The Genomic Landscape of Pediatric Ewing Sarcoma

Brian D. Crompton1, Chip Stewart2, Amaro Taylor-Weiner2, Gabriela Alexe1,2,3, Kyle C. Kurek4, Monica L. Calicchio4, Adam Kiezun2, Scott L. Carter2, Sachet A. Shukla5, Swapnil S. Mehta1, Aaron R. Thorner1, Carmen de Torres6, Cinzia Lavarino6, Mariona Suñol6, Aaron McKenna2, Andrey Sivachenko2, Kristian Cibulskis2, Michael S. Lawrence2, Petar Stojanov2,7, Mara Rosenberg2, Lauren Ambrogio2, Daniel Auclair2, Sara Seepo4, Brendan Blumenstiel2, Matthew DeFelice2, Ivan Imaz-Rosshandler8, Angela Schwarz-Cruz y Celis8, Miguel N. Rivera2,9, Carlos Rodriguez-Galindo1, Mark D. Fleming4, Todd R. Golub1,2, Gad Getz2,9,10, Jaume Mora5, and Kimberly Stegmaier1,2
ABSTRACT

Pediatric Ewing sarcoma is characterized by the expression of chimeric fusions of EWS and ETS family transcription factors, representing a paradigm for studying cancers driven by transcription factor rearrangements. In this study, we describe the somatic landscape of pediatric Ewing sarcoma. These tumors are among the most genetically normal cancers characterized to date, with only EWS–ETS rearrangements identified in the majority of tumors. STAG2 loss, however, is present in more than 15% of Ewing sarcoma tumors; occurs by point mutation, rearrangement, and likely nongenetic mechanisms; and is associated with disease dissemination. Perhaps the most striking finding is the paucity of mutations in immediately targetable signal transduction pathways, highlighting the need for new therapeutic approaches to target EWS–ETS fusions in this disease.

SIGNIFICANCE: We performed next-generation sequencing of Ewing sarcoma, a pediatric cancer involving bone, characterized by expression of EWS–ETS fusions. We found remarkably few mutations. However, we discovered that loss of STAG2 expression occurs in 15% of tumors and is associated with metastatic disease, suggesting a potential genetic vulnerability in Ewing sarcoma. Cancer Discov. 4(11): 1326–41. © 2014 AACR.

INTRODUCTION

The sequencing of cancer genomes in aggressive pediatric solid tumors, such as rhabdoid tumors and retinoblastoma, has revealed remarkably stable genomes (1, 2). In both cases, oncogenesis is thought, at least in part, to be driven by epigenetic deregulation. It has been hypothesized that pediatric solid tumors notable for recurrent translocations involving transcription factors will also have relatively stable genomes, as has already been observed in acute myeloid leukemia and rhabdomyosarcoma with recurrent cytogenetic translocations (3, 4). Studies to systematically address this hypothesis are just beginning to emerge. Ewing sarcoma is an aggressive, poorly differentiated tumor, typically arising from bone, and is characterized by highly recurrent translocations involving ETS transcription factors, with EWS–FLI and EWS–ERG translocations being the most common (5, 6). Ewing sarcoma cells have a strict dependency on EWS–ETS fusion proteins, supporting the notion that these translocations are a critical oncogenic event in this disease (7). Although the EWS–FLI rearrangement was described over 20 years ago, effective therapy directly targeting this transcription factor abnormality has yet to be developed. New therapeutic strategies for this disease are needed because the vast majority of children with metastatic or relapsed Ewing sarcoma will die despite intensive, multimodality therapy (8, 9). Moreover, even for children who are cured, the long-term morbidity of cytotoxic treatment is significant (10).

