repair (39), may be altered in the context of *STAG2* mutations, a topic also worthy of exploration in future studies.

Several other hematologic and solid tumor malignancies have been associated with mutations in the cohesin complex, including *STAG2*. Some of these cancers are driven by oncogenic transcription factors, such as AML1-ETO in AML (10,11). One might speculate that mutations in cohesin complex members could alter the oncogenic effects of other transcriptional oncoproteins with which they are associated. Therefore, our data may represent a more general cancer-associated mechanism for modulating oncogenic transcription.

Finally, while *STAG2* loss appears to promote this more aggressive phenotype, it also creates a new dependency on *STAG1* expression in these cells, a vulnerability that could be exploited therapeutically in Ewing sarcoma and other cancers with *STAG2* mutations. In both of our genome-wide CRISPR-Cas9 screens, *STAG1* was unequivocally the most striking synthetic lethal dependency in the context of *STAG2* loss. While *STAG1* does not possess enzymatic activity, making a drug discovery effort challenging, new degradation

approaches to target repression may provide a tractable strategy for “drugging” this target (40).

## **METHODS**

Cell culture, cell viability, western immunoblotting, and co-immunoprecipitation of proteins were all performed using published standard techniques. Details of each assay are documented in the Supplemental Material. All cell lines used in this study were previously genotyped and confirmed to express the appropriate EWS-rearrangement using either a combination of whole-exome sequencing and transcriptome sequencing or a combination of STR profiling and RT-PCR (7).

### *CRISPR Cas9 Gene Editing*

The CRISPR plasmids pSpCas9(BB)-2A-GFP (PX458) and lentiCRISPR v2 were acquired from Addgene. Guide sequences against *STAG1* and *STAG2* were designed using the CRISPR design tool (crispr.mit.edu) and cloned into both vectors according to previously published protocols (41,42). Control sgRNAs were chosen from previously validated non-targeting sequences and a LacZ-targeting sequence (43,44). Target sequences for each guide are available in (Supplemental Table S9).

Ewing sarcoma cells were treated with CRISPR Cas9 gene editing plasmids and targeting guides by both transfection and viral transduction. A673 cells were transfected with either control guides or *STAG2*-targeting CRISPR guides packaged with PX458 constructs (10 ug) using X-tremeGENE HP. For CRISPR transduction, lentivirus was produced by

transfection of 293T cells using X-tremeGENE HP with lentiCRISPR v2 plasmid cloned to contain the targeting guides of interest with the packaging plasmids (pCMV8.9 and pCMV-VSVG). Ewing sarcoma cells were transduced with viral supernatant and 8 ug/mL of polybrene (Sigma-Aldrich) for 2 h, after which fresh media was added and the cells incubated for 48 h before antibiotic selection. In all cases, genetic knockout of target genes was confirmed by western immunoblotting. Genome-wide CRISPR-Cas9 screening data has been previously published and a summary of those methods and analytical approaches are outlined in the Supplemental Material.

#### *Clonal selection of cells with CRISPR Cas9 knockout of target genes*

Cells treated with PX458 constructs were selected 72 h after cells were transfected and flow sorted for GFP expression. Sorted cells were returned to medium and allowed to return to a normal growth rate. Cells treated with lentiCRISPR v2 virus were selected for 48 h with puromycin. In both cases, selected cells were then seeded in ClonaCell-TCS Medium (Stemcell Technologies) such that individual colonies were grown from single cells. Numerous individual clones were extracted by pipette and expanded in 24-well cell culture plates. Once expanded, the clones were tested for gene knockout by western blot.

#### *Urea gradient*

The strength of protein-protein interactions was tested by the addition of urea across a range of concentrations. After cells were lysed using the cohesin-specific protocol outlined above, urea was added to lysates to achieve the target urea concentration and

incubated at room temperature for 15 minutes. The reaction was then quenched by the addition of lysis buffer. Co-immunoprecipitation and western immunoblotting was then performed as described.

### *Transcriptome and chromatin profiling*

We performed genome-wide expression profiling (RNA-Seq), proteomic profiling by mass spectrometry, chromatin immunoprecipitation with sequencing (ChIP-Seq), assay for transposase-accessible chromatin using sequencing (ATAC-Seq), and HiChIP using published techniques with minimal modifications. Details of the methods used are detailed in the Supplemental Methods. Analysis of the data generated for each profiling technique are detailed below.

### *GSEA analysis of transcriptome and proteomic data*

GSEA v3.0 software (45) was used to identify functional associations of the molecular phenotypes induced by *STAG2* KO vs. *STAG2* wild-type with the collection c2 of 4,738 curated pathways and experimental gene sets and the collection c5 of 4,436 Gene Ontology biological processes signatures available from the MSigDB v6.0 database (45,46). The molecular phenotypes induced by *STAG2* KO vs. *STAG2* wild-type were measured by (i) RNA-Seq expression of the A673 and TC71 cells, (ii) proteome mass spectrometry of A673 cells, and (iii) RNA-Seq expression of Ewing sarcoma tumors with loss of function mutations in *STAG2* compared to tumors expressing wild type *STAG2*.(7) For each of these datasets, the hg19 genes were ranked based on the expression fold change in *STAG2* KO vs. *STAG2* wild type phenotypes. The goal of GSEA was to identify

the MSigDB gene sets that are distributed at the top or at the bottom of the ranked list of genes based on the Kolmogorov-Smirnov enrichment test. Gene sets with less than 15 genes or more than 1500 genes were excluded from the analysis. Gene sets with a nominal  $P \leq 0.05$  and an FDR  $\leq 0.25$  for the Kolmogorov-Smirnov test were considered significant hits. The results were visualized on volcano plots for the normalized enrichment score (NES) vs.  $-\log_{10}(P)$  and on GSEA plots.

#### *ChIP-Seq data processing and peak calling*

The fragment length for all the reads ranged between 190 to 250 base pairs. Quality control tests for unmapped sequences were performed based on the FastQC v.0.11.5 software (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). All of the ChIP-Seq data sets were aligned to the GRCh37/hg19 human genes using bwa v0.7.17 (<https://github.com/lh3/bwa>) with the mem -M options. PCR duplicates were removed with the Picard v2.18.2 MarkDuplicates tool (47). For each mark the reads mapped on the *STAG2* wild-type clones A673.sgNT-1c4 and A673.sgNT-2c3 were merged and labeled as “*STAG2* WT” and similarly, the reads mapped on the *STAG2* knockout clones A673.sgSTAG2-1c6, A673.sgSTAG2-4c5 were merged and labeled as “*STAG2* KO”.

The mapped reads for individual and merged clones were normalized in units of Reads Per Kilobase per Million (RPKM or rpm/bp) and coverage tracks for the RPKM signal were created as bigwig files for bins of size 20 base pairs by using the *bamCoverage* tool available in deepTools v2.5.3. (48). Area under Curve (AUC) RPKM normalized signal

across genomic regions (e.g., peaks or extended peaks) was computed with the bwtool summary software (49).

Peak calling was performed using the model-based MACS v1.4.3 software, with the cut-offs  $P \leq 1e-09$  for H3K27Ac and  $P \leq 1e-05$  for all other marks (50). The peaks were identified against input control except for SMC1a. The ENCODE black-listed regions for hg19 (available at <https://www.encodeproject.org/annotations/ENCSR636HFF/>) were removed from each set of peak regions.

Quality control tests for the mapped reads were performed by using the CHIPQC library available from Bioconductor v3.5 (51). The distances between replicates for *STAG2* WT and *STAG2* KO clones were estimated based on the irreproducible discovery rate (IDR) scores (available from the R v3.4.1 package `idr`) and visualized on correlation heatmaps and PCA plots. The EWS/FLI peaks with low binding signal ( $AUC < 50$ ) were disregarded for further analyses and so 841 out of 5,550 peaks on *STAG2* WT clones and 514 out of 3,101 peaks on *STAG2* KO clones were filtered out.

The number of reads (millions of base pairs) and the fraction of reads in peaks (FRiP) for the merged *STAG2* WT and *STAG2* KO clones are listed in (Supplemental Table S10). The peaks were annotated with the closest hg19 genes by using the `annotatePeaks` function available in the Homer v4.7 package (52) and the GREAT annotation platform (53). ChIP-Seq data for this study are available for download from the Gene Expression Omnibus (GEO) repository (GSE 116495) upon manuscript publication.

### *ATAC-Seq data processing and peak calling*

ATAC-Seq data were collected using paired-end 50 bp reads from HiSeq, Illumina at the Center for Cancer Genome Discovery at the Dana-Farber Cancer Institute (2015). Quality control tests for unmapped reads were performed based on the FastQC v.0.11.5 software (Babraham Bioinformatics, <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>).

