The clear cell sarcoma functional genomic landscape

The Clear Cell Sarcoma Functional Genomic Landscape

Emanuele Panza[1,2,†], Benjamin B. Ozenberger[3,†], Krystal Straessler[1,3,†], Jared J. Barrott[3], Li Li[3], Yanliang Wang[3], Mingchao Xie[4], Anne Boulet[1], Simon W. A. Titen[1], Clinton C. Mason[5], Alexander J. Lazar[6], Li Ding[4], Mario R. Capecchi[1]*, Kevin B. Jones[3]*

[1]Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, Utah, USA.
[2]Department of Medical and Surgical Sciences, University of Bologna, Italy.
[3]Departments of Orthopaedics and Oncological Sciences, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, Utah, USA.
[4]Departments of Medicine and Genetics, McDonnell Genome Institute, Siteman Cancer Center, Washington University School of Medicine, St. Louis, Missouri, USA.
[5]Department of Pediatrics, University of Utah School of Medicine, Salt Lake City, Utah, USA.
[6]Departments of Pathology and Genomic Medicine, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA.
[†]These authors contributed equally.

Conflict of Interest Statement: The authors have declared that no conflict of interest exists.

*Corresponding Authors:

Mario R. Capecchi, PhD
Department of Human Genetics
University of Utah School of Medicine
15 North 2030 East, Room 5440
Salt Lake City, UT, 84112
email: capecchi@genetics.utah.edu
phone: 801-581-7096

Kevin B. Jones, MD
Depts. of Orthopaedics and Oncological Sciences
University of Utah School of Medicine
2000 Circle of Hope Drive, Room 3726
Salt Lake City, UT 84112, USA
email: kevin.jones@hci.utah.edu
phone: 801-585-0300

Abstract

Clear Cell Sarcoma (CCS) is a deadly malignancy affecting adolescents and young adults. It is characterized by reciprocal translocations resulting in the expression of the chimeric EWSR1-ATF1 or EWSR1-CREB1 fusion proteins, driving sarcomagenesis. Besides these characteristics, CCS has remained genomically uncharacterized. Copy number analysis of human CCSs showed frequent amplifications of the *MITF* locus and chromosomes 7 and 8. Few alterations were shared with Ewing sarcoma or desmoplastic small round cell tumors, other *EWSR1*-rearranged tumors. Exome sequencing in mouse tumors generated by expressing *EWSR1-ATF1* from the *Rosa26* locus demonstrated no other repeated pathogenic variants. Additionally, we generated a new CCS mouse by Cre-*loxP*-induced chromosomal translocation between *Ewsr1* and *Atf1,* resulting in copy number loss of chromosome 6 and chromosome 15 instability, including amplification of a portion syntenic with human chromosome 8, surrounding *Myc*. Additional experiments in the *Rosa26* conditional model demonstrated that *Mitf* or *Myc* can contribute to sarcomagenesis. Copy number observations in human tumors and genetic experiments in mice render, for the first time, a functional landscape of the CCS genome. These data advance efforts to understand the biology of CCS with innovative models, in which we can eventually validate preclinical therapies, necessary to move toward longer and better survival of the young victims of this disease.

Keywords: Clear cell sarcoma, Ewing sarcoma, translocation, mouse models, copy number variation analysis.

Introduction

Specific chromosomal translocations associate with defined malignancies, especially among leukemias and sarcomas. In some of these translocation-associated cancers, few genetic perturbations other than the characteristic fusion oncogene are identified. While this observation provides strong circumstantial evidence that the fusion protein generated by each chromosomal translocation drives its associated cancer, experimental evidence is preferable.

Early attempts to create mouse models to test the oncogenicity of specific products of chromosomal translocations expressed fusion gene products of chromosomal translocations as randomly inserted transgenes under the control of general promoters (1). Others placed fusion genes under Cre-*LoxP*-conditionally controlled expression from general promoters (2-5). Still others modified the 5' parent locus by the targeted addition of coding sequence from the 3' parent locus, moved into position by Cre-mediated recombination (6). The Cre-*loxP* system has also been used to generate chromosomal translocations in mouse somatic cells (7-10), but only one resulted in the associated leukemia (9).

CCS is a rare, but particularly deadly type of soft-tissue sarcoma that arises most commonly in adolescents and young adults. It was initially characterized by tissue features reminiscent of melanoma, but later found to associate consistently with either t(12;22) or t(2;22) balanced chromosomal translocations, generating *EWSR1-ATF1* or *EWSR1-CREB1*, respectively. Each of these fusion oncogenes was later demonstrated to associate with a slightly broader range of neoplasia, not consistently including the melanocytic differentiation of the original, clinically-defined entity of CCS. A mouse model expressing the human fusion gene, *EWSR1-ATF1* was found to generate a faithful recapitulation of human CCS, as well as other variants of EWSR1-ATF1-related neoplasia, depending on the cell of origin selected for conditional expression of the fusion (5, 11). This observation made the strong argument that

EWSR1-ATF1 is sufficient for induction of CCS and functions as a true oncogene. The presence of other secondary genetic changes in CCS has not been previously characterized.

As much as translocation-driven sarcomas are considered to be otherwise genetically quiet, those that have been better characterized than CCS are not completely lacking other alterations (12-21). For example, Ewing sarcoma is a pediatric bone tumor driven by a related chimeric fusion oncogene, *EWSR1-FLI1.* Beyond the pathognomonic fusion, a minority of tumors show loss of *STAG2* or *TP53* (12, 16, 21). Ewing sarcomas also harbor recurrent copy number alterations (CNAs), such as amplification of 1q and trisomies of chromosomes 8 and 12 (22).

A number of questions arise when such secondary alterations are observed repeatedly in a single type of tumor. Do they represent factors that help to drive tumor formation, or merely passenger mutations that accompany it? If they are co-factors to transformation driven by a specific translocation fusion oncogene, do they overcome some vulnerability engendered by that fusion, or synergize with that fusion to drive oncogenesis? Are the specific secondary changes selected for by necessity or just by increasing the rate of cell proliferation? Can these accompanying genetic changes reveal details about the mechanism of the fusion oncoprotein's function in a cell? Testing the necessity or sufficiency of these secondary changes requires a model for tumorigenesis to which the observed secondary change can be added or removed. Such a model is generally lacking for Ewing sarcoma, but readily available for CCS.

We set out to profile the copy number alterations in a cohort of human CCS tumors, then compare these findings to both observations and experimental manipulations of the genome in mouse genetic models of CCS, including a new model driven by an induced chromosomal translocation.

Results

Human CCS tumors demonstrate recurrent copy number alterations

Formalin-fixed paraffin-embedded (FFPE) human CCS tissue specimens and matched adjacent tissue controls were used to retrieve tumor and germline DNA. Thirteen of these pairs rendered sufficient quantity and quality of DNA to be hybridized onto the OncoScan CNV Plus array, a microarray-based assay of copy number alterations and loss of heterozygosity, which is intended for FFPE-retrieved DNA. The assay enables the detection of relevant data across the whole genome and specifically checks for the presence of 64 hotspot mutations in nine cancer-related genes (*BRAF*, *EGFR*, *IDH1*, *IDH2*, *KRAS*, *NRAS*, *PIK3CA*, *PTEN*, and *TP53*). Only a single tumor demonstrated any of these point mutations, an activating mutation in *EGFR* (Figure 1A).

Across the CCS samples, the most frequent large copy number increase (ranging in length from >10 Mb to an entire chromosome arm) occurred for chromosomes 8 and 9. Ten of the 13 samples (76.9%) demonstrated a copy number increase in 8q and seven of the 13 samples (53.8%) had either gains or deletions of 8p and 9p (Figure 1A). The long arm of chromosome 7 also had a high percentage of copy number aberrations (61.5%, all gains) as did 7p, 9q, and 17q. Analysis for co-occurrence or exclusion found only three significant pairings or associations: co-occurrence of 9q and 11q ($p = 0.035$), 17q and age < 20 years ($p = 0.035$), and exclusion of 8q and 14q ($p = 0.038$). Copy number increases at *MITF* and *FAT1* were observed in six and five of the 13 samples, respectively. No focal loss of *CDKN2A* was observed, though 5 of the 13 samples had deletions of 9p which contains *CDKN2A*.

Some Human CCS CNAs are shared with other tumors associated with *EWSR1* fusions

We performed the same OncoScan Plus array analysis on 8 desmoplastic small round cell tumors (DSRCTs) with matched control tissues. These tumors were chosen because they are characterized by the translocation-generated *EWSR1-WT1* fusion oncogene, thus sharing with CCS the *EWSR1* gene involvement. For DSRCT samples, 5 of the 8 (62.5%) had gains of chromosome 5, which were nearly absent in the CCS samples (Figure1B). Deletions of 8p or 16q (in 3 of the 8, 37.5%, incomplete 16q arm deletion in one of the three) and copy number increases on 18p or chromosome 20 (also in 3 of the 8, 37.5%, incomplete 18p arm deletion in one of the three) were also observed. No co-occurrence or exclusion of gains or deletions were observed at an unadjusted $p < 0.05$ level. The gains at *MITF* and *FAT1* observed in CCS were absent in DSRCT.

Overall, the frequencies of large (>10 Mb) increases in copy number on 8q and 8p, as well as focal gains involving *MITF*, were greater in CCS than DSRCT, while gains of 5q, 5p and 20p, and loss of the X chromosome were significantly more frequent in DSRCT samples.

We compared the frequency of common copy number alterations in Ewing sarcoma (ES) with the frequencies seen in CCS and DSRCT. Trisomy of chromosome 8 is reported in the majority of *EWSR1-FLI1* tumors, along with amplifications of 1q and whole chromosome 12 (Figure 2). Using the findings from Tirode et al. (21) on 112 ES patient tumors as a comparator, we found no significant difference in the prevalence of copy number increases in 1q and chromosomes 8 and 12 between ES and CCS. The prevalence of deletions of *CDKN2A* and 16q in ES and CCS were also not significantly different. The same findings of no difference in prevalence were observed when comparing DSRCT to ES for these specific alterations, except for a significantly higher prevalence of chromosome 8 copy number increases in ES compared to DSRCT ($p = 0.02$).