Efforts to discover collaborating, and potentially targetable, genetic alterations in Ewing sarcoma tumors have primarily focused on the study of copy-number alterations, with several studies reporting copy-number gains of chromosome 8 in nearly 50% of tumors (11). However, with the exception of focal deletions at the CDKN2A locus on 9p21, specific genes of relevance have yet to be elucidated from these studies (12, 13). Furthermore, few recurrent somatic mutations have been reported in Ewing sarcoma, with the notable exceptions of TP53 (5%–20%) and more recently the cohesin complex family member STAG2 (20%), with neither of these imminently druggable (13–16). We sought to define the genomic landscape of pediatric Ewing sarcoma by integrating multiple next-generation sequencing methods applied to a large panel of human Ewing sarcoma tumors.

RESULTS

Massively Parallel Sequencing of Ewing Sarcoma

We collected 116 pediatric Ewing sarcoma tumors from 112 patients and extracted high-quality DNA from 96 tumors for sequencing, including 4 patients with multiple tumors. Clinical data and demographics were available for a subset of these patients, which were consistent with epidemiologic studies of pediatric Ewing sarcoma (Table 1 and Supplementary Table S1A–S1D; ref. 17). We performed whole-exome sequencing (WES) of tumor–normal pairs from 26 patients,
tumors from 66 patients, and 11 cell lines. We also performed whole-genome sequencing (WGS), genotyping array (SNP array), and transcriptome sequencing (RNASeq) on a subset of these samples (Fig. 1A; Supplementary Fig. S1; Supplementary Table S1A–S1D). We found only 12 genes with non-silent mutations in more than one sample of the 26 tumors (Fig. 1A; Supplementary Fig. S1; Supplementary Table S3; Supplementary Fig. S2). We also identified 10 genes with mutation sites listed in the Catalogue of Somatic Mutations in Cancer (COSMIC; v51), but of these only TP53 was mutated in more than one sample (Supplementary Table S2A–S2D; ref. 18). We used the MutSig 2CV algorithm to detect significantly mutated genes from 26 tumors with normal pairs and found that only STAG2 was significantly mutated (Supplementary Table S5; ref. 19). With an overall mutation rate of 0.62 mutations per megabase (Mb), this analysis had 90% power to detect significantly mutated genes present in 30% or more of patients (Supplementary Fig. S3; Supplementary Table S6A; ref. 19).

To increase our power to detect less commonly mutated genes, we developed an approach to analyze WES for Ewing sarcoma tumors (13, 14). We found nonsilent coding mutations in chromatin-modifying genes. Enrichment of mutated genes in gene signatures defined by tyrosine kinase activity or pathways reported to be active in Ewing sarcoma (i.e., GLI1, MAPK, and PI3K–AKT–mTOR pathways) when considering either the 26 tumors with normal pairs or the entire collection of 92 tumors (Supplementary Table S9; refs. 24–32). Given the lack of highly recurrent mutations, it is also notable that there was no enrichment of mutations in chromatin-modifying genes.

Finally, somatic copy-number alterations (SCNA) were identified from SNP-array data and by read-depth analysis of WES data. GISTIC2.0 analysis identified significant copy-number gains of chromosome 8 in 45% (41 of 92) of cases and focal deletions at 9p21.3 in 28% of tumors (5 of 18 by SNP array and 22% or 20 of 92 by SNP array plus WES), consistent with loss of CDKN2A (Fig. 1A–C; Supplementary Fig. S4A; Supplementary Table S10; ref. 33). No significant focal amplifications were found. One recent study reported recurrent focal deletions of SMARCB1 and RELN in Ewing sarcoma tumors, but we were unable to identify these aberrancies by SNP array or WES in our cohort (34). As pointed out by the authors of this previously published study, these deletions may be more prevalent in Ewing sarcoma tumors with either atypical features or mixed features of both Ewing sarcoma and rhabdoid.