All of the ATAC-Seq data sets were aligned to the GRCh37/hg19 human genes using bwa v0.7.17 (<https://github.com/lh3/bwa>) with the mem -M options (47). The mitochondrial reads (chr M) were removed with samtools after alignment. PCR duplicates were removed with Picard v2.18.2 MarkDuplicates tool.

The reads mapped on the *STAG2* wild-type clones sgNT-1c4 and sgNT-2c3 were merged and labeled “*STAG2* WT”, and similarly, the reads mapped on the *STAG2* knockout clones, sgSTAG2-1c6, and sgSTAG2-4c5 were merged and labeled “*STAG2* KO”. The mapped reads were normalized in units of Reads Per Kilobase per Million (RPKM or rpm/bp) and coverage tracks for the RPKM signal were created as bigwig files for bins of size 20 base pairs by using the *bamCoverage* tool available in deepTools v2.5.3 (48). Model-based peaks were identified using MACS v1.4.3(50) with the cut-off 1e-05 for the *P*-value. The peaks were annotated with the closest hg19 genes by using the *annotatePeaks* function available in the Homer v4.7 package (52). The number of reads (million of base pairs) and the fraction of reads in peaks (FRiP) scores for the merged *STAG2* WT and *STAG2* KO clones after chr M and duplicate reads removal are listed in (Supplemental Table S10). The ATAC-Seq data for this study is available for download

from the Gene Expression Omnibus (GEO) repository (GSE 116495) upon manuscript publication.

#### *ChIP-Seq and ATAC-Seq visualization and analysis*

ChIP-Seq binding peaks and normalized binding signal were visualized on the Integrative Genomic Viewer (IGV) v2.4.0 platform.<sup>(54)</sup> Gene promoter regions were defined as the +/- 3 kb intervals around the hg19 gene transcription start site (TSS). Enhancer regions were required to exist outside of the promoter regions defined as +/- 3 kb from a TSS. Heatmaps of AUC ChIP-Seq normalized signal occupancy on genomic regions were created using the *computeMatrix* and the *plotHeatmap* tools available in the deepTools v2.5.3 suite. The *plotProfile* tool from deepTools v2.5.3 was used to create metaplots based on the average normalized scores across genomic regions. The BEDTools v2.27<sup>(55)</sup> function *intersect* was used to screen for overlap between sets of genomic regions. The BEDTools function *merge* was used to merge sets of genomic regions, by combining overlapping and “book-ended” features within a short distance (e.g, 1-3 base pairs) into a single interval which spans all of the combined features.

Union lists of *STAG2* WT and *STAG2* KO binding regions were created separately for each mark by merging the lists of peaks for the *STAG2* WT clones with the list of peaks for the *STAG2* KO clones for the mark. The merging operation was performed with *bedtools merge* by combining into a single interval any *STAG2* WT and *STAG2* KO peaks within  $\leq 1$  base pair distance. The EWS/FLI binding sites were extended with 500 base pairs either side prior to computing AUC signal occupancy for marks other than EWS/FLI.

Binding regions for EWS/FLI, SMC1a, H3K27Ac and ATAC-Seq were visualized in heatmaps and analyzed for genome-wide “delta” difference scores in genome-wide AUC signal in *STAG2* KO vs. *STAG2* WT. The genome-wide *STAG2* KO vs. *STAG2* WT changes in signal occupancy were classified as “high”, “low” or “unchanged” based on the absolute cut-off 1.5 for the delta z-scores.

To determine whether some cohesin binding sites are unique in Ewing sarcoma cells, SMC1a ChIP-Seq data was compared to publicly available data of cohesin binding sites on chromatin. SMC3 ChIP-Seq data from A549, GM12878, HeLa-S3, and SK-N-SH cell lines, differentiated neural cells, and 16 week primary fetal cells were downloaded from the ENCODE repository (<https://www.encodeproject.org/>). RAD21 peak data from A549, HeLa-S3, and SK-N-SH cell lines, H1-hESC stem cells, differentiated neural cells, liver cells, and 16 week primary fetal cells were also downloaded from the ENCODE repository. A consensus “non-Ewing” list of cohesin binding sites was defined by intersecting these publicly available SMC3 and RAD21 data sets.

### *HiChIP Analysis*

HiChIP raw reads were aligned to hg19 human reference genome using HiC-Pro (56). Each of the four replicate samples was sequenced to a depth > 150M reads. Moreover, each sample passed stringent quality control with a minimum of 23% of all reads mapping to intrachromosomal loci. High-confidence loop calls were inferred using hichipper using a union of SMC1 ChIP-Seq peaks from both the WT and *STAG2* KO cell line were combined with HiChIP-specific peaks, which we have previously shown to be an

efficacious strategy to account for HiChIP-intrinsic biases (57). This consensus loop anchor set was composed of 103,632 SMC1 ChIP-Seq peaks from the *STAG2* WT, 44,966 SMC1 ChIP-Seq peaks from the *STAG2* KO, and 20,363 peaks from the combined self-ligation reads across all four HiChIP replicates. In total, 129,453 unique genomic loci were considered to be possible loop anchors. Long range interactions spanning two anchor regions, termed DNA loops, were derived from linked paired-end reads that overlapped restriction fragments containing this consensus peak of possible loop anchors. In total, 1,125,468 interactions spanning pairs of 79,850 genomic loci were observed among the four samples.

As a majority of these interactions represent background proximity ligation, we performed stringent filtering to identify putative biologically functional DNA loops. We called a set of 22,680 loops naive to any additional genetic or epigenetic annotation that met more stringent criteria. These loops contained at least four reads with paired-end tags (PETs) in two or more replicates and were statistically-significant at a distance-dependent FDR of 1%. This set of more stringently filtered loops served as a basis for differential loop calling between the wild-type and *STAG2* KO mutant cell lines. Loop edges were annotated for their overlap with ChIP-Seq binding sites of EWS/FLI and CTCF as well as gene promoter regions. The EWS/FLI binding sites used for this annotation was generated by analysis of all EWS/FLI ChIP-Seq data from A673 *STAG2* WT and KO cells combined with EWS/FLI binding sites identified from a published ChIP-Seq data set of unmanipulated A673 cells (GSM 1517562) (23).

We extracted a list of loops connecting unique transcription start sites to a distal EWS-FLI peaks (Supplemental Table S5), a resource for understanding gene regulation of the EWS-FLI fusion gene in Ewing Sarcoma. Each of these loops were supported by at least four PETs in two or more replicates and were statistically significant using a distance-dependent bias correction at a False Discovery Rate (FDR) < 1% (58). Moreover, each of these loop anchors directly overlapped a transcription start site at one end and an EWS-FLI peak at the other end. To determine if loops specific to Ewing sarcoma but absent from other cell types may be facilitated through the binding of the EWS-FLI fusion protein, we examined whether loops identified to have edges co-bound by EWS/FLI were present in a non-Ewing cell line, GM12878 (SRR3467175-SRR3467178). Loops that were differentially present in A673 vs. GM12878, differential loops were called using the *diffloop* package (59) restricting for loops with an  $\text{abs}(\log\text{FC}) > 2$  and an *diffloop* FDR < 1%.

To identify loops that were lost or gained in A673 with *STAG2* KO compared to *STAG2* WT cells, differential loops were called at using the *diffloop* (59) package (Supplementary Fig. 6). The median loop length of differential loops bound by EWS-FLI and CTCF were also identified and examined for differences in length based on whether loops were weaker or stronger in the *STAG2* KO condition by Wilcoxon rank sum test. To determine whether differential loop edges were more likely to be bound by EWS/FLI, we performed a Fisher's Exact Test.

**Acknowledgments:** This work was funded by the NCI R01 CA204915 (KS), Curing Kids' Cancer (KS), St. Baldrick's Foundation (BDC) and the NCI K08 CA188073 (BDC).