EWSR1-ATF1 is sufficient for murine sarcomagenesis without additional protein coding mutations

Mice engineered to express human *EWSR1-ATF1* conditionally from the *Rosa26* locus were induced using *Bmi1Cre*^{ERT2}, with tamoxifen administration at a variety of ages (Figure 3A-C). This specific lineage is optimally suited to produce clonal sarcomas with clear cell morphology and the characteristic CCS melanoma-like features (5). When exomes from 34 tumors harvested from six mice (Figure 3D, and each large enough to permit isolation of high-quality DNA and RNA samples) were sequenced to a minimum depth of 20X coverage, and compared to matched host germline exomes sequenced to a minimum depth of 8X coverage, 7846 variants were identified. Of these, high-quality variants were defined as having no more than 1 variant read in the host germline and at least 2 variant reads in the tumor, yielding 265 potential somatic variant alleles (Supplemental Table 1). This list of potential somatic variants was subjected to targeted sequencing to a minimum depth of 100X coverage for both germline controls and tumors. A variant allele frequency (VAF) of no more than 0.01 in a matched control and at least 0.05 in a tumor yielded 107 variants in 26 of the tumors, with eight tumors demonstrating no variants reaching these thresholds (Figure 3E, Supplemental Table 2).

KEGG pathway analysis of the genes implicated by the 82 variants with VAF greater than 0.1 identified shared involvement in calcium signaling, chemical carcinogenesis, and aldosterone synthesis/secretion pathways, none reaching more than minimal significance. Because the tumor cell prevalence in mouse CCS tumor tissues is consistently greater than 80 percent, a VAF of 0.4 was considered the lowest that might represent a clonal variant across the tumor cells, which could potentially have contributed to initial tumorigenesis. The expression level of the 8 genes with greater than 0.4 VAF was tested in a subset of 22 of these 34 tumors by RNA sequencing; only 3 showed significant expression, a level set at one percent of *Gapdh* expression and comparable to that of *Mitf* (Figure 3F, Supplemental Tables 3 and 4). Thus, less

than 10 percent of tumors bore even a single mutation that was found to be clonal across the tumor cells and that affected a gene that was expressed generally by this tumor type.

Exome data were also analyzed with regard to copy number variation. No repeated changes were identified across the group of tumors (Figure 3G). No whole chromosome gains or losses were noted in any tumor's exome. No large fragment changes reached significance.

Induced chromosomal translocation in mice drives tumorigenesis and reveals other secondary changes

We next sought to develop a new mouse model that could recapitulate CCS more closely at the genetic level, by an induced chromosomal translocation. In the mouse, both the *Ewsr1* and *Atf1* loci bear significant homology to each respective parent locus in the human genome (5). However, in contrast to their human counterparts, the mouse genes are in opposite orientation with respect to each host chromosome's centromere, such that recombination generates an acentric chromosome bearing the *Ewsr1-Atf1* fusion gene as well as most of chromosome 11 and a small fragment of chromosome 15, and a dicentric chromosome bearing most of chromosome 15 and a small fragment of chromosome 11 (Figure 4A). Although different from the human t(12;22) translocation, the mouse translocation would therefore permit investigation of two important aspects of oncogene-driven sarcomagenesis. First, retention, or possibly amplification, of an acentric chromosome fragment with the fusion oncogene would indicate that either tumor formation and growth were dependent on the presence of the fusion oncogene, or that there was selection for tumor cells in which the fusion oncogene was present. Second, dicentric chromosome generation is thought to drive aneuploidy due to double strand breaks that develop during cytokinesis after the two centromeres are pulled to different daughter cells. This should encourage copy number variation and even chromothripsis, which might uncover secondary changes that accompany murine clear cell sarcomagenesis driven by a chromosomal translocation. Therefore, using homologous recombination, we targeted a single

loxP site into the 7th intron of the *Ewsr1* locus on mouse chromosome 11 and another into the 4th intron of the *Atf1* locus on mouse chromosome 15 (Figure 4A).

Double heterozygous mice that also had *HprtCre*, *Prx1Cre*, or limb injection of the protein TATCre generated tumors at similar penetrance (Figure 4B). Tumors arose in mesenchymal tissues, similar to the *Rosa26-EA1* models, but at longer latencies and lower prevalence (Figure 4C-D). Histologically, translocation-generated tumors matched the range of tumors produced in the *Rosa26* models as well as the range of neoplasia associated with *EWSR1-ATF1* in humans, showing the typical clear cytoplasm and enlarged, irregular hyperchromatic nuclei (Figure 4E).

We isolated total RNA from 8 tumors, generated complementary DNA (cDNA) libraries using the RiboZero method, then sequenced these on an Illumina HiSeq instrument to generate transcriptomes comparable to a prior sequencing effort of many mouse genetic cancer models, including the *Rosa26-EA1* model initiated by a variety of drivers for Cre-recombinase, as well as a set of mixed lineage mesenchymal controls from the chest wall (with skeletal muscle, bone, cartilage, tendon, aponeurosis in each). CCS tumors from each induction method shared a somewhat similar expression profile overall and clustered near each other on principal component analysis compared to the other tumor types (Figure 5A-C).

Next, for a tighter comparison, we sequenced in a single batch, transcriptomes from 5 tumors induced by *HprtCre*-initiated translocations and 5 tumors induced by TATCre injection to activate the *Rosa26-EA1* allele. Again, these clustered somewhat closely to each other, relative to a control, albeit with slightly wider variation among the translocation-initiated tumors by PCA and whole transcriptome distance clustering (Figure 5D-E).

We investigated copy number alteration in four of the translocation-generated tumor samples, using a low-read-depth whole genome sequencing approach (Figure 5F). The data demonstrated that the genome segments that comprised the dicentric chromosome (a very short segment immediately adjacent to the chromosome 11 centromere and most of the body of

chromosome 15) had variable copy number, consistent with the expected breakage of this chromosome during unbalanced cytokinesis. However, rather than whole genome chromothripsis, which might be the expected result of double strand breaks, the only other associated CNA was a copy number decrease of chromosome 6 in the translocation model. The bulk of chromosome 11, which accompanied the translocation-generated *Ewsr1-Atf1* fusion on the acentric derivative chromosomal fragment maintained an approximately diploid copy number, showing neither significant amplification or loss.

Mitf contributes modestly to *EWSR1-ATF1*-driven tumorigenesis

Surprisingly, the chromosome 6 copy number decrease that was detected in the translocation-initiated mouse model mapped to syntenic portions of human chromosome 7 that showed copy number increases in human CCS. None of the chromosomal regions demonstrating copy number decreases in more than one human CCS tumor correlated with syntenic segments lost in chromosome 6 in the mouse.

The *Mitf* locus, homologous to the region of focal copy number increases in human CCS, is also on mouse chromosome 6 (Figure 6A). Prior research has identified *MITF* (melanocyte-inducing transcription factor) to be responsible for the pigment generation and melanocytic morphology of CCS (23). Work in cell lines suggested that *MITF* drives progression of CCS, functioning as an additional oncogene, similar to a role defined for it in melanoma (24). We previously noted that *EWSR1-ATF1*-induced mouse sarcomas that exhibited melanocytic features expressed *Mitf (5)*. In order to test the oncogenic contribution of *Mitf*, we procured a mouse strain with a natural loss-of-function point mutation in *Mitf*, termed *Mitf* ^{vit}, for the vitiligo (loss of pigment) that develops patchily in heterozygotes, but completely in homozygotes (25).

Fortuitously, in this naturally occurring point mutant, a BsiEI endonuclease digestion site was disrupted by the *Mitf* ^{vit} mutation such that enzymatic digestion of a PCR-generated fragment incorporating this region could be used for genotyping (Figure 6B). The *Mitf* ^{vit} and

Rosa26 loci are both on mouse chromosome 6 at a 15 Mb distance. After nine matings and nearly 100 progeny screened, meiotic recombination generated a single chromosome 6 with both conditionally activatable *EWSR1-ATF1* and *Mitf* ^{vit} segregating together. Mice bearing this doubly targeted chromosome 6 and a second chromosome 6 with only the *Rosa26-EA1* allele were then bred to *Mitf* ^{vit} heterozygotes to generate littermate cohorts of mice heterozygous for *Rosa26-EA1* and either homozygous for wildtype *Mitf* (*Mitf* $^{w}/_{w}$) or homozygous for the vitiligo allele (*Mitf* $^{vit}/_{vit}$). These cohorts were injected with TATCre protein at age 4 weeks and monitored for tumorigenesis (Figure 6C). Sarcomagenesis to the point of morbidity occurred at a latency that was significantly longer in the homozygous vitiligo mice; similarly sized tumors were produced, but at longer latencies (Figure 6D-E). Blinded assessment of histology of the tumors demonstrated a subtly reduced frequency of the clear cell morphology that dominated the tumors in the wildtype *Mitf* group.

In order to test the potential contribution of higher baseline expression of *Mitf* in a potential cell of origin for *EWSR1-ATF1*-mediated tumorigenesis, we crossed mice bearing the conditional expression allele, *Rosa26-EA1*, to mice bearing Cre-recombinase under the control of the *tyrosinase* gene (*Tyr*^{CreERT2}), which is expressed in melanoblasts, which have very high native expression of *Mitf*. Topical administration of tamoxifen to induce Cre-mediated recombination in dermal melanoblasts generated no tumors. Systemic administration of tamoxifen at an age of 4 weeks led to a few tumors developing at a latency of greater than one year. While these tumors demonstrated GFP signal, suggesting expression of the *EA1* allele at *Rosa26*, they developed at much longer latency than any other model of CCS induction from *Rosa26*.

Myc overexpression contributes to _EWSR1-ATF1_-driven sarcomagenesis, but alters tumor phenotype

It was also noted that the CNA distribution in the translocation-initiated tumors, across the length of mouse chromosome 15 included on the dicentric chromosome, was not random. Specific segments of mouse chromosome 15 were amplified in all four tumors and others demonstrated consistent copy number losses. Approximately a third of human chromosome 8, including the portion that was amplified in a majority of the human CCS tumors, is syntenic to the part of mouse chromosome 15 that underwent CNA. A sub-region of this syntenic region was amplified in all 4 translocation tumors, but significantly so in 3. This included a number of potential oncogenes, such as *Fzd6*, *Angpt1*, *Borg*, *Baalc*, *Azin1*, and *Dcaf13* (Figure 7A). *Myc* was in a portion that exhibited neither increased nor decreased copy number in mouse chromosome 15 (Figure 7B).

Copy number amplification of the *MYC* locus was observed in 9 of the 13 human CCS tumors (Figure 7A). In order to test the potential contribution of a locus commonly thought to be the reason for selection for chromosome 8 amplification in other tumors, we bred the *Rosa26*^{EA1}/$_{EA1}$ mice to mice heterozygous for a conditionally expressed activated mutant allele of *Myc* knocked into the *Igs2* pseudogene locus (Figure 7C-D). Littermate cohorts were injected with TATCre at age 4 weeks and sarcomagenesis was monitored. There was a dramatic shortening of the latency to tumorigenesis in mice heterozygous for the conditionally activated *Myc* allele (Figure 7E). Histologically, however, tumors with added *Myc* overexpression had features that were strikingly dissimilar from the *EA1*-only tumors, including poor recapitulation of CCS characteristics, with predominance of a nested histomorphology that is very rarely identified in human CCS (Figure 7F-G).