### Significantly Mutated Genes in Ewing Sarcoma

Mutations in TP53 have been reported in 3% to 14% of Ewing sarcoma tumors (13, 14). We found nonsilent coding aberrancies of TP53 in 6% (6 of 96 samples including the Project (ESP) as well as a panel of more than 2,900 normal exomes to remove germline events and recurrent artifacts, which reduced this rate to 6.1 variants/Mb (Supplementary Tables S6B and S7; refs. 20, 21). After combining these results with the 26 tumor-normal pairs (also now filtered by the panel of normal samples), the entire cohort had an average variant rate of 4.6/Mb. This improved our sensitivity to detect significantly mutated genes (90% power) present in 15% or more of patients (Supplementary Fig. S3; ref. 19). We hypothesized that the remaining rare germline events would be widely distributed across the exomes, allowing MutSig to detect significant somatic events. Variants were analyzed with the MutSig 2CV algorithm, and consistent with the above hypothesis, only four significantly mutated genes were identified (Supplementary Table S8). STAG2 and TP53 were the only two genes known to be cancer-associated. RPTN, a member of the epidermal differentiation complex, and PRR2, a member of the human salivary proline-rich protein family, were also significantly mutated on the basis of this analysis (22, 23). Neither of these genes, however, was expressed in any of the Ewing sarcoma tumors or cell lines based on the RNASeq data, suggesting that these variants are unlikely to be driver events in Ewing sarcoma. These results confirm that there are few significantly mutated genes in Ewing sarcoma other than EWS–ETS rearrangements at frequencies greater than 15%.

We also noted a stark paucity of mutations in genes involved in signaling pathways. In fact, of the 26 tumor-normal pairs, EPHA7 was the only tyrosine kinase recurrently mutated, and there was only 1 patient (SJDES001-R) with a candidate targetable lesion (PTEN mutation coupled with deletion of the other allele). Furthermore, there was no enrichment of mutated genes in gene signatures defined by tyrosine kinase activity or pathways reported to be active in Ewing sarcoma (i.e., GLI1, MAPK, and PI3K–AKT–mTOR pathways) when considering either the 26 tumors with normal pairs or the entire collection of 92 tumors (Supplementary Table S9; refs. 24–32). Given the lack of highly recurrent mutations, it is also notable that there was no enrichment of mutations in chromatin-modifying genes.

### Table 1. Demographics

<table>
<thead>
<tr>
<th>Demographic</th>
<th>Percentage (number/total)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>55% (51/92)</td>
</tr>
<tr>
<td>Mean age (min–max)</td>
<td>11 y (4 mo–21 y)</td>
</tr>
<tr>
<td>White</td>
<td>96% (25/26)</td>
</tr>
<tr>
<td>Black</td>
<td>4% (1/26)</td>
</tr>
<tr>
<td>Axial primary (chest, vertebrate, pelvis)</td>
<td>42% (25/59)</td>
</tr>
<tr>
<td>Tumors with EWS-FLI (vs. EWS-ERG)</td>
<td>85% (22/26)</td>
</tr>
<tr>
<td>Survival (patients sampled at diagnosis)</td>
<td>69% (11/16)</td>
</tr>
<tr>
<td>Survival (patients sampled after treatment failure)</td>
<td>25% (2/8)</td>
</tr>
</tbody>
</table>

**NOTE:** Summary of available clinical data (details are given in Supplementary Table S1).
The Genomics of Ewing Sarcoma