## References

1. Solomon DA, Kim T, Diaz-Martinez LA, Fair J, Elkahloun AG, Harris BT, *et al.* Mutational inactivation of STAG2 causes aneuploidy in human cancer. *Science* **2011**;333(6045):1039-43 doi 10.1126/science.1203619.
2. Balbas-Martinez C, Sagrera A, Carrillo-de-Santa-Pau E, Earl J, Marquez M, Vazquez M, *et al.* Recurrent inactivation of STAG2 in bladder cancer is not associated with aneuploidy. *Nat Genet* **2013**;45(12):1464-9 doi 10.1038/ng.2799.
3. Cancer Genome Atlas Research N, Ley TJ, Miller C, Ding L, Raphael BJ, Mungall AJ, *et al.* Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* **2013**;368(22):2059-74 doi 10.1056/NEJMoa1301689.
4. Kon A, Shih LY, Minamino M, Sanada M, Shiraishi Y, Nagata Y, *et al.* Recurrent mutations in multiple components of the cohesin complex in myeloid neoplasms. *Nat Genet* **2013**;45(10):1232-7 doi 10.1038/ng.2731.
5. Solomon DA, Kim JS, Bondaruk J, Shariat SF, Wang ZF, Elkahloun AG, *et al.* Frequent truncating mutations of STAG2 in bladder cancer. *Nat Genet* **2013**;45(12):1428-30 doi 10.1038/ng.2800.
6. Brohl AS, Solomon DA, Chang W, Wang J, Song Y, Sindiri S, *et al.* The genomic landscape of the Ewing Sarcoma family of tumors reveals recurrent STAG2 mutation. *PLoS Genet* **2014**;10(7):e1004475 doi 10.1371/journal.pgen.1004475.
7. Crompton BD, Stewart C, Taylor-Weiner A, Alexe G, Kurek KC, Calicchio ML, *et al.* The genomic landscape of pediatric Ewing sarcoma. *Cancer discovery* **2014**;4(11):1326-41 doi 10.1158/2159-8290.CD-13-1037.

8. Tirode F, Surdez D, Ma X, Parker M, Le Deley MC, Bahrami A, *et al.* Genomic landscape of Ewing sarcoma defines an aggressive subtype with co-association of STAG2 and TP53 mutations. *Cancer discovery* **2014**;4(11):1342-53 doi 10.1158/2159-8290.CD-14-0622.
9. Leiserson MD, Vandin F, Wu HT, Dobson JR, Eldridge JV, Thomas JL, *et al.* Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. *Nat Genet* **2015**;47(2):106-14 doi 10.1038/ng.3168.
10. Faber ZJ, Chen X, Gedman AL, Boggs K, Cheng J, Ma J, *et al.* The genomic landscape of core-binding factor acute myeloid leukemias. *Nat Genet* **2016**;48(12):1551-6 doi 10.1038/ng.3709.
11. Duployez N, Marceau-Renaut A, Boissel N, Petit A, Bucci M, Geffroy S, *et al.* Comprehensive mutational profiling of core binding factor acute myeloid leukemia. *Blood* **2016**;127(20):2451-9 doi 10.1182/blood-2015-12-688705.
12. Viny AD, Ott CJ, Spitzer B, Rivas M, Meydan C, Papalexi E, *et al.* Dose-dependent role of the cohesin complex in normal and malignant hematopoiesis. *J Exp Med* **2015**;212(11):1819-32 doi 10.1084/jem.20151317.
13. Mullenders J, Aranda-Orgilles B, Lhoumaud P, Keller M, Pae J, Wang K, *et al.* Cohesin loss alters adult hematopoietic stem cell homeostasis, leading to myeloproliferative neoplasms. *J Exp Med* **2015**;212(11):1833-50 doi 10.1084/jem.20151323.

14. Haering CH, Farcas AM, Arumugam P, Metson J, Nasmyth K. The cohesin ring concatenates sister DNA molecules. *Nature* **2008**;454(7202):297-301 doi 10.1038/nature07098.
15. Nasmyth K. Segregating sister genomes: the molecular biology of chromosome separation. *Science* **2002**;297(5581):559-65 doi 10.1126/science.1074757.
16. Remeseiro S, Cuadrado A, Gomez-Lopez G, Pisano DG, Losada A. A unique role of cohesin-SA1 in gene regulation and development. *EMBO J* **2012**;31(9):2090-102 doi 10.1038/emboj.2012.60.
17. Losada A, Yokochi T, Kobayashi R, Hirano T. Identification and characterization of SA/Scp3p subunits in the *Xenopus* and human cohesin complexes. *J Cell Biol* **2000**;150(3):405-16.
18. Downen JM, Fan ZP, Hnisz D, Ren G, Abraham BJ, Zhang LN, *et al.* Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* **2014**;159(2):374-87 doi 10.1016/j.cell.2014.09.030.
19. Wendt KS, Yoshida K, Itoh T, Bando M, Koch B, Schirghuber E, *et al.* Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **2008**;451(7180):796-801 doi 10.1038/nature06634.
20. Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **2010**;467(7314):430-5 doi 10.1038/nature09380.
21. Benedetti L, Cereda M, Monteverde L, Desai N, Ciccarelli FD. Synthetic lethal interaction between the tumour suppressor STAG2 and its paralog STAG1. *Oncotarget* **2017**;8(23):37619-32 doi 10.18632/oncotarget.16838.

22. van der Lelij P, Lieb S, Jude J, Wutz G, Santos CP, Falkenberg K, *et al.* Synthetic lethality between the cohesin subunits STAG1 and STAG2 in diverse cancer contexts. *Elife* **2017**;6 doi 10.7554/eLife.26980.
23. Riggi N, Knoechel B, Gillespie SM, Rheinbay E, Boulay G, Suva ML, *et al.* EWS-FLI1 utilizes divergent chromatin remodeling mechanisms to directly activate or repress enhancer elements in Ewing sarcoma. *Cancer Cell* **2014**;26(5):668-81 doi 10.1016/j.ccell.2014.10.004.
24. Creighton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* **2010**;107(50):21931-6 doi 10.1073/pnas.1016071107.
25. Mumbach MR, Rubin AJ, Flynn RA, Dai C, Khavari PA, Greenleaf WJ, *et al.* HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat Methods* **2016**;13(11):919-22 doi 10.1038/nmeth.3999.
26. Kinsey M, Smith R, Lessnick SL. NR0B1 is required for the oncogenic phenotype mediated by EWS/FLI in Ewing's sarcoma. *Mol Cancer Res* **2006**;4(11):851-9 doi 10.1158/1541-7786.MCR-06-0090.
27. Kojic A, Cuadrado A, De Koninck M, Gimenez-Llorente D, Rodriguez-Corsino M, Gomez-Lopez G, *et al.* Distinct roles of cohesin-SA1 and cohesin-SA2 in 3D chromosome organization. *Nat Struct Mol Biol* **2018**;25(6):496-504 doi 10.1038/s41594-018-0070-4.

28. Muroi H, Nakajima M, Satomura H, Takahashi M, Yamaguchi S, Sasaki K, *et al.* Low PHLDA3 expression in oesophageal squamous cell carcinomas is associated with poor prognosis. *Anticancer Res* **2015**;35(2):949-54.
29. Ohki R, Saito K, Chen Y, Kawase T, Hiraoka N, Saigawa R, *et al.* PHLDA3 is a novel tumor suppressor of pancreatic neuroendocrine tumors. *Proc Natl Acad Sci U S A* **2014**;111(23):E2404-13 doi 10.1073/pnas.1319962111.
30. Rickman DS, Millon R, De Reynies A, Thomas E, Wasyluk C, Muller D, *et al.* Prediction of future metastasis and molecular characterization of head and neck squamous-cell carcinoma based on transcriptome and genome analysis by microarrays. *Oncogene* **2008**;27(51):6607-22 doi 10.1038/onc.2008.251.
31. Bidus MA, Risinger JI, Chandramouli GV, Dainty LA, Litzi TJ, Berchuck A, *et al.* Prediction of lymph node metastasis in patients with endometrioid endometrial cancer using expression microarray. *Clin Cancer Res* **2006**;12(1):83-8 doi 10.1158/1078-0432.CCR-05-0835.
32. Marques Howarth M, Simpson D, Ngok SP, Nieves B, Chen R, Siplashvili Z, *et al.* Long noncoding RNA EWSAT1-mediated gene repression facilitates Ewing sarcoma oncogenesis. *J Clin Invest* **2014**;124(12):5275-90 doi 10.1172/JCI72124.
33. Willingham SB, Volkmer JP, Gentles AJ, Sahoo D, Dalerba P, Mitra SS, *et al.* The CD47-signal regulatory protein alpha (SIRPa) interaction is a therapeutic target for human solid tumors. *Proc Natl Acad Sci U S A* **2012**;109(17):6662-7 doi 10.1073/pnas.1121623109.
34. Hu-Lieskovan S, Zhang J, Wu L, Shimada H, Schofield DE, Triche TJ. EWS-FLI1 fusion protein up-regulates critical genes in neural crest development and is