Discussion

The genome of human CCS is characterized most prominently by copy number increases focally at the *MITF* locus and of whole chromosomes 7 and 8, with each of these copy number alterations identified in a majority of tumors. The functional character of these observed alterations in copy number was informed by comparisons with other groups of tumors: two human tumor types with alternative *EWSR1* rearrangements and two mouse genetic models of neoplasia induced by expression of *EWSR1-ATF1*, either as a conditional transgene or the product of a conditional chromosomal translocation of the homologous mouse loci. The last of these, the successful generation of a sarcoma from a Cre-*loxP*-induced chromosomal translocation, was a singular achievement in itself, but also instructive with regard to the function of the CCS genome.

These copy number data provide the first glimpse into the human CCS genome. Because of the small numbers of CCS tumors available at any single center, we utilized FFPE samples collected over two decades, and thus were limited to copy number variation assessments, instead of full genome or exome sequencing. Given that sarcomas typically progress toward genomic complexity by copy number change, rather than point mutation burden (14), and that only a single event affecting any of the hot-spot mutations in known cancer genes was identified in a single CCS tumor, copy number data may provide an acceptable genome landscape for this tumor type. Some altered copy number observations were shared with ES and DSRCT, both of which bear low mutational burdens (12, 13, 15, 16). Naturally, the significance of these comparisons is limited by small numbers in the CCS and DSCRT cohorts, advising caution in over-interpretation of the numbers.

Although we did not observe focal loss of *CDKN2A* in our cohort of 13 human samples, 5 of them showed deletions of 9p, which contains the *CDKN2A* locus. Three others (none overlapped with 9p deletion) showed amplification of the portion of chromosome 12 that

contains the *MDM2* locus. Amplification of *MDM2* or loss of *CDKN2A* can reduce the stability of p53. Disruption of p53 stability is considered an important step for many cells undergoing malignant transformation driven by the expression of an oncogene, as they must overcome oncogene-induced senescence. In particular, Komura and colleagues (26) demonstrated that escape from senescence is an important hurdle to pass during initial growth of EWSR1-ATF1-expressing cells. These findings from human CCS genomes are at least a nod in support of that line of thinking.

The most important conclusion from each of these assembled and compared datasets is that no single secondary genetic alteration is strictly required for *EWSR1-ATF1*-driven sarcomagenesis. In the human tumors, even the most prevalent alterations were still absent in a strong minority of tumors. Whole exome sequencing and copy number analyses of mouse sarcomas initiated from *EWSR1-ATF1* expression demonstrated no single change that was repeated in more than one tumor. The observation that tumors developed at all in a mouse model in which an acentric chromosomal fragment was host to the *Ewsr1-Atf1* fusion testifies to the profound transforming capacity of this particular oncogene.

While two solid tumor types have been generated previously in the mouse by CRISPR/Cas9-induced translocations (27, 28), we know of no prior solid tumors arising from Cre-*loxP*-induced translocations and no other previously reported opportunities to compare tumorigenesis from induced translocations to that from expression of the fusion gene coding sequence alone.

A chromosomal translocation achieves multiple changes, any of which may contribute to oncogenesis. First, newly associated exons express a fusion oncoprotein that wields impacts not shared with either parent locus. Second, translocation positions the new fusion gene under the genetic and epigenetic control of a specific promoter and enhancers, which regulate the level of expression and responsiveness to other novel conditions. Third, translocation disrupts

one allele of each of the parent loci. Fourth, translocation may broadly affect other nearby genes by reposition.

Obviously, the impact of changes other than the fusion oncogene will vary with each specific type. Not all parent alleles will have sufficient homology with their human counterparts to be oncogenic in the mouse. Further, the regional impacts of the translocation may be more or less homologous depending on how widely these impacts reach. For example, *Atf1* is on mouse chromosome 15, where only a small region immediately surrounding the locus is syntenic to the host chromosome of *ATF1* on human chromosome 12 (29). The breadth of these syntenic regions will determine whether enhancer elements and other regionally important sequences are shared sufficiently to make the model of these larger shifts in mouse chromosomal segments a true mimic of the human translocation.

The translocation mouse model for CCS provided additional data. As noted, the importance of *Ewsr1-Atf1* in sarcomagenesis was strengthened by its retention in cells even on an acentric chromosome that required selective pressure to be present. However, also of interest was the fact that this acentric chromosome with the fusion was not amplified in the tumors. The *Rosa26*-driven model also showed no amplification of fusion expression locus, which was noted as the only other secondary change in the exomes of a similar model of synovial sarcoma (4). This suggests that the continued presence of the fusion is critical in the tumor cells, but that there is no apparent selection for increased levels of expression.

The loss of one copy of chromosome 6 in the translocation tumors was especially interesting because loss of chromosome 6 would have been artifactually impacted in the *Rosa26*-driven model. Because the fusion expression locus maps to chromosome 6, negative selection pressure would likely prevent loss of one of the two copies of the chromosome in that model. The mouse translocation model's copy number decrease of chromosome 6 and the tumors that formed even in mice homozygous for $Mitf^{vit}$ in the *Rosa26* model suggest that melanocytic MITF is not strictly required for clear cell sarcomagenesis. However, evidence from

another CCS mouse model, where the fusion *EWSR1-ATF1* is expressed from the *Col1A1* locus on mouse chromosome 11 (26) corroborates our human CNA analysis showing *MITF* amplification and the subtle—but significant—increase in latency to morbid tumorigenesis that we observed in the *Mitf*^{vit} experiments. MITF likely facilitates tumor formation from *EWSR1-ATF1* expression, possibly also impacting tumor phenotype. This last point is important given that the human samples profiled in this group were tumors fitting the traditional pathological definition of CCS, including melanoma-related features. It is known that non-melanocytic human *EWSR1-ATF1*-driven tumors are less dependent on *MITF* than *EWSR1-ATF1*-driven neoplasia that includes melanocytic features (30). On mouse chromosome 6, there may be other tumor suppressors or factors that would prohibit maintenance of the acentric chromosomal fragment critical to this particular translocation model, but mouse chromosome 6 has no synteny with the few significant copy number decreases observed in human CCS. In spite of this subtle copy number loss of the chromosome, *Mitf* expression was strong in the translocation tumors by RNAseq.

The CNAs in mouse chromosome 15, syntenic to the regions of human chromosome 8 that were so frequently gained in human CCS tumors, also suggested additional drivers to consider. *MYC* is often considered as a driving oncogene to lend selective pressure for human chromosome 8 copy number gains. However, this was not specifically amplified in the translocation-driven mouse tumors. Furthermore, while amplified *Myc* expression and stability was clearly capable of enhancing tumorigenesis when added to *EWSR1-ATF1* expression, it did not faithfully maintain the tumor phenotype of CCS, suggesting that while *MYC* is—not surprisingly—oncogenic, it may not specifically be a driver of clear cell sarcomagenesis, as it is a driving pathway in many types of cancer (31). The other genes in the cross-species amplified regions deserve additional attention in future investigation.

Finally, the amplifications in chromosome 7 in human CCS also require additional investigation. Of note, *EGFR*, located in the consistently amplified regions of human

chromosome 7 is also present on syntenic regions of the mouse acentric chromosomal fragment (including most of mouse chromosome 11) in the translocation-driven model. Again, this was specifically retained in tumors, but not amplified. *EGFR* contained the only hot spot mutation that was detected in any of the tested human tumors. This is another locus for specific future attention, especially given the potential for therapeutic targeting of EGFR.

Overall, these tumors—like others in the family of translocation-associated sarcomas—have a relatively quiet genome, quite distinct from melanomas, to which they are also related. While no other mutations are strictly required in the process of sarcomagenesis, some of the frequently observed secondary alterations may provide additional insights into relative genetic dependencies for this cancer or suggest means by which it can circumvent attempts to target the fusion biology directly. The EWSR1-ATF1-associated family of neoplasia includes very aggressive malignancies that have a predilection for adolescents and young adults. We have no targeted therapies available for these deadly sarcomas. More efforts to understand the biology of these tumors and innovative models in which to validate therapies preclinically are needed to move toward longer and better survival for the young victims of these diseases.

Methods

DNA microarray and CNA analysis on human FFPE tumors

Extracted DNA was assessed by the OncoScan FFPE Assay Kit by Agilent, Inc (CA, USA). The raw microarray data for each sample were then used to generate OSCHP files for analysis by Nexus Express, version 10.0 (BioDiscovery, CA, USA). Initial calls of copy number gains and loss were made with the FASST2 method and default settings. Distributions of intensity values and B-Allele Frequency (BAF) for each sample were manually inspected to

identify diploid chromosomes and subsequent re-centering of probes based on these diploid regions was performed. For each sample, the p and q arms of each chromosome were assessed by Nexus and manual inspection for regions of gain or loss extending 10Mb or more and present in an estimated 10% or more of cells, allowing for analysis of substantial clones being present in addition to near complete clonal expansion. Focal gains and deletions involving individual genes not part of 10Mb gains or deletions or frequent in other sarcomas was also determined, with additional manual inspection of the most prevalent genes. Concomitance of gain or loss across the p and q arms of each chromosome as well as associations with age, gender, primary tumor status, and metastatic tumor status were assessed with Fisher's Exact test and the Cochran-Mantel-Haenszel statistic.

The quality of each processed sample was assessed by its osmapd score – which showed a clear separation of values. Samples having a mapd score > 0.5 were excluded as well as a sample having a mapd score near this threshold (0.41) which showed potential sample contamination. Paired germline samples were not available for all persons but when present, the frequencies of copy number events were assessed and aberrations present in both germline and tumor were excluded. All samples passing these preliminary quality control checks were then assessed to be from unique persons by comparing 4,948 common SNPs. The *MITF* region was found to have frequent copy number gain calls, and also had frequent loss of heterozygosity. To identify whether LOH in this region was specific to CCS or more likely to be a common germline event, two publicly available datasets processed on the identical microarray were also processed and compared (https://www.ncbi.nlm.nih.gov/gds ; GSE147353, GSE110026). LOH in the MITF region was found to be a common event in both of these sets, though gains were nearly absent in the SCRC data, but found to be called somewhat frequently in the PNBX data. Assessment of significant concomitance or exclusion across arms and with covariates was performed with Fisher's Exact test. Only 3 tests were found to meet an

unadjusted P<0.05 threshold – losing significance when adjusted for multiple comparisons. As a result, we share these top 3 results under the caveat that they will need future replication before being considered bona fide.

Exome sequencing of *Bmi1Cre^{ERT2}-Rosa26* murine tumors

Paired-end libraries were produced with enrichment using Agilent Sure Select, followed by alignment to the UCSC mm9 mouse reference (derived from NCBI Mouse Build 37) with BWA (version 0.5.9 with parameters -t 4 -q 5). A total of 34 tumor samples were paired with 6 matched normals and processed through our somatic variation pipeline using Samtools, Varscan, GATK and Pindel.