RESEARCH ARTICLE

Figure 1. Mutational landscape of Ewing sarcoma tumors. A, sample data, sequencing method, and recurrent aberrancies detected by massively parallel sequencing are indicated for 92 Ewing sarcoma tumors and 11 Ewing sarcoma cell lines. Box colors for each panel are indicated in the key. SNV, single-nucleotide variant. B, arm-level SCNAS from 92 Ewing sarcoma tumors determined by segmentation analysis of WES. C, focal SCNAS for 18 Ewing sarcoma tumors determined from SNP-array analysis. B and C, plotted are the $-\log_{10}$ values for copy-number losses (blue) and gains (red) for each chromosomal arm or focal region calculated with GISTIC2.0 (33). Gains of chromosome 8 were the most significant arm-level SCNA, and CDKN2A is located at the most significant focal loss. RNASeq data demonstrated that ABCC12 was likely not expressed in any sample regardless of copy-number status. Focal deletions at 11p24.3 and 22q12.2 involve the FLI and EWS genes in the EWS–FLI rearranged samples. D, the locations of variants in STAG2 and TP53 indicated in linear protein domain models. Domain annotations are taken from recent publications (15, 73). Numbers below each protein name indicate the number of amino acids in the full-length protein. Variant details are listed in Supplementary Table S11. Solid boxes, mutations identified in tumors with normal pairs (including multiple tumors from the same patient); open triangles, variants from tumors without normal pairs after germline filtering; and open circles, variants from cell lines after filtering. Mutations: MS, missense; FS, frameshift; NS, nonsense; SS, splice site. Protein domains: STAG, STAG superfamily; SCD, stromalin conservative domain; TADI/II, transcriptional-activation I and II; DNA, DNA binding; OD, tetramerization; CTD, carboxyl terminus.

Tumor trios described below; 95% confidence interval (CI), 2%–13% of Ewing sarcoma samples (Fig. 1A and D and Supplementary Table S11A–S11D). TP53 variants were associated with a greater rate of SCNAS ($P = 0.025$ by Mann–Whitney test) across the genome but no significant change in tumor ploidy (Supplementary Fig. S4B and S4C). None of the 51 tumors acquired at the time of patient diagnosis had a TP53 mutation, whereas 25% of tumors acquired after treatment (3 of 12; 95% CI, 6%–57%) were mutated at TP53 ($P = 0.006$ by Fisher exact test). Furthermore, relapses from two different time points from a single patient (SJDES007-R1 and SJDES007-R2) acquired unique TP53 mutations (p.C176F and p.C135F, respectively). Along with the striking observation that Ewing sarcoma is rarely diagnosed in families with Li-Fraumeni syndrome, these data support the hypothesis that TP53 mutations play a greater role in treatment resistance...
or disease recurrence than in the initiation of Ewing sarcoma, as previously suggested in the literature (13).

Mutations in STAG2 were recently identified in several types of cancer, including Ewing sarcoma (4, 15, 16, 35–37). We identified eight nonsilent single-nucleotide variants (SNV) or indels in STAG2 in 96 tumors, including two distinct STAG2 mutations in separate relapses from the same patient (SJDES007-R1 and SJDES007-R2), as well as a previously unreported STAG2–MAP7D3 fusion in a Ewing sarcoma cell line (Fig. 1A and D and Supplementary Table S11A–S11D). We measured STAG2 expression by IHC in 73 tumors (including an extension cohort of 20 tumors not sequenced) and found that 15% (11 of 73; 95% CI, 8%–25%) of tumors had complete or heterogeneous loss of STAG2 expression. We also demonstrated loss of full-length STAG2 expression in Ewing sarcoma cell lines by IHC and/or Western immunoblotting and confirmed that TTC466 expressed STAG2–MAP7D3 fusion transcripts, resulting in a loss of STAG2 expression (Fig. 2A–C and Supplementary Tables S4D and S12). Interestingly, one tumor and three cell lines were observed to have loss of STAG2 expression without evidence of somatic mutations, suggesting that there are other mechanisms of STAG2 silencing in Ewing sarcoma.