- responsible for the observed phenotype of Ewing's family of tumors. *Cancer Res* **2005**;65(11):4633-44 doi 10.1158/0008-5472.CAN-04-2857.
35. Miyagawa Y, Okita H, Nakaijima H, Horiuchi Y, Sato B, Taguchi T, *et al.* Inducible expression of chimeric EWS/ETS proteins confers Ewing's family tumor-like phenotypes to human mesenchymal progenitor cells. *Mol Cell Biol* **2008**;28(7):2125-37 doi 10.1128/MCB.00740-07.
36. Liao BB, Sievers C, Donohue LK, Gillespie SM, Flavahan WA, Miller TE, *et al.* Adaptive Chromatin Remodeling Drives Glioblastoma Stem Cell Plasticity and Drug Tolerance. *Cell Stem Cell* **2017**;20(2):233-46 e7 doi 10.1016/j.stem.2016.11.003.
37. Iniguez AB, Alexe G, Wang EJ, Roti G, Patel S, Chen L, *et al.* Resistance to Epigenetic-Targeted Therapy Engenders Tumor Cell Vulnerabilities Associated with Enhancer Remodeling. *Cancer Cell* **2018**;34(6):922-38 e7 doi 10.1016/j.ccell.2018.11.005.
38. Franzetti GA, Laud-Duval K, van der Ent W, Brisac A, Irondelle M, Aubert S, *et al.* Cell-to-cell heterogeneity of EWSR1-FLI1 activity determines proliferation/migration choices in Ewing sarcoma cells. *Oncogene* **2017**;36(25):3505-14 doi 10.1038/onc.2016.498.
39. Meisenberg C, Pinder SI, Hopkins SR, Wooller SK, Benstead-Hume G, Pearl FMG, *et al.* Repression of Transcription at DNA Breaks Requires Cohesin throughout Interphase and Prevents Genome Instability. *Mol Cell* **2018** doi 10.1016/j.molcel.2018.11.001.

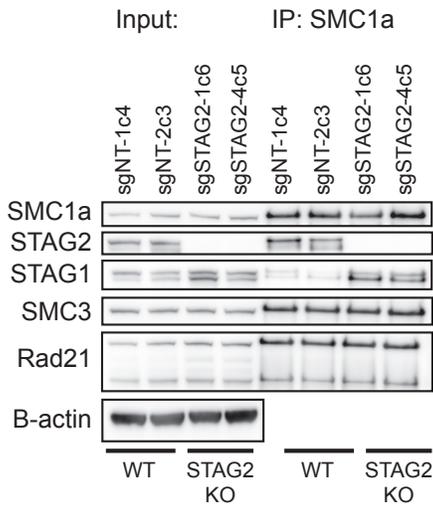
40. Lai AC, Crews CM. Induced protein degradation: an emerging drug discovery paradigm. *Nat Rev Drug Discov* **2017**;16(2):101-14 doi 10.1038/nrd.2016.211.
41. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **2013**;8(11):2281-308 doi 10.1038/nprot.2013.143.
42. Sanjana NE, Shalem O, Zhang F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat Methods* **2014**;11(8):783-4 doi 10.1038/nmeth.3047.
43. Wang T, Wei JJ, Sabatini DM, Lander ES. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **2014**;343(6166):80-4 doi 10.1126/science.1246981.
44. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **2013**;152(5):1173-83 doi 10.1016/j.cell.2013.02.022.
45. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **2005**;102(43):15545-50 doi 10.1073/pnas.0506580102.
46. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **2015**;1(6):417-25 doi 10.1016/j.cels.2015.12.004.
47. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **2010**;26(5):589-95 doi 10.1093/bioinformatics/btp698.

48. Ramirez F, Ryan DP, Gruning B, Bhardwaj V, Kilpert F, Richter AS, *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **2016**;44(W1):W160-5 doi 10.1093/nar/gkw257.
49. Pohl A, Beato M. bwtool: a tool for bigWig files. *Bioinformatics* **2014**;30(11):1618-9 doi 10.1093/bioinformatics/btu056.
50. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **2008**;9(9):R137 doi 10.1186/gb-2008-9-9-r137.
51. Carroll TS, Liang Z, Salama R, Stark R, de Santiago I. Impact of artifact removal on ChIP quality metrics in ChIP-seq and ChIP-exo data. *Front Genet* **2014**;5:75 doi 10.3389/fgene.2014.00075.
52. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **2010**;38(4):576-89 doi 10.1016/j.molcel.2010.05.004.
53. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* **2010**;28(5):495-501 doi 10.1038/nbt.1630.
54. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **2013**;14(2):178-92 doi 10.1093/bib/bbs017.
55. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **2010**;26(6):841-2 doi 10.1093/bioinformatics/btq033.

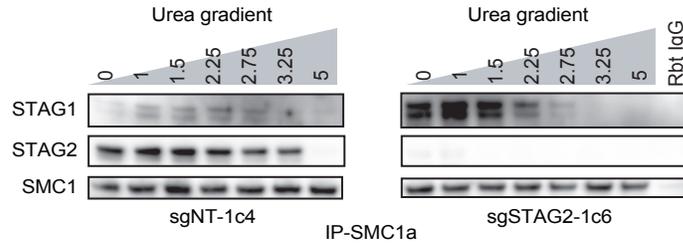
56. Servant N, Varoquaux N, Lajoie BR, Viara E, Chen CJ, Vert JP, *et al.* HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol* **2015**;16:259 doi 10.1186/s13059-015-0831-x.
57. Lareau CA, Aryee MJ. hichipper: a preprocessing pipeline for calling DNA loops from HiChIP data. *Nat Methods* **2018**;15(3):155-6 doi 10.1038/nmeth.4583.
58. Phanstiel DH, Boyle AP, Heidari N, Snyder MP. Mango: a bias-correcting ChIA-PET analysis pipeline. *Bioinformatics* **2015**;31(19):3092-8 doi 10.1093/bioinformatics/btv336.
59. Lareau CA, Aryee MJ, Berger B. diffloop: a computational framework for identifying and analyzing differential DNA loops from sequencing data. *Bioinformatics* **2018**;34(4):672-4 doi 10.1093/bioinformatics/btx623.