All putative somatic variants were annotated using transcripts identified in release 58 of Ensembl (58_37K). A single representative transcript was selected for annotation of each variant based on the most extreme functional effect and longest transcript isoform. Filters were applied removing variants that affected only 3'UTR, 5'UTR, intronic sequence, and intergenic sequence, in addition to any coding SNPs that affect transcripts without a complete open reading frame.

Using supporting read counts we identified 323 high-quality nonsynonymous somatic variants. High quality variants had a minimum of 8 reads coverage in the normal and 20 reads coverage in the tumor, a maximum of 1 supporting read in the normal or 1% VAF (variant allele fraction), and a minimum of 2 unique supporting reads in the tumor with a minimum tumor VAF of 10%.

KEGG pathway analysis was performed and corroborated by two programs, g:Profiler and Enrichr (32-34). Copy number analysis was derived from exome-sequencing data by

Bioconductor's DNAcopy package (Seshan VE, Olshen A, 2020. *DNAcopy: DNA copy number data analysis.* R package version 1.62.0)

Mouse studies and generation of *loxP* animals

The *Rosa26*^{EA1} (5), *HprtCre* (35) and *Prx1Cre* (36) mice were all previously described. Mice for these experiments were obtained from the laboratory colony. The first line was generated in the laboratory, but the latter two lines were obtained originally from The Jackson Laboratory (Bar Harbor, Maine, USA). All mice were maintained on a mixed 129/SvJ and Bl/6 strain background, utilizing littermate controls for any comparative experiments between groups. Sex of mice was not intentionally varied between groups and all groups had both sexes represented. Ages of mice are noted in all tumorigenesis experiments.

Using Gene Targeting, we inserted a *loxP* element in intron 7 of the *Ewsr1* gene, and in intron 4 of the *Atf1* gene. In this way we generated two distinct animal lines bearing *loxP* engineered chromosomes. We designed gene targeting vectors using the Gene Construction Kit software (GCK, TextCo Biosoftware). To generate the vectors we used a combination of traditional cloning techniques and recombineering (37, 38). The procedures for the electroporation of mouse ES cells, the selection of positive clones, the blastocyst injection, the generation of chimeras and the germline transmission has been performed as previously described (38).

Heterozygous animals *Ewsr1 loxP/WT* and *Atf1 loxP/WT* were crossed to obtain double heterozygous and finally double homozygous *Ewsr1 loxP/loxP* , *Atf1 loxP/loxP* animals. The genotype of the animals was confirmed by PCR using specific primers (*Ewsr1*-Forward TAGGGCTGGCATTCTTTAAG and *Ewsr1*-Reverse TTAAGAGCCTTAAGAGTCGG;

Atf1-Forward TGGATTGGTTATGGAGCAGG and *Atf1*-Reverse GTAGGCCTTGGATATTACCC). In both the *Ewsr1* and *Atf1* case, with these oligos the targeted allele gives an amplicon of 100 bp longer than the wild type allele.

TATCre injection

TATCre is a soluble form of the Cre protein that can reach the nuclei of cells through the TAT epitope. TATCre animal injections have been performed as previously described(5).

Magnetic resonance imaging

All imaging was performed on euthanized animals, using an HT Bruker BioSpec 7.1 T horizontal-bore MRI instrument.

RNA extraction

One biopsy was immediately collected in RNA later (ThermoFisher Scientific, USA) and stored at 4°C for 3 days to preserve the RNA integrity until RNA sequencing was performed. RNA extraction was performed with Qiagen RNeasy kit (Qiagen, Hilden Germany), following manufacturer recommendations.

DNA extraction

A small minced biopsy was treated with Proteinase K (GoldBio, Gold Biotechnology, USA) at 37°C overnight and DNA was extracted using a traditional phenol chloroform procedure. DNA was quantitated with spectrophotometer and 1µl was run on an agarose gel for quality check.

Histology

Tumor biopsies were reduced into 1 cubic centimeter pieces and fixed 3 days in a 4% formaldehyde in PBS (Phosphate-Buffered Saline) solution. Samples for histological analysis were prepared using traditional techniques. Briefly, samples were washed of the fixative by rinsing them in PBS two times for 10 minutes with gentle rolling. The dehydration phase was performed with sequential incubation in 25, 50 70, 95 and 100% ethanol solutions. Two hours of Xylene incubation was used to clarify the samples before the specimens were paraffin embedded. Paraffin blocks were then sectioned at 5-8 μm thickness with a microtome, mounted on slides, and aged overnight at 40 °C. The next day, the slides were washed in Xylene to remove the paraffin and rehydrated to water through an ethanol series (100%, 95%, 70%, 50%, 25%, H2O). The samples were H&E stained and cover glass was mounted to preserve the integrity of the samples following standard protocols.

All assessments of hematoxylin and eosin stained sections were performed with the pathologist blinded to genotype of the tumor. For human correlate pathology, we used a previously collected and H&E stained human CCS tissue microarray, which included doubled 3 mm cores from 20 cases of molecularly confirmed CCS.

RNA Sequencing

RNA sequencing libraries were prepared with the RiboZero method, then sequenced either using an Illumina HiSeq instrument as previously described (5) or on a NovaSeq instrument.

***EWSR1-ATF1* expression level calculation for human correlates**

Total RNA was isolated from three human tumor formalin-fixed and paraffin-embedded scrolls as well as from a frozen aliquot of the human SU-CCS-1 cell line (obtained from ATCC). Isolation of RNA from the scrolls utilized the Recoverall Total Nucleic Acid Isolation kit (ThermoFisher Scientific, USA). SuperScript IV First-Strand Reverse Transcription kit

(ThermoFisher Scientific, USA) was used to generate cDNAs, which were subjected to quantitative polymerase chain reaction (RT-qPCR) with primers specific for the fusion (Forward 5'-CATGAGCAGAGGTGGGCG, Reverse 5'-CCCCGTGTATCTTCAGAAGATAAGTC) and for the housekeeping gene beta-2-microglobulin (Forward 5'-TGACTTTGTCACAGCCCAAGATA, Reverse 5'-AATCCAAATGCGGCATCTTC). The mean RPM level for mouse *B2m* was used from the mouse tumors RNAseq, along with the fold-expression of the fusion in human tumors and cell line to present reference levels in Figure 4A.

Low read-depth whole genome sequencing

Genomic DNA was isolated and libraries generated using Nextera DNA flex library prep with UDI. Each library was sequenced on a Novaseq instrument to a level of approximately 40M reads, in order to generate approximately 2-3X depth coverage of the whole genome. Data were aligned on the mm10 genome and run through CNVkit with the CBS method and a bin size of 100 kb.

Statistics

All data are presented as individual points with means also represented. Each statistical test utilized is noted in the figure legend. Log-rank tests were performed with the evanmiller.org calculator. Sample sizes are either noted explicitly in the figure legend, clearly discernible by the dot-plot data presented, or both.

Study Approval

All of the animal work was performed according to ethical standards and under the approval of the University of Utah Institutional Animal Care and Use Committee. All the human

samples utilized were obtained from consenting patients with Institutional Review Board approval at the M.D. Anderson Cancer Center.

Raw data availability

Sequencing was uploaded to the Gene Expression Omnibus at GSE158726.

Author Contributions

The work was conceived of and supported by EP, KS, AB, SWAT, CCM, LD, AJL, MRC and KBJ. Human patient samples were collected by AJL. Mouse genetic experiments with the *Rosa26* model were performed by KS, JJB, and YW. The translocation mouse was designed and generated by EP, KS, and MRC. Experiments with the translocation mouse were performed by EP, AB, SWAT, MRC and KBJ. Mouse exome data were analyzed by KS, MX, LD, KBJ. Human CNA data were analyzed by BO, CCM, KBJ. Transcriptomic data were analyzed by KS and LL. All authors reviewed, edited, and approved the final manuscript.

Acknowledgments

Brian Dalley from the High-Throughput Genomics core facility assisted with transcriptome and genome sequencing and Chris Stubben and Tim Parnell from the Bioinformatics Shared Resource with analysis. BioDiscovery kindly provided software and analysis support for the OncoScan array hybridizations. All animal imaging was performed at the Preclinical Imaging Core Facility.

The work was supported by the Cure Childhood Cancer Foundation, the Pablove Foundation, and Sara's Cure (KBJ), as well as R01 MH093595 (MRC) and the Halt Cancer at X

Foundation (SWAT). CCM was supported by the Pediatric Cancer Program which is funded by the Intermountain Healthcare and Primary Children's Hospital Foundations and the Department of Pediatrics at the University of Utah.

References

1. Hanahan D, Wagner EF, and Palmiter RD. The origins of oncomice: a history of the first transgenic mice genetically engineered to develop cancer. *Genes Dev.* 2007;21(18):2258-70.

2. Goodwin ML, Jin H, Straessler K, Smith-Fry K, Zhu JF, Monument MJ, Grossmann A, Randall RL, Capecchi MR, and Jones KB. Modeling alveolar soft part sarcomagenesis in the mouse: a role for lactate in the tumor microenvironment. *Cancer Cell.* 2014;26(6):851-62.

3. Haldar M, Hancock JD, Coffin CM, Lessnick SL, and Capecchi MR. A conditional mouse model of synovial sarcoma: insights into a myogenic origin. *Cancer Cell.* 2007;11(4):375-88.

4. Jones KB, Barrott JJ, Xie M, Haldar M, Jin H, Zhu JF, Monument MJ, Mosbruger TL, Langer EM, Randall RL, et al. The impact of chromosomal translocation locus and fusion oncogene coding sequence in synovial sarcomagenesis. *Oncogene.* 2016;35(38):5021-32.

5. Straessler KM, Jones KB, Hu H, Jin H, van de Rijn M, and Capecchi MR. Modeling clear cell sarcomagenesis in the mouse: cell of origin differentiation state impacts tumor characteristics. *Cancer Cell.* 2013;23(2):215-27.

6. Keller C, Arenkiel BR, Coffin CM, El-Bardeesy N, DePinho RA, and Capecchi MR. Alveolar rhabdomyosarcomas in conditional Pax3:Fkhr mice: cooperativity of Ink4a/ARF and Trp53 loss of function. *Genes Dev.* 2004;18(21):2614-26.

7. Buchholz F, Refaeli Y, Trumpp A, and Bishop JM. Inducible chromosomal translocation of AML1 and ETO genes through Cre/loxP-mediated recombination in the mouse. *EMBO Rep.* 2000;1(2):133-9.

8. Collins EC, Pannell R, Simpson EM, Forster A, and Rabbitts TH. Inter-chromosomal recombination of Mll and Af9 genes mediated by cre-loxP in mouse development. *EMBO Rep.* 2000;1(2):127-32.