In recent studies, there have been conflicting reports about the association of STAG2 loss with increased aneuploidy in cancer (15, 35, 36, 38). We found that tumors with loss of STAG2 expression had a slightly higher mean proportion of the genome with SCNAs (0.15 for STAG2 positive vs. 0.36 for STAG2 loss; P = 0.009 by Mann–Whitney test) but no difference in ploidy (Fig. 2D and E). However, we also noted that some STAG2 mutations were identified in tumors with TP53 mutations and/or from patients who had previously received treatment, factors known to induce chromosomal instability. In fact, we found that TP53 mutations occurred significantly more often than expected in samples with aberrant STAG2, as determined by IHC or by the identification of a mutation when IHC was unavailable (P = 0.014 by Fisher exact test). When limiting our analysis to tumors that are TP53 wild-type and obtained at diagnosis, we found that there was no significant difference in the rate of SCNAs in tumors with or without loss of STAG2 expression (0.14 for STAG2 positive vs. 0.27 for STAG2 loss; P = 0.108 by Mann–Whitney test; Fig. 2F). Furthermore, we did not observe any copy-number changes in one tumor sample with a clonal STAG2 mutation (SJDES006; Supplementary Fig. S4A). Rather, we observed that 88% of patients (7 out of 8; 95% CI, 47%–100%) with STAG2 loss at the time of diagnosis also presented with clinical evidence of metastatic disease compared with only 27% (14 of 51; 95% CI, 16%–42%) of those whose diagnostic tumor sample expressed STAG2 (P = 0.002 by Fisher exact test), demonstrating that STAG2 loss is associated with disease dissemination in Ewing sarcoma (Fig. 2G).

Furthermore, we also performed single-sample gene set enrichment analysis (ssGSEA) to identify gene signatures that were enriched in samples with loss of STAG2 expression. Because of the small number of tumors with both loss of STAG2 expression and available RNASeq data, we combined RNASeq data from tumors and cell lines for this analysis. Two of the 11 signatures that are enriched in samples with loss of STAG2 expression were gene sets associated with metastasis in cancer (Fig. 2H–J and Supplementary Table S13).

Alterations in ETS Transcription Factors

We identified EWS–ETS rearrangements for all samples with WGS and EWS–ETS fusion transcripts in all samples with RNASeq data (Supplementary Table S4A–S4C). Reciprocal ETS–EWS fusion transcripts were detected in only four samples, each with five or fewer reads. In some cases, RNASeq detected multiple EWS–ETS fusions, a phenomenon previously reported in Ewing sarcoma (39–41). In all samples, at least one fusion variant was expected to result in the translation of the ETS portion of the chimeric protein in-frame. These in-frame variants were always found to have higher expression than out-of-frame variants in these samples, reaffirming that alternative splicing is an effective mechanism for expressing EWS–ETS fusion proteins for rearrangements otherwise predicted to yield out-of-frame and truncated fusion forms (P < 0.0001 by Mann–Whitney test; Supplementary Table S4C). We also found that wild-type FLI and ERG exons were poorly expressed in EWS–ETS–expressing tumors (Fig. 3A–C and Supplementary Table S14A and S14B). These data indicate that EWS–ETS rearrangements are necessary for the expression of FLI and ERG and that alternative splicing of EWS–ETS rearrangements is sometimes necessary for the in-frame translation of these ETS transcription factors in Ewing sarcoma.

Recent work in prostate cancer demonstrated that nonrearranged ETS transcription factor aberrancies play a role in the development of disease even in tumors expressing ETS fusions and that ETS2 repressor factor (ERF) expression has also been reported to attenuate ETS-associated tumorigenesis (42, 43). Therefore, we looked for evidence of other ETS aberrances in our samples. Although mutations in ETS family genes were not statistically enriched among the 92 tumors sequenced by GSEA, we identified three samples with nonsilent coding variants in ERF (3%; 3 of 92; 95% CI, 1%–9%), leading to a relative decrease in the expression of the ERF transcript for the only patient with RNASeq data and an ERF mutation, SJDES012 [reads per kilobase per million (RPKM) = 20.7, wild-type ERF RPKM; range, 23.3–57.6; Supplementary Table S11A–S11D]. We also identified recurrent, focal deletions of the ETS1 gene (11%; 10 out of 92; 95% CI, 5%–19%), originating at the FLI breakpoint, in Ewing sarcoma tumors. None of these samples, however, had RNA available for determining the effect of deletion on ETS1 expression, and we did not identify point mutations in this gene (Supplementary Fig. S5 and Supplementary Table S15).