Figure 1

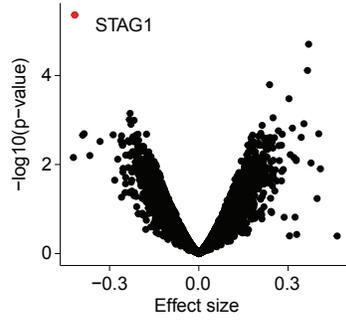
A



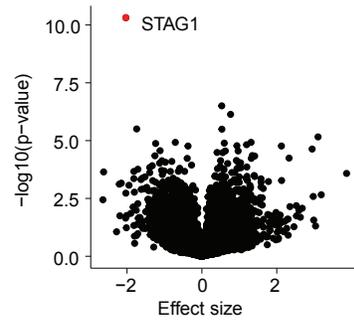
B



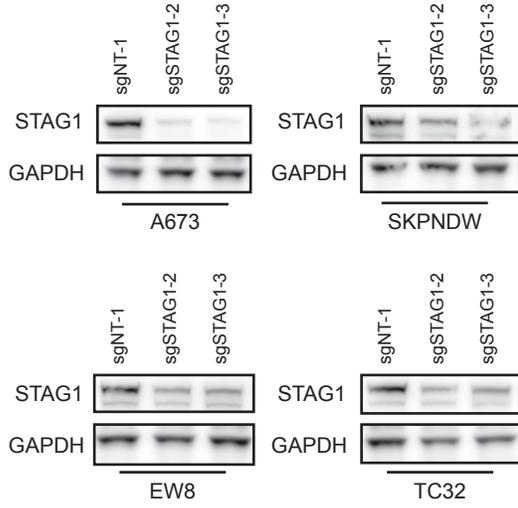
C



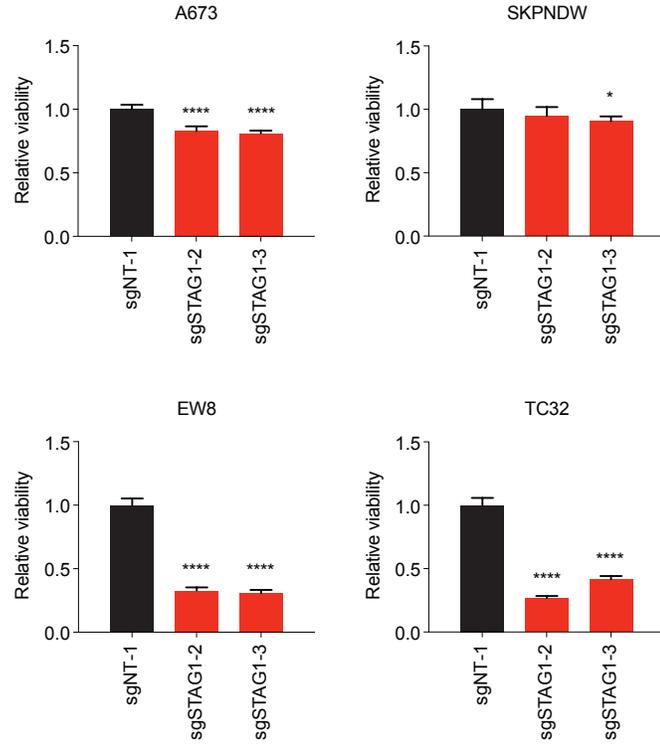
D



E



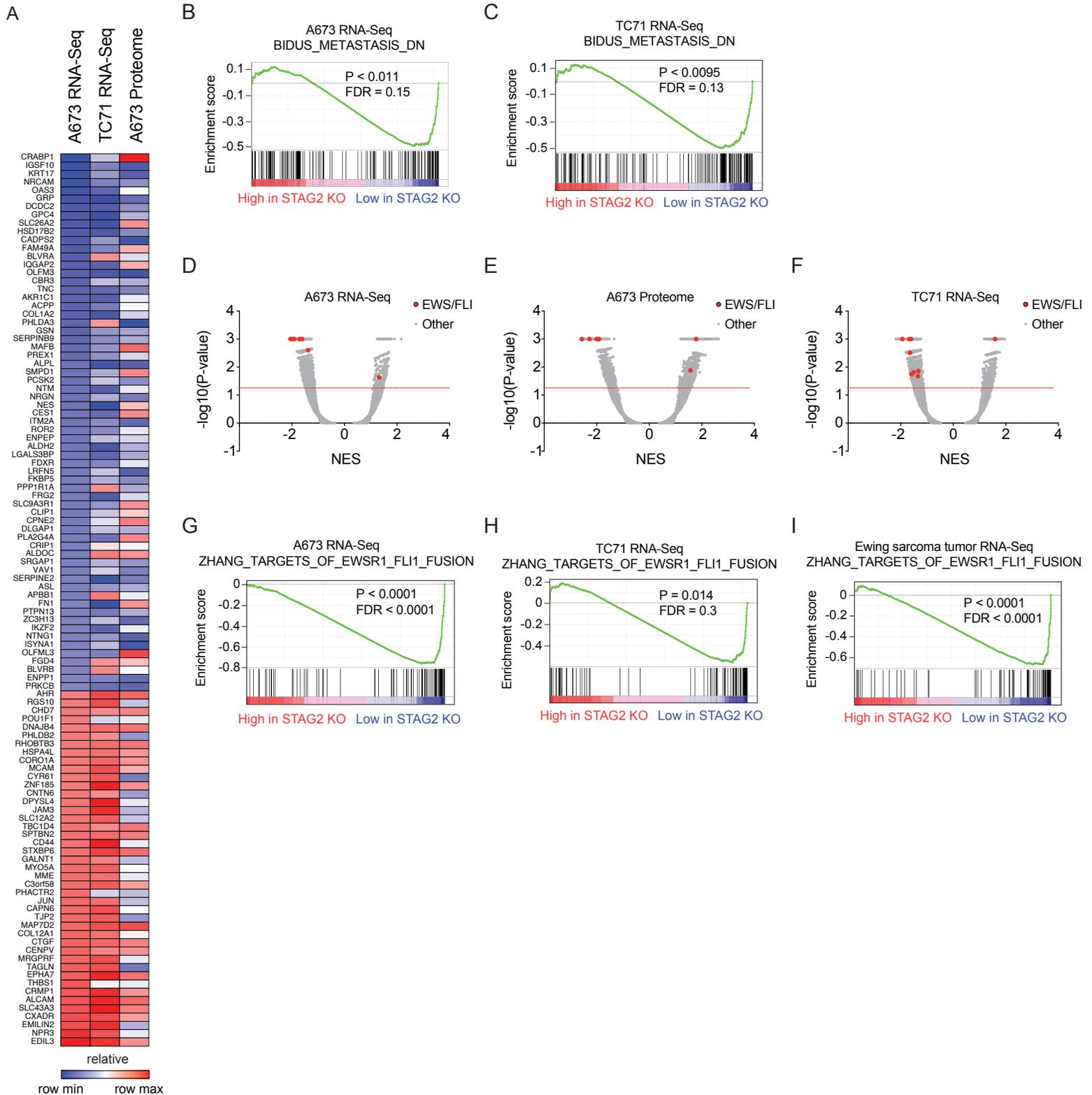
F



**Figure 1. Ewing sarcoma cells with STAG2 loss are dependent on STAG1.**

**A**, Western blot of A673 *STAG2* wild-type (WT) and *STAG2* knockout (KO) clones demonstrating cohesin protein abundance in WCL and after co-IP for SMC1a. **B**, Western blot of A673 *STAG2* wild-type (sgNT-1c4) and *STAG2* knockout (sgSTAG2-1c6) cells treated with the indicated molar concentrations of urea and subjected to co-IP for SMC1a demonstrating differential affinity of STAG1 and STAG2 for cohesin. **C**, Scatter plot of relative gene dependencies in genome-scale CRISPR-Cas9 screen of 7 Ewing sarcoma cells that are *STAG2* mutated compared to 4 Ewing sarcoma cells that express wild-type *STAG2*. **D**, Scatter plot of relative gene dependencies in genome-scale CRISPR-Cas9 screen of A673 cells clonally selected for *STAG2* knockout compared to clones selected for expression of wild-type *STAG2* using two biologic replicates for each class. **C-D**, Gene dependencies are plotted on the x-axis to the left of zero and genes whose loss results in an increased proliferation score to the right of zero. The significance of the deviation is plotted on the y-axis. **E**, Western blot demonstrating *STAG1* downregulation in the indicated Ewing sarcoma cell lines including two *STAG2* wild-type lines (A673 and SKPNDW) and two *STAG2* mutated lines (EW8 and TC32). Cells were treated with non-targeting (sgNT-1) or *STAG1*-targeting (sgSTAG1-2 and sgSTAG1-3) CRISPR-Cas9 guides. **F**, Cell viability effect of *STAG1* knockout across Ewing sarcoma lines expressing wild-type *STAG2* and those that are *STAG2* mutated.

Figure 2

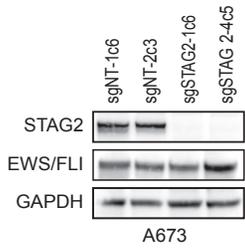


**Figure 2. Gene expression changes associated with loss of STAG2.**

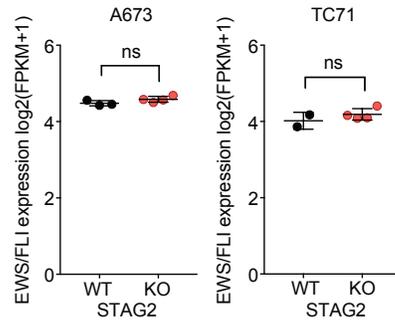
**A**, Heatmap of gene expression changes as measured by RNA-Seq (A673 and TC71) and by mass spectrometry (A673) induced by loss of STAG2 expression in clonally selected cells treated with *STAG2* targeting CRISPR-Cas9. Genes in heatmap had a significant change in expression by RNA-Seq in A673 cells after *STAG2* knockout. **B-C**, Enrichment plots for a select metastatic signature in **B**, A673 cells and **C**, TC71 cells with *STAG2* knockout compared to the wild-type *STAG2* condition as measured by RNA-Seq. **D-F**, Normalized enrichment scores (NES) for genes with significantly altered expression in gene signatures from c2 and c5 collections in MSigDB for cells with *STAG2* knockout as measured in **D**, A673 by RNA-Seq, **E**, A673 by mass spectrometry and **F**, TC71 by RNA-Seq. Red dots indicate the subset of signatures generated from the downregulation or ectopic expression of *EWS/FLI*. Corresponding data in Supplemental Tables S1-3 **G-I**, Enrichment plots for a representative *EWS/FLI* signature in **G**, A673 cells **H**, TC71 cells and **I**, Ewing sarcoma primary patient tumors with *STAG2* loss compared to the wild-type *STAG2* condition as determined by RNA-Seq.