9. Forster A, Pannell R, Drynan LF, McCormack M, Collins EC, Daser A, and Rabbitts TH. Engineering de novo reciprocal chromosomal translocations associated with Mll to replicate primary events of human cancer. *Cancer Cell.* 2003;3(5):449-58.

10. Tanaka M, Yamaguchi S, Yamazaki Y, Kinoshita H, Kuwahara K, Nakao K, Jay PY, Noda T, and Nakamura T. Somatic chromosomal translocation between Ewsr1 and Fli1 loci leads to dilated cardiomyopathy in a mouse model. *Sci Rep.* 2015;5(7826.

11. Yamada K, Ohno T, Aoki H, Semi K, Watanabe A, Moritake H, Shiozawa S, Kunisada T, Kobayashi Y, Toguchida J, et al. EWS/ATF1 expression induces sarcomas from neural crest-derived cells in mice. *J Clin Invest.* 2013;123(2):600-10.

12. Brohl AS, Solomon DA, Chang W, Wang J, Song Y, Sindiri S, Patidar R, Hurd L, Chen L, Shern JF, et al. The genomic landscape of the Ewing Sarcoma family of tumors reveals recurrent STAG2 mutation. *PLoS Genet.* 2014;10(7):e1004475.

13. Bulbul A, Shen JP, Xiu J, Tamayo P, and Husain H. Genomic and Proteomic Alterations in Desmoplastic Small Round Blue-Cell Tumors. *JCO Precis Oncol.* 2018;2(

14. Cancer Genome Atlas Research Network. Electronic address edsc, and Cancer Genome Atlas Research N. Comprehensive and Integrated Genomic Characterization of Adult Soft Tissue Sarcomas. *Cell.* 2017;171(4):950-65 e28.

15. Chow WA, Yee JK, Tsark W, Wu X, Qin H, Guan M, Ross JS, Ali SM, and Millis SZ. Recurrent secondary genomic alterations in desmoplastic small round cell tumors. *BMC Med Genet.* 2020;21(1):101.

16. Crompton BD, Stewart C, Taylor-Weiner A, Alexe G, Kurek KC, Calicchio ML, Kiezun A, Carter SL, Shukla SA, Mehta SS, et al. The genomic landscape of pediatric Ewing sarcoma. *Cancer Discov.* 2014;4(11):1326-41.

17. Lagarde P, Przybyl J, Brulard C, Perot G, Pierron G, Delattre O, Sciot R, Wozniak A, Schoffski P, Terrier P, et al. Chromosome instability accounts for reverse metastatic outcomes of pediatric and adult synovial sarcomas. *J Clin Oncol.* 2013;31(5):608-15.

18. Orbach D, Mosseri V, Pissaloux D, Pierron G, Brennan B, Ferrari A, Chibon F, Bisogno G, De Salvo GL, Chakiba C, et al. Genomic complexity in pediatric synovial sarcomas (Synobio study): the European pediatric soft tissue sarcoma group (EpSSG) experience. *Cancer Med.* 2018;7(4):1384-93.

19. Vlenterie M, Hillebrandt-Roeffen MH, Flucke UE, Groenen PJ, Tops BB, Kamping EJ, Pfundt R, de Bruijn DR, Geurts van Kessel AH, van Krieken HJ, et al. Next generation sequencing in synovial sarcoma reveals novel gene mutations. *Oncotarget.* 2015;6(33):34680-90.

20. Taylor BS, Barretina J, Maki RG, Antonescu CR, Singer S, and Ladanyi M. Advances in sarcoma genomics and new therapeutic targets. *Nat Rev Cancer.* 2011;11(8):541-57.

21. Tirode F, Surdez D, Ma X, Parker M, Le Deley MC, Bahrami A, Zhang Z, Lapouble E, Grossetete-Lalami S, Rusch M, et al. Genomic landscape of Ewing sarcoma defines an aggressive subtype with co-association of STAG2 and TP53 mutations. *Cancer Discov.* 2014;4(11):1342-53.

22. Jahromi MS, Putnam AR, Druzgal C, Wright J, Spraker-Perlman H, Kinsey M, Zhou H, Boucher KM, Randall RL, Jones KB, et al. Molecular inversion probe analysis detects novel copy number alterations in Ewing sarcoma. *Cancer Genet.* 2012;205(7-8):391-404.

23. Li KK, Goodall J, Goding CR, Liao SK, Wang CH, Lin YC, Hiraga H, Nojima T, Nagashima K, Schaefer KL, et al. The melanocyte inducing factor MITF is stably expressed in cell lines from human clear cell sarcoma. *Br J Cancer.* 2003;89(6):1072-8.

24. Davis IJ, Kim JJ, Ozsolak F, Widlund HR, Rozenblatt-Rosen O, Granter SR, Du J, Fletcher JA, Denny CT, Lessnick SL, et al. Oncogenic MITF dysregulation in clear cell sarcoma: defining the MiT family of human cancers. *Cancer Cell.* 2006;9(6):473-84.

25. Steingrimsson E, Moore KJ, Lamoreux ML, Ferre-D'Amare AR, Burley SK, Zimring DC, Skow LC, Hodgkinson CA, Arnheiter H, Copeland NG, et al. Molecular basis of mouse microphthalmia (mi) mutations helps explain their developmental and phenotypic consequences. *Nat Genet.* 1994;8(3):256-63.

26. Komura S, Ito K, Ohta S, Ukai T, Kabata M, Itakura F, Semi K, Matsuda Y, Hashimoto K, Shibata H, et al. Cell-type dependent enhancer binding of the EWS/ATF1 fusion gene in clear cell sarcomas. *Nat Commun.* 2019;10(1):3999.

27. Lagutina IV, Valentine V, Picchione F, Harwood F, Valentine MB, Villarejo-Balcells B, Carvajal JJ, and Grosveld GC. Modeling of the human alveolar rhabdomyosarcoma Pax3-Foxo1 chromosome translocation in mouse myoblasts using CRISPR-Cas9 nuclease. *PLoS Genet.* 2015;11(2):e1004951.

28. Maddalo D, Manchado E, Concepcion CP, Bonetti C, Vidigal JA, Han YC, Ogrodowski P, Crippa A, Rekhtman N, de Stanchina E, et al. In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system. *Nature.* 2014;516(7531):423-7.

29. Kolishovski G, Lamoureux A, Hale P, Richardson JE, Recla JM, Adesanya O, Simons A, Kunde-Ramamoorthy G, and Bult CJ. The JAX Synteny Browser for mouse-human comparative genomics. *Mamm Genome.* 2019;30(11-12):353-61.

30. Hallor KH, Micci F, Meis-Kindblom JM, Kindblom LG, Bacchini P, Mandahl N, Mertens F, and Panagopoulos I. Fusion genes in angiomatoid fibrous histiocytoma. *Cancer Lett.* 2007;251(1):158-63.

31. Schaub FX, Dhankani V, Berger AC, Trivedi M, Richardson AB, Shaw R, Zhao W, Zhang X, Ventura A, Liu Y, et al. Pan-cancer Alterations of the MYC Oncogene and Its Proximal Network across the Cancer Genome Atlas. *Cell Syst.* 2018;6(3):282-300 e2.

32. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR, and Ma'ayan A. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics.* 2013;14(128.

33. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 2016;44(W1):W90-7.

34. Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, and Vilo J. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* 2019;47(W1):W191-W8.

35. Tang SH, Silva FJ, Tsark WM, and Mann JR. A Cre/loxP-deleter transgenic line in mouse strain 129S1/SvImJ. *Genesis.* 2002;32(3):199-202.

36. Logan M, Martin JF, Nagy A, Lobe C, Olson EN, and Tabin CJ. Expression of Cre Recombinase in the developing mouse limb bud driven by a Prxl enhancer. *Genesis.* 2002;33(2):77-80.

37. Liu P, Jenkins NA, and Copeland NG. A highly efficient recombineering-based method for generating conditional knockout mutations. *Genome Res.* 2003;13(3):476-84.

38. Wu S, Ying G, Wu Q, and Capecchi MR. A protocol for constructing gene targeting vectors: generating knockout mice for the cadherin family and beyond. *Nat Protoc.* 2008;3(6):1056-76.

Figures

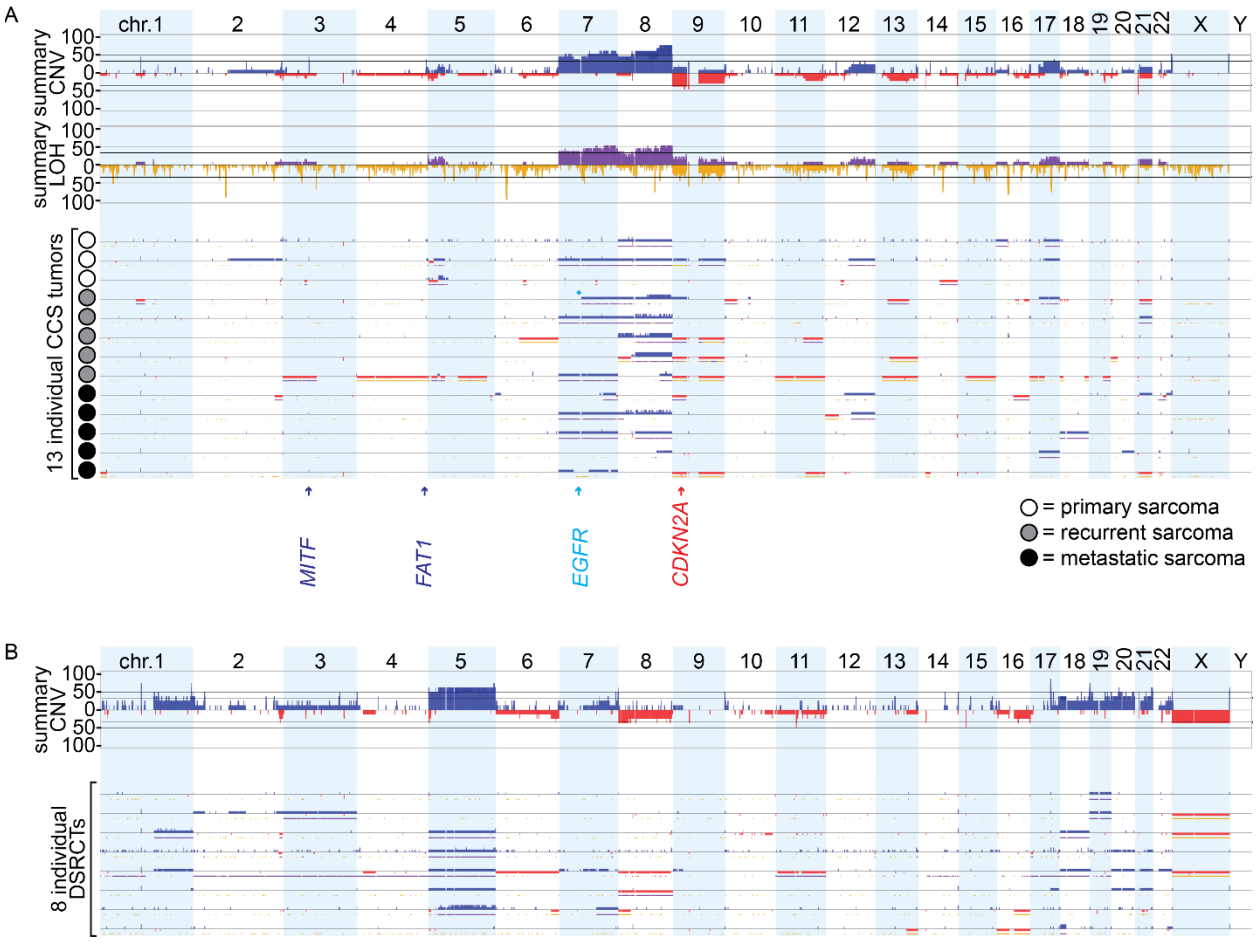

Figure 1. Human clear cell sarcomas demonstrate a few repeated copy number aberrations across the genome.