Ewing Sarcoma Genomics after Treatment and in Established Cell Lines

Of the 26 tumors with available normal controls, 16 patients were sampled at diagnosis and eight sampled after treatment. For this cohort, upfront treatment used the same chemotherapy agents for all patients except one (Supplementary Table S1A–S1D). Diagnostic and treated tumors differed significantly in their mutation rates (0.37 vs. 1.11 mutations per Mb, respectively; P = 0.0017 by Mann–Whitney test; Fig. 4A and B; Supplementary Table S6A). SNVs were predominantly C to T transitions with the majority subcategorized as *CpG to T mutations, a common pattern attributed to spontaneous deamination of 5-methyl-cytosine and associated with cancers with high levels of methylation (44, 45). In treated tumors, there was a relative increase in substitutions of cytosine at CpG.
The Genomics of Ewing Sarcoma

Figure 2. Loss of STAG2 expression in Ewing sarcoma. 

A, IHC staining for STAG2 in Ewing sarcoma cell lines and tumors demonstrates a loss of STAG2 expression in mutated samples (red asterisk). The A673 Ewing cell line, expressing wild-type STAG2, stained positive for STAG2 expression, whereas TC32, mutated at STAG2, does not express STAG2 protein. Also shown are two primary Ewing sarcoma tumor samples stained for STAG2 expression. The tumor with a STAG2 mutation does not stain for STAG2 expression, whereas the tumor with wild-type STAG2 stains positive for STAG2 expression. 

B, Western immunoblot analysis of STAG2 protein expression in Ewing sarcoma cell lines with mutated (black box), rearranged (green box), and wild-type (white box) STAG2. Mutations and rearrangements of STAG2 result in loss of expression of protein, but some cell lines with wild-type STAG2 also have loss of protein expression. 

C, top, schematic illustrating novel STAG2-MAP7D3 fusion transcripts detected by RNASeq in the Ewing sarcoma cell line TTC466. Bottom, RT-PCR detected the presence of both STAG2-MAP7D3 fusion forms in TTC466 but not in CADO-ES1 (negative control) Ewing sarcoma cell lines. 

D, scatter plot of tumor ploidy for tumors staining positive for STAG2 versus STAG2 loss (mean: 2.12 vs. 2.02; P = 0.585 by Mann–Whitney test). 

E, scatter plot of the fraction of the genome affected by SCNAs in all Ewing tumors staining positive for STAG2 versus STAG2 loss by IHC (mean: 0.15 vs. 0.27; P = 0.002 by Fisher exact test). 

F, the fraction of genome affected by SCNAs in TP53 wild-type diagnostic tumors staining positive for STAG2 versus STAG2 loss (mean: 0.14 vs. 0.27; P = 0.108 by Mann–Whitney test). 

G, 88% of patients with STAG2 loss by IHC at diagnosis presented with metastatic disease compared with 27% of patients whose diagnostic tumors expressed STAG2 (P = 0.002 by Fisher exact test). 

H, heatmap of ssGSEA enrichment z scores for the significant gene sets (SNR ≥ 1.5 and FDR q value < 0.05) positively correlated with STAG2 loss (top) or STAG2 expression (bottom). Geneset names are listed in Supplementary Table S13.