Figure 3

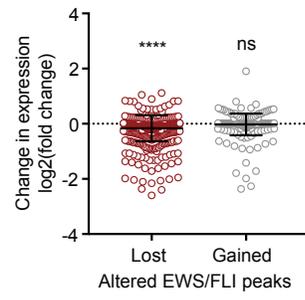
A



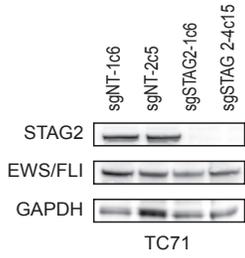
C



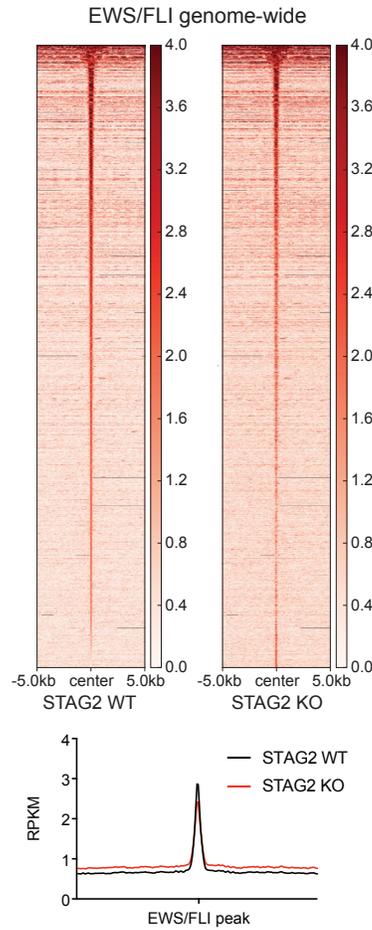
E



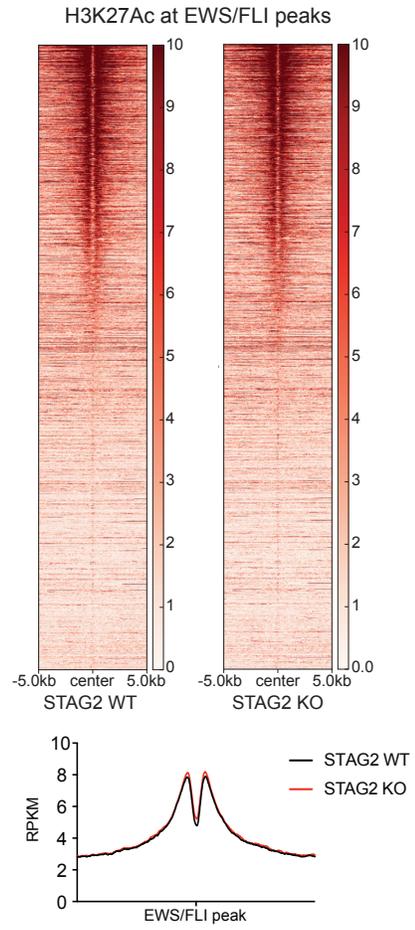
B



D



F

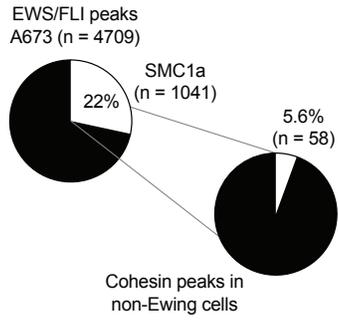


**Figure 3. STAG2 loss has minimal effect on EWS/FLI expression, chromatin binding, and enhancer marking.**

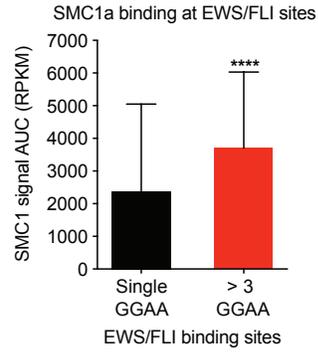
Western blot for EWS/FLI protein levels in **A**, A673 and **B**, TC71 clonally selected for expression of wild-type *STAG2* or *STAG2* knockout. **C**, Expression levels of EWS/FLI as measured by RNA-Seq in A673 and TC71 cells clonally selected for wild-type *STAG2* (WT) and *STAG2* knockout (KO). **D**, Genome-wide heatmaps of EWS/FLI ChIP-Seq signal in A673 cells expressing *STAG2* (left) or *STAG2* knockout (right) with each row centered horizontally on significant peaks identified in either or both conditions and sorted vertically by signal in the *STAG2* wild-type condition. Below the heatmaps is the metaplot of genome-wide signal for each condition. **E**, Gene expression changes plotted for genes closest to an EWS/FLI peak with a significant loss or gain in peak intensity in *STAG2* knockout compared to *STAG2* wild-type A673. The Wilcoxon Signed Rank test was used to determine if either column deviated significantly from the *STAG2* wild-type condition **F**, Heatmaps and metaplot of H3K27Ac ChIP-Seq signal centered horizontally on significant EWS/FLI peaks identified as in (**D**) and sorted vertically by signal in the *STAG2* wild-type condition. For all box plots, a horizontal bar indicates the mean and whiskers indicate the standard deviation (SD). Mann-Whitney test was used to test for significant differences between conditions.