A. The frequency of copy number gains (blue) or losses (red), loss of heterozygosity (yellow), and point mutant (cyan diamond), as called by BioDiscovery Nexus Express is shown for Clear Cell Sarcoma (n = 13).

B. The frequency of copy number gains (blue) or losses (red) is shown for Desmoplastic Small Round Cell Tumor (n = 8), analyzed in parallel to CCS for accurate comparison.

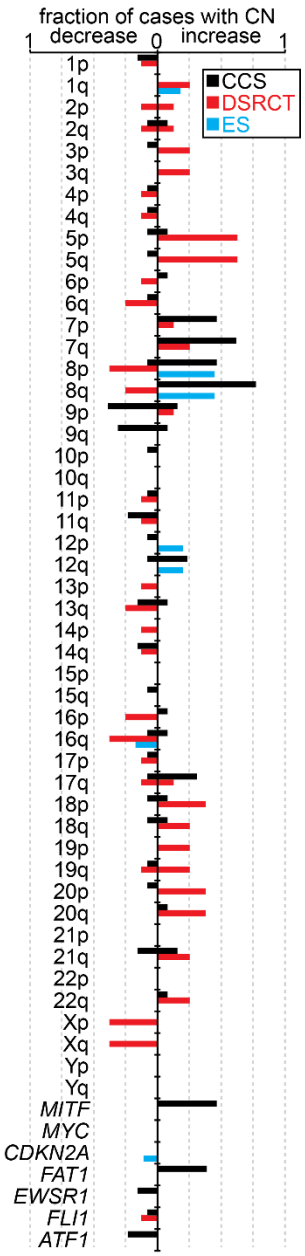

Figure 2. Clear cell sarcoma shares few secondary genomic alterations with other *EWSR1*-rearranged cancers.

Summary of copy number alterations by chromosome arm and genes of interest, displayed as the fraction (0-1) for each sarcoma subtype, CCS (n = 13), DSRCT (n = 8), and Ewing Sarcoma (n = 112). Prevalence of gains and losses for CCS and DSRCT reflects heterozygous or homozygous loss or copy number gains present with at least 10% estimated clonality.

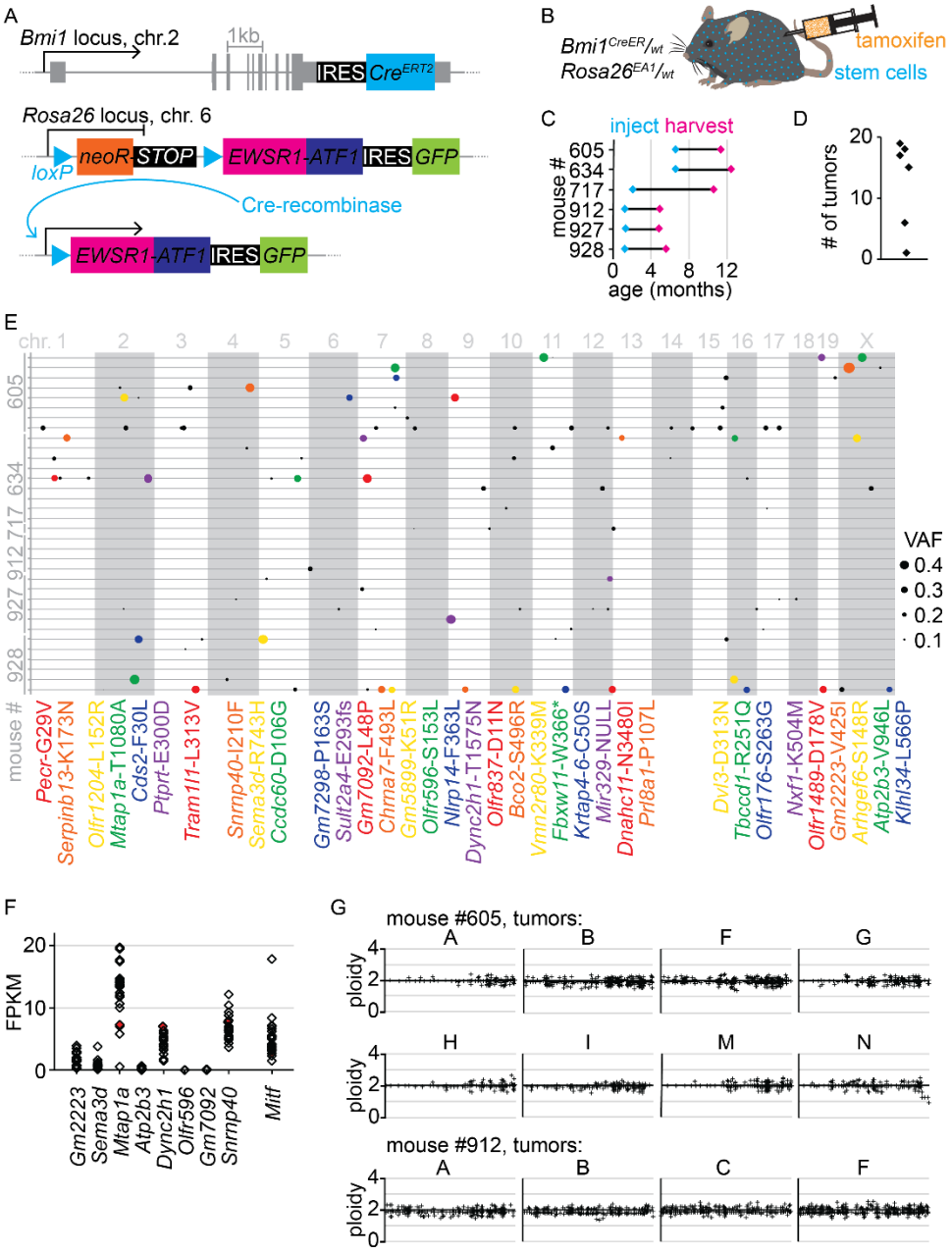

Figure 3. Exome sequencing of *EWSR1-ATF1*-expression-initiated mouse tumors reveals that no secondary alterations are strictly required to complete sarcomagenesis.

A. Schematic illustrating *Cre*^{ERT2} engineered at the *Bmi1* locus in conjunction with Cre-inducible human-*EWSR1-ATF1* engineered at the *Rosa26* locus. Cre-recombinase removes the neoR-

STOP cassette to induce expression of *EWSR1-ATF1*, and the GFP reporter confirms *EWSR1-ATF1* expression.

B. Tamoxifen injection induces *Cre*^{ERT2} activity in *Bmi1*-expressing stem cells (in blue), and then in turn, Cre-recombinase induces *Rosa26*-mediated expression of *EWSR1-ATF1* (*EA1*).

C. Time course showing age at injection and age at tumor harvest (n = 6).

D. Summary of total numbers of tumors detected and harvested per mouse (n = 6).

E. Exome sequencing of 34 mouse tumors, with one tumor's exome presented on each line, clustered by host mouse, and each variant allele denoted by a circle with size corresponding to variant allele frequency (VAF). All variants with VAF > 0.25 are also identified by corresponding colors in the list of gene symbols and protein amino acid substitutions below.

F. RNA sequencing rendered expression levels relative to *Gapdh* for the 8 genes with variants present at fractions higher than 0.4. *Mitf* expression level is included for reference.

G. Exome-wide copy number variation is inferred from exome sequencing data, as shown for 12 mouse tumors. Normal diploidy is represented by the x-axis crossing at a value of y = 2.