Figure 4

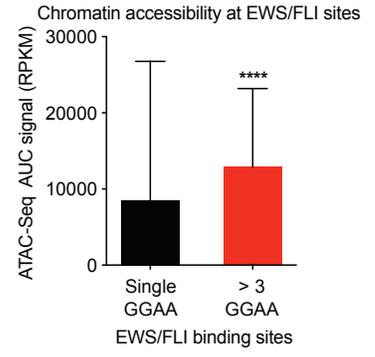
A



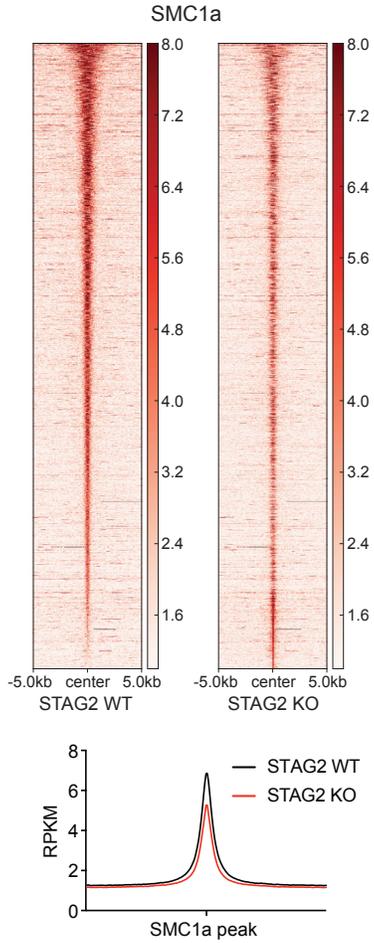
B



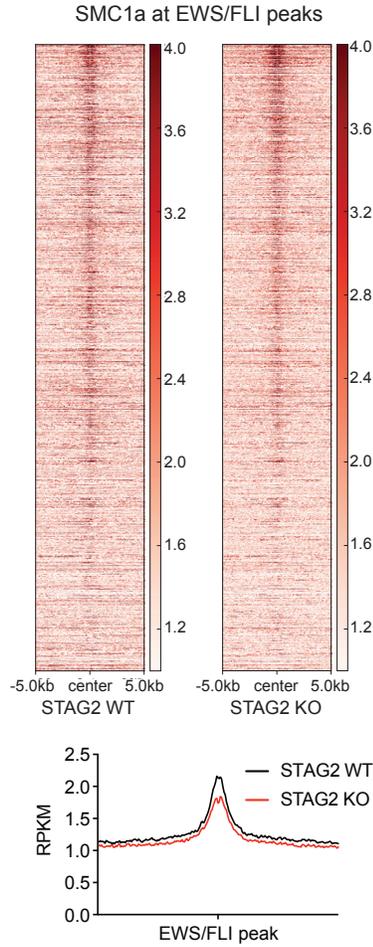
C



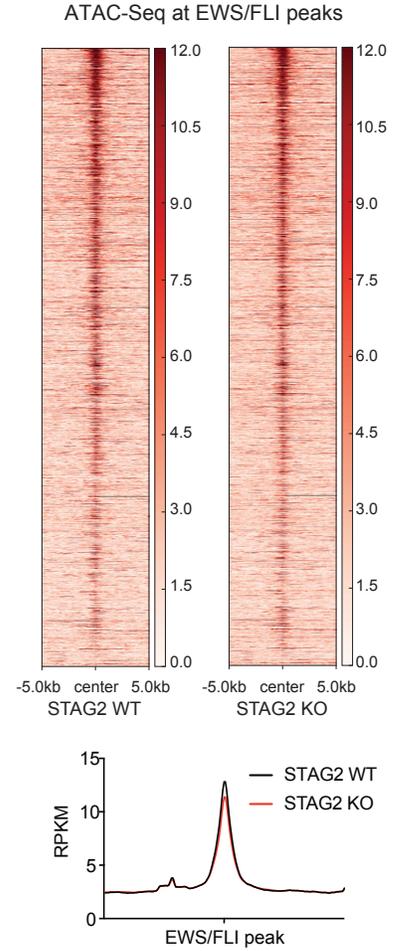
D



E



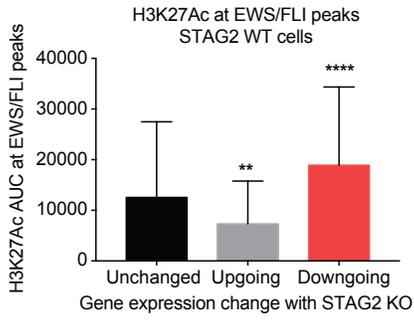
F



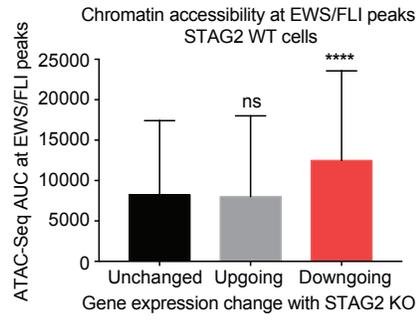
**Figure 4. Cohesin binding is decreased at EWS/FLI-bound chromatin in cells with loss of STAG2.** **A**, Pie charts demonstrating (left) the overlap in high-confidence EWS/FLI binding sites and high-confidence cohesin (SMC1a) binding sites determined by ChIP-Seq for the respective targets and (right) the percent of sites co-bound by EWS/FLI and cohesin in A673 cells which are reported to be bound by cohesin in other (non-Ewing) cell types. **B**, SMC1a signal or **C**, chromatin accessibility overlapping high-confidence EWS/FLI binding sites containing either a single GGAA sequence or a sequence of greater than 3 GGAA repeats. **D**, Genome-wide heatmaps and metaplots of cohesin ChIP-Seq signal in A673 cells expressing wild-type (WT) *STAG2* (left) or *STAG2* knockout (KO; right) centered on a significant cohesin peaks identified in either or both conditions. **E-F**, Heatmaps and metaplots of **E**, cohesin signal and **F**, chromatin accessibility centered horizontally on significant EWS/FLI peaks identified as in (**Fig. 3D**).

Figure 5

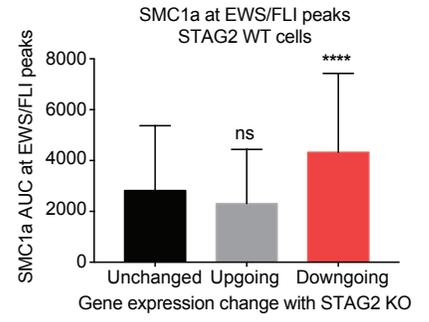
A



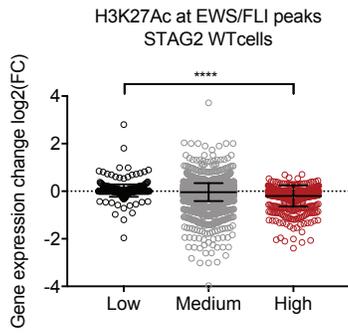
B



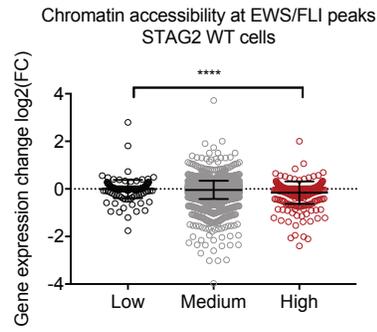
C



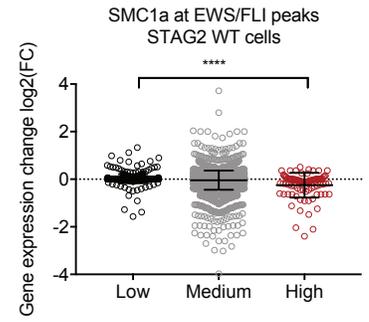
D



E



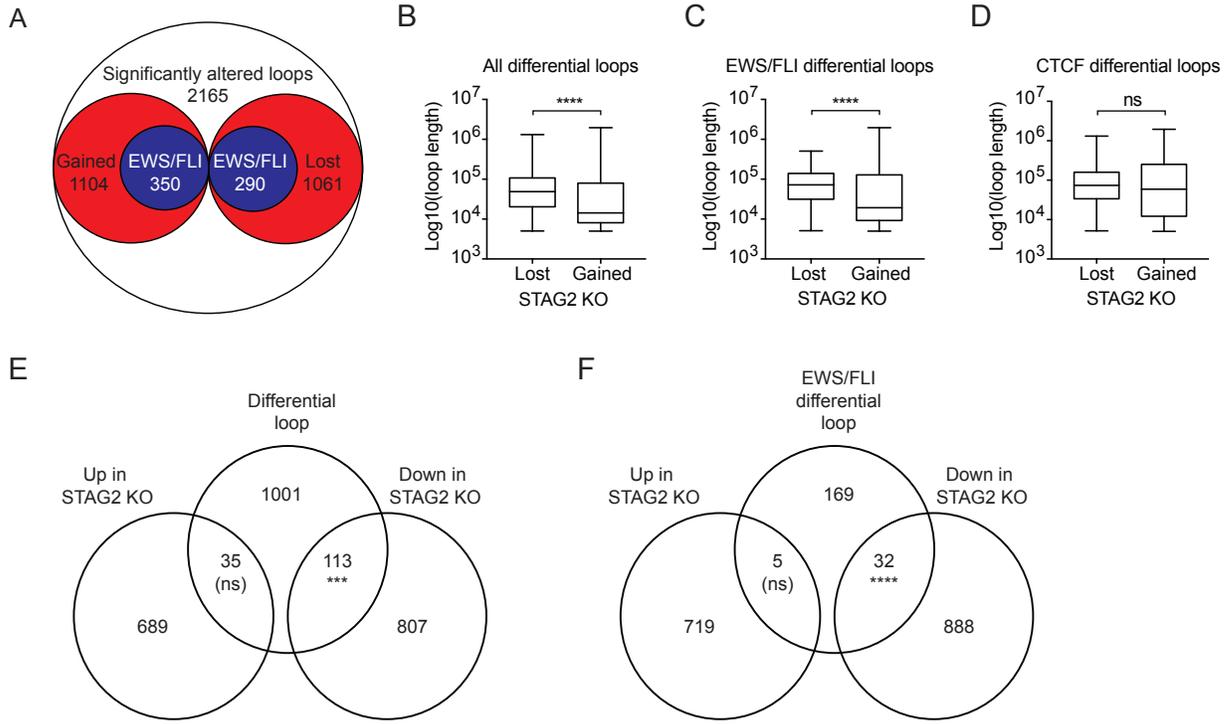
F



**Figure 5. Genes with repressed expression when STAG2 is lost are located near EWS/FLI sites with high cohesin binding, enhancer markings, and chromatin accessibility**

**A-C**, ChIP-Seq signal measured at the nearest EWS/FLI peak in *STAG2* wild-type cells for genes that are unchanged, upgoing, or downgoing after *STAG2* loss for **A**, H3K27Ac **B**, chromatin accessibility and **C**, cohesin. For all bar graphs, the mean and SD are plotted. The Mann-Whitney test was used for comparing two conditions and a one-way ANOVA with Sidak's multiple comparison test used for more than two conditions (ns = not significant, \*\*\*\*  $P < 0.0001$ , \*\*  $P < 0.01$ ). **D-F**, Change in gene expression for genes nearest high-confidence EWS/FLI binding sites based on the level of **D**, H3K27Ac binding **E**, chromatin accessibility or **F**, SMC1a binding at each EWS/FLI site. Low is defined as a normalized z-score of signal intensity equal to or less than -1.5, High as equal to or greater than 1.5, and Medium as everything else. Plotted is the mean and SD. One-way ANOVA with Sidak's multiple comparison test was used for comparing multiple conditions (\*\*\*\*  $P < 0.0001$ ).