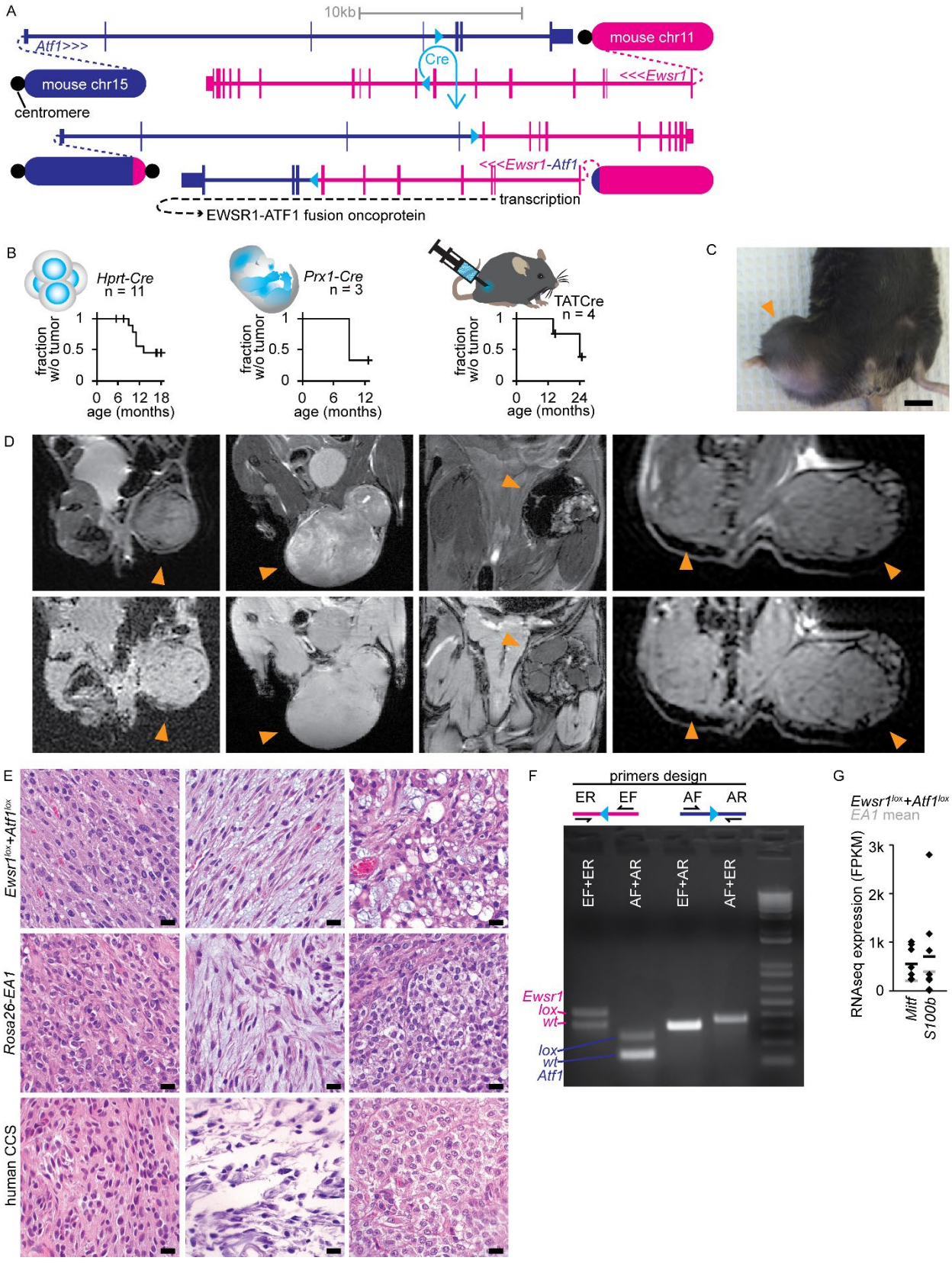
A
10kb
mouse chr11
Atf1>>>
Cre
<<<Ewsr1
mouse chr15
centromere
<<<Ewsr1-Atf1
transcription
EWSR1-ATF1 fusion oncoprotein
B
Hprt-Cre
n = 11
Prx1-Cre
n = 3
TATCre
n = 4
fraction w/o tumor
age (months)
C
D
E
Ewsr1lox+Atf1lox
Rosa26-EA1
human CCS
F
primers design
ER
EF
AF
AR
EF+ER
AF+AR
EF+AR
AF+ER
Ewsr1
lox
wt
Atf1
G
Ewsr1lox+Atf1lox
EA1 mean
RNAseq expression (FPKM)
Mitf
S100b

Figure 4. Cre-mediated chromosomal translocation induces sarcomagenesis in the mouse.

A. Schematic representing the *loxP* sites targeted to *Ewsr1* intron 7 and *Atf1* intron 4, as well as the 2 products of Cre-mediated chromosomal translocation.

B. Schematics and Kaplan-Meier survival plots of tumorigenesis in mice heterozygous for *Ewsr1-loxP* and *Atf1-loxP*, induced by three different Cre-recombinase delivery methods, a knock-in allele (*HprtCre*), a transgenic allele (*Prx1Cre*), and an injection of Cre-recombinase protein (TATCre).

C. Gross photo of a hindlimb tumor (arrow) forming in a *Prx1Cre*-induced mouse (scale bar = 10mm).

D. Magnetic resonance images of tumors (indicated by arrows) forming in the thigh, dorsal pelvis, thigh, and both pelvis and contralateral thigh of mice induced by *HprtCre* or TATCre (upper panels, T2-weighted; lower panels, proton-density-weighted).

E. Representative H&E photomicrographs of histological sections of tumors from translocation (above), *Rosa26-EA1* expression (middle), and human CCS tumors (below), with short spindle cell, myxoid, and clear cell morphologies (left to right; scale bars = 10um).

F. Schematic of PCR amplification strategy with schematic of primers that amplify across each translocation site, tested in genomic tumor DNA from a TATCre-induced tumor, with wildtype and *lox* alleles as well as each translocation product detected by specified primer combinations.

G. Graph of the RNAseq-determined expression in fragments per kilobase per million reads (FPKM) of two melanocytic marker genes, with 8 *HprtCre*-induced translocation tumors and their mean noted in black and the mean of 13 EA1-induced tumors in grey.

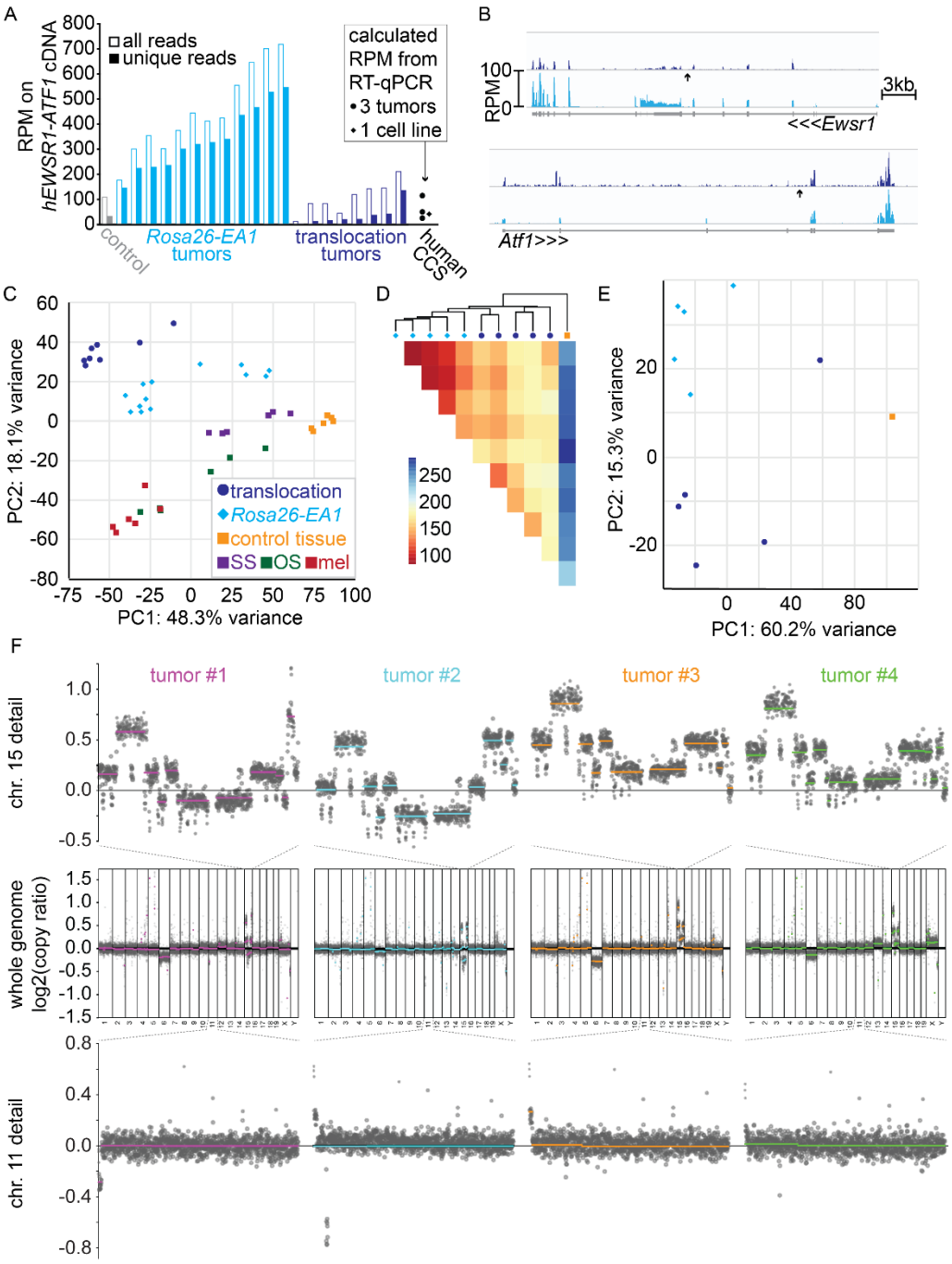

Figure 5. Cre-mediated translocation model of clear cell sarcomagenesis mimics the transcriptome of *EWSR1-ATF1*-expression model, but has additional genome copy number alterations.

A. Graph of the reads per million mapped reads (RPM) that aligned onto the human *EWSR1-ATF1* fusion oncogene cDNA coding sequence expressed conditionally from the *Rosa26* locus

and from the translocation (*HprtCre*-induced). For reference, also noted are the calculated RPM levels of *EWSR1-ATF1* expression by RT-qPCR from the human SU-CCS1 cell line and 3 formalin-fixed and paraffin-embedded human CCS tumor specimens, using beta-2-microglobulin as a control in the human samples and its average expression across the mouse samples to calculate an RPM estimate.

B. RPM alignments across the genomic sequence for *Ewsr1* and *Atf1*, averaged across 8 translocation generated tumors (blue) and 12 *Rosa26-EA1* tumors (cyan), demonstrating overall lower expression (compared to cDNA expression in the EA1 tumors) and a reduced 3' bias in the exons 3' to the translocation point (arrow) of *Ewsr1* relative to the same in *Atf1* in the translocation-generated tumors, which may represent reduced expression of the exons 3' to the translocation in *Ewsr1* or increased expression of the exons 3' to the translocation in *Atf1*.

C. Principal component (PC) analysis demonstrating relative clustering of 8 translocation-driven sarcomas with *Rosa26-EA1*-driven comparators, separate from the clusterings of other mouse cancer subtypes sequenced in the same batch as the EA1 tumors.

D. Heatmap of the Pearson correlation distance between single batch sequenced transcriptomes of the *Rosa26-EA1*-driven (cyan diamonds) or *HprtCre*-induced translocation-driven (blue circles) sarcomas, as well as one control tissue sample.

E. PC analysis of the same transcriptomes as in (D).

F. Copy number alteration (CNA) analysis on 4 *HprtCre*-induced translocation tumors, using a low-read-depth whole genome sequencing approach. The upper row shows higher resolution images of chromosome 15 from each tumor's sequencing, the middle row CNA across the entire genome and lower row chromosome 11 from each.

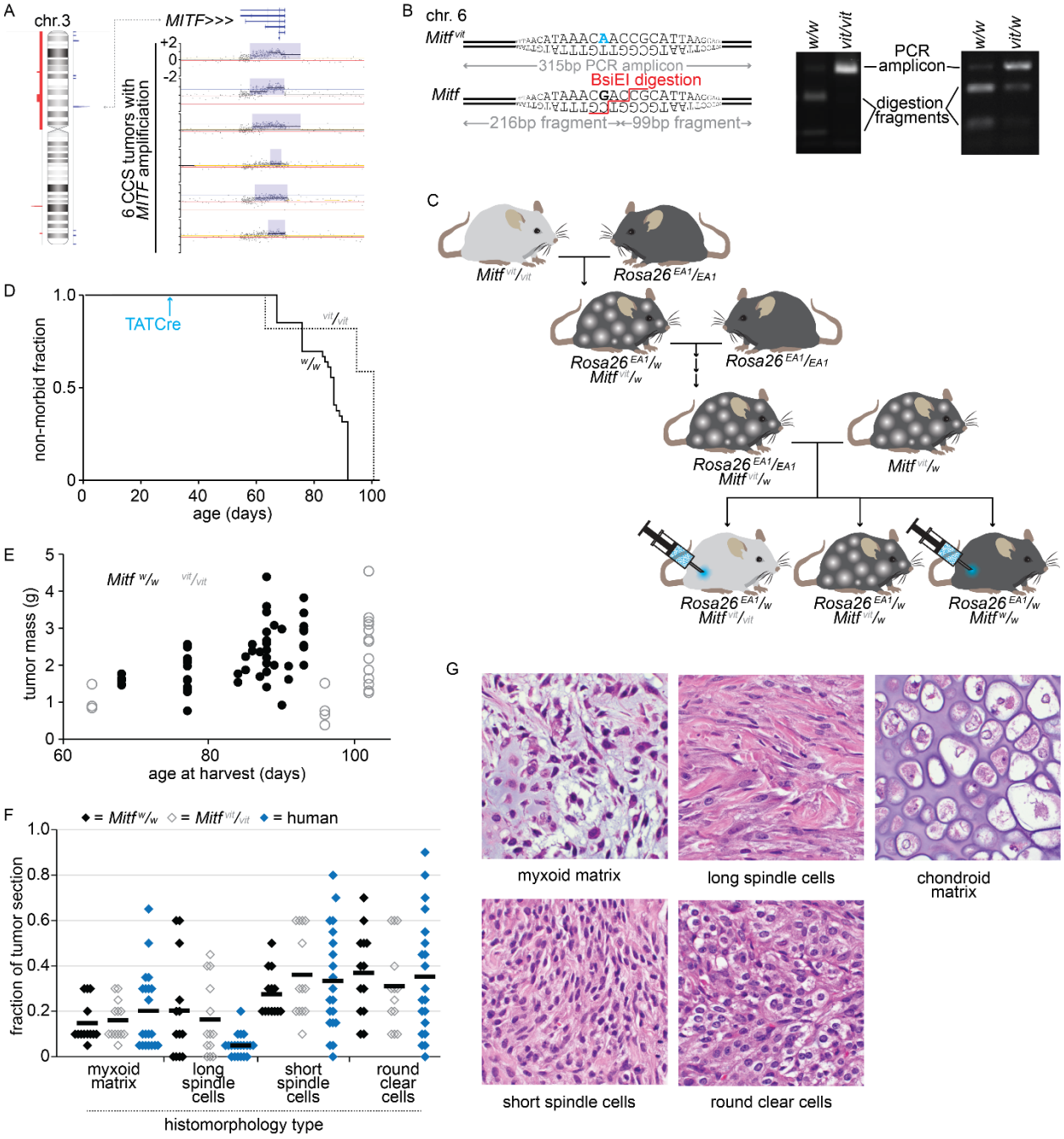

Figure 6. MITF contributes to oncogenesis driven by *EWSR1-ATF1*

A. Left, human chromosome 3 is shown with the spike in amplification (in blue) occurring in 6/13 human CCS tumors. Right, copy number microarray data showing the recurrent, focal amplification occurs at the *MITF* locus in all 6 tumors.

B. Top, mouse DNA sequences depicting the *Mitf* vitiligo (*Mitf* ^{vit}) mutation (in blue) and the normal *Mitf* sequence. BsiEI restriction enzyme is used to confirm *Mitf* ^{vit} after PCR (below) because it recognizes and cuts wildtype *Mitf* but not *Mitf* ^{vit}.

C. Mouse breeding schematic showing strategy to generate EA1 heterozygotes with homozygous mutant or homozygous wildtype *Mitf* as littermates, which we injected with TATCre to induce *EWS-ATF1* expression.

D. Kaplan-Meier curve shows achievement of morbidity of *Mitf* $^{w/w}$ mice (solid line) and *Mitf* $^{vit/vit}$ mice (dotted line) following TATCre injection at 28 days old. (n = 37 and 15). The 50% median survival time to morbidity for *Mitf* $^{vit/vit}$ compared to wildtype was 102 days compared to 87 days. A log-rank test was performed and the difference was deemed statistically significant (p value < 0.0001, z score value of 4.52).

E. Tumor mass measurements are depicted for *Mitf* $^{w/w}$ mice (black dots) and *Mitf* $^{vit/vit}$ mice (gray circles). The tumors were not significantly different in size (t-test, 2-tailed, p = 0.6).

F. Graph of the blinded quantitation of histological features in mice with tumors developing from *EA1* expression in either *Mitf* $^{w/w}$ mice (black diamonds, n = 37) or *Mitf* $^{vit/vit}$ mice (gray diamonds, n = 15), with human CCSs on a tissue microarray (n = 20, blue diamonds). By two-tailed Student's t-test, with $\alpha = 0.05$, none of these comparisons reached statistical significance between the two mouse groups).

G. Photomicrographs demonstrating example histomorphologies in hematoxylin and eosin stained tissue sections of *EA1*-expressing tumors. (Each photomicrograph is a 100μm square.)

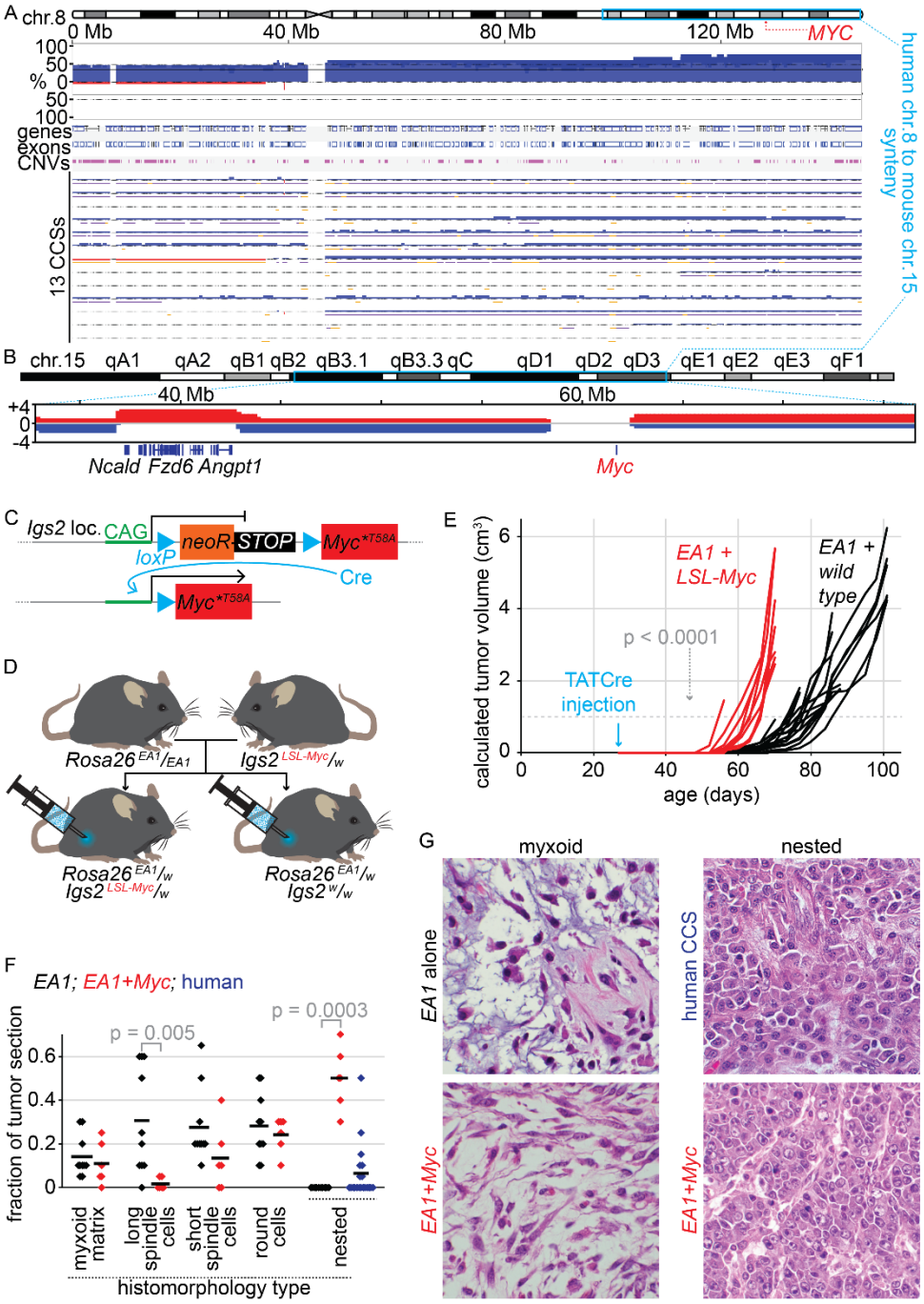

Figure 7. MYC stabilization enhances sarcomagenesis from *EWSR1-ATF1*, but alters tumor phenotypes.

A. BioDiscovery data regarding copy number alterations in regions of human chromosome 8 that are syntenic to amplified regions of mouse chromosome 15 and that flank *MYC*

B. Copy number alterations in regions of mouse chromosome 15 syntenic to human chromosome 8 and surrounding *Myc*.

C. Schematic showing the floxed stop cassette which allows for Cre-inducible expression of Myc^{T58A} at the *Igs2* locus.

D. Breeding strategy to generate mice that express both *EWSR1-ATF1* (*EA1*) and Myc^{T58A} upon TATCre injection, as well as $Igs2^{w/w}$ littermate control.

E. Tumor growth curves for $Rosa26^{EA1}/_{WT}$ mice with either $Igs2^{LSL\text{-}Myc/w}$ (red) or $Igs2^{w/w}$ (black), following TATCre injection at 28 days.

F. Graph of the blinded quantitation of hematoxylin and eosin stained slides for histologic features distinguishing *EA1*-only and *EA1*+Myc^{T58A}-expressing tumors (red), with the prevalence of the nested morphology for reference observed in human CCSs on a tissue microarray (n = 20).

G. Photomicrographs of hematoxylin and eosin stained tissue sections demonstrating the two variants on myxoid features, one in each genotype, as well as the nested histomorphology observed focally in 7 human tumors (from n = 20) and all *Myc*-activating mouse tumors (each photomicrograph is a 100μm square).