Figure 6

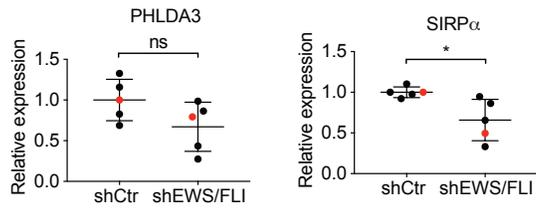


**Figure 6. Alterations in chromatin interactions induced by STAG2 loss are enriched for EWS/FLI enhancer-promoter contacts.**

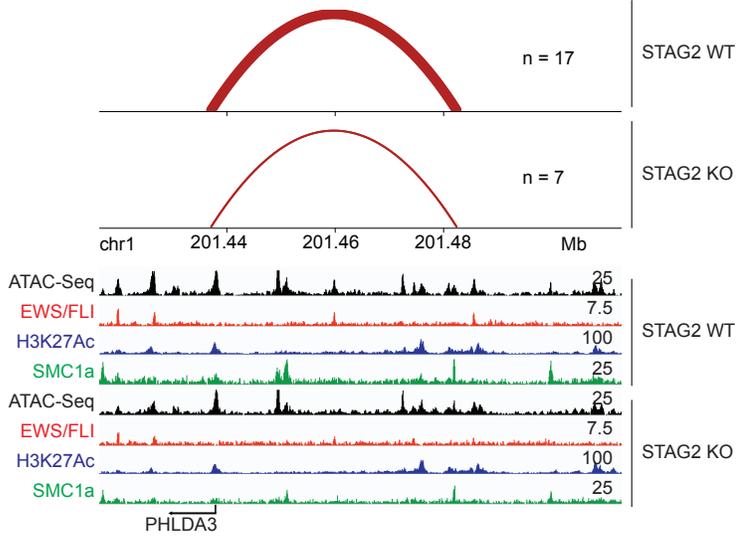
**A**, Diagram depicting overlap of significantly altered chromatin loops after STAG2 loss in A673 cells identified by SMC1a HiChIP. All altered loops represented by white circle, loops gained and lost represented by red circles, and loops anchored at one or more ends by a high-confidence EWS/FLI peak represented by blue circles. **B-D**, Box plots of the distance between loop anchors of significantly altered loops either lost or gained in *STAG2* knockout (KO) A673 cells for **B**, all differential loops ( $P < 0.0001$ ) **C**, loops anchored by EWS/FLI on one or more sides ( $P < 0.0001$ ) and **D**, loops anchored by CTCF on one or more sides ( $P = 0.46$ ). Central line indicates median, boxes 25<sup>th</sup>-75<sup>th</sup> percentiles, whiskers extend to minimal and maximal value and the Mann-Whitney test was used for comparing conditions. **E**, Diagram depicting 1149 genes with the promoter located at the anchor of a differential loop (middle circle) and the overlap of those genes with genes significantly up regulated (left circle) or downregulated (right circle,  $P = 0.001$ ) in *STAG2* knockout. **F**, Diagram depicted 206 genes associated with differential loops anchored by EWS/FLI on the loop edge distal from the gene promoter (middle circle) and overlapping with genes significantly upregulated (left circle) or downregulated (right circle,  $P < 0.001$ ) in *STAG2* knockout.

Figure 7

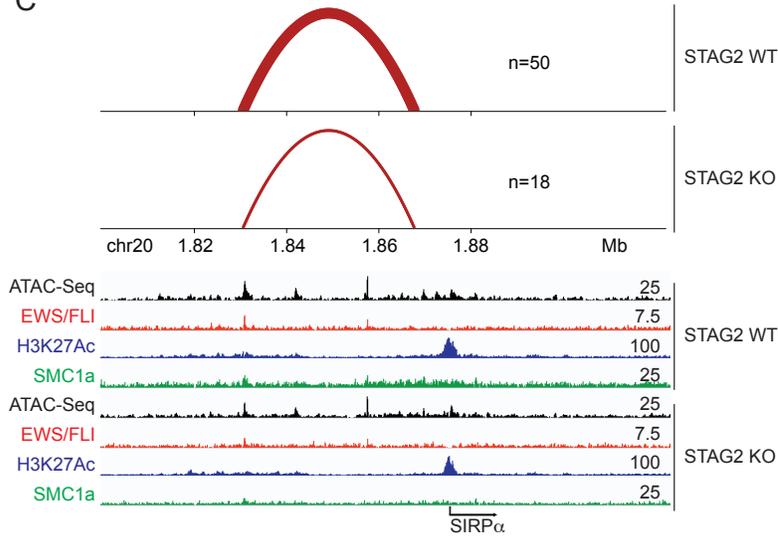
A



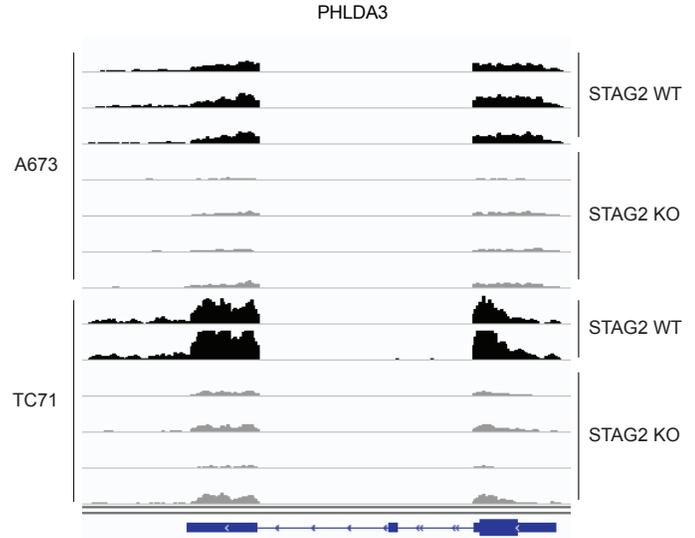
B



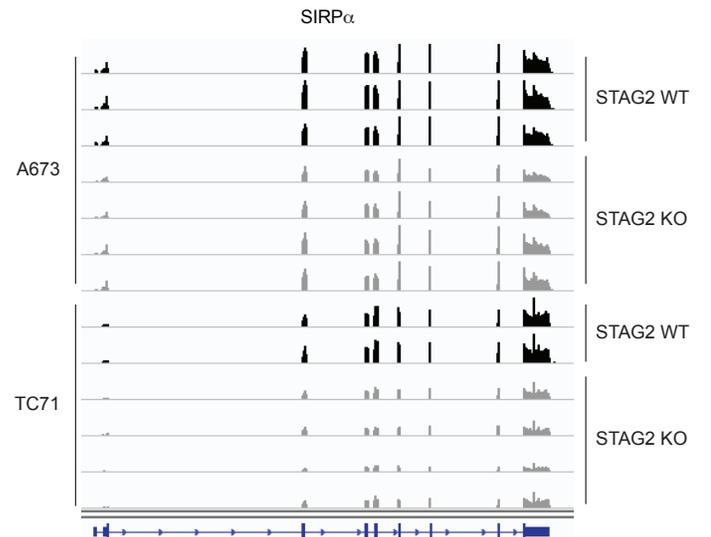
C



D



E



**Figure 7. EWS/FLI contact with a subset of target genes is weakened with STAG2 loss.**

**A**, Relative gene expression levels of *PHLDA3* and *SIRP $\alpha$*  as measured by RNA-Seq in Ewing sarcoma cells treated with control and EWS/FLI-targeting shRNA. Black dots are data from Marques Howarth, M. et al. 2014 and red dots are data from Riggi, N. et al. 2014. **B-C**, Visualization of HiChIP and ChIP-Seq data from for A673 cells expressing wild-type (WT) versus *STAG2* KO for the region around the transcription start site for **B**, *PHLDA3* and **C**, *SIRP $\alpha$* . Top shows arched red lines depicting chromatin contacts between genomic regions at the edge of each arc. Line thickness is proportional to number of contacts detected by HiChIP. Each line is the combination of technical replicates. Bottom shows normalized signal for the indicated ChIP-Seq or ATAC-Seq experiment. Each row is the combination of biological replicates. **D-E**, Visualization of RNA-Seq FPKM normalized signal for **D**, *PHLDA3* and **E**, *SIRP $\alpha$* . (ns = not significant, \*\*\*\* P < 0.0001, \*\* P < 0.01)