I and J, GSEA plots of running enrichment scores for two metastatic signatures (l and k) significantly enriched in STAG2 loss samples. **, P < 0.01.
dinucleotides, a pattern attributed to the activity of APOBEC enzymes in regions of DNA damage. However, there was no evidence of kataegis or increased APOBEC expression in treated samples, patterns usually associated with APOBEC-induced mutations (Fig. 4C and D; refs. 44–47). Although the analysis of translocations was limited by the relative paucity of WGS data analyzed (Supplementary Table S12), we detected an increase in somatic aberrancies in Ewing sarcoma tumors. Because paired normal controls for cell lines are unavailable, WES data were subjected to the same filtering process as tumors for which we did not have matched normal samples (Supplementary Table S16). We detected an EWS–ETS fusion for all cell lines with RNASeq data (Supplementary Table S4C). We found a significantly increased rate of STAG2 mutations (36%; 4 of 11; 95% CI, 11%–69%; \( P = 0.016 \)), STAG2 loss by Western blot analysis or IHC (70%; 7 of 10; 95% CI, 35%–93%; \( P = 0.0006 \)), TP53 mutations (64%; 7 of 11; 95% CI, 31%–89%; \( P < 0.0001 \)), CDKN2A deletion (55%; 6 of 11; 95% CI, 23%–83%; \( P = 0.028 \)), and chromosome 1q copy-number gain (55%; 6 of 11; 95% CI, 23%–83%; \( P = 0.011 \)) in cell lines compared with tumors for which the Fisher exact test. There was no significant difference in the rate of copy-number gains of chromosome 8 (6 of 11), and although there were more cell lines from females (8 of 11) than from males, this did not reach statistical significance when compared with the gender of patients in our study (Fig. 5A). We also performed principal component analysis using the RNASeq data from all tumors and cell lines. Projection of the first two principal components, accounting for the largest amount of variance in the samples, revealed that cell lines and tumors clustered separately (Fig. 5B). However, only the first principal component (PC1) was found to significantly discriminate tumors and cell lines (\( P < 1 \times 10^{-14} \))
The Genomics of Ewing Sarcoma

**Figure 4.** Ewing sarcoma tumors acquire somatic aberrancies with treatment. **A,** coding mutations in 24 Ewing sarcoma tumor samples identified from WES and WGS of patient tumors and paired normal tissue with available clinical data. Bars indicate the number of coding-region mutations in each tumor and are grouped by disease state (diagnostic or treated) and ordered by number of mutations (high to low) within each group. **B,** the rate of coding mutations is higher in tumors sequenced after treatment than in tumors from diagnostic samples. Plotted are the average mutation rates for diagnostic and treated samples ± the standard deviation (SD). \( p = 0.0017 \) by Mann–Whitney test. **C,** treated tumors have a lower fraction of CpG to T transitions and a higher fraction of substitutions at cytosines in TpC dinucleotides. The NpN nomenclature indicates the dinucleotide sequence involved in the indicated mutation and \* indicates the mutated base. Plotted are the fractions of SNVs for each mutational category ± the 95% CI; *, \( p < 0.05 \); **, \( p < 0.01 \) by \( z \) test. **D,** shown is the rate of SNVs by 3-base context from tumors acquired at diagnosis (left) or after treatment (right). Colors indicate the single-nucleotide change, and the location of each bar in the matrix indicates the flanking bases. **E** and **F,** circo plots of seven tumors sequenced by WGS. Chromosomes displaying cytobands are arranged end-to-end in the outer ring with the inner ring showing segmented copy-number variations generated from exome sequencing. Interchromosomal rearrangements are displayed as purple arcs and intrachromosomal as green arcs. Rearrangements at t(21;22)(q22;q12) in SJDES010 and SJDES018-R indicate the EWS–ERG fusions in these samples. All other tumors demonstrate a translocation at t(11;22)(q24;q12), consistent with the EWS–FLI fusion.
We then applied ssGSEA after ranking the genes by their weighted contribution to PC1 and found that cell lines were enriched for gene sets associated with cell proliferation and metabolism, whereas tumors were enriched for gene sets associated with the immune system or signaling from the extracellular matrix (Fig. 5D and Supplementary Table S17). In summary, these data suggest that although Ewing sarcoma cell lines possess the spectrum of recurrent molecular events observed in EWS–ETS–positive primary tumors, they may be more representative of high-risk disease because aberrancies associated with aggressive disease (e.g., TP53 mutations and CDKN2A deletions) occur at a higher frequency in the cell lines. Moreover, the transcriptional analysis suggests, not surprisingly, that there is an interaction of the primary tumors with the microenvironment, including infiltration with immune cells, and that there are differences in growth properties between cell lines and primary Ewing sarcoma tumors. However, additional differences in transcription between the two were not identified in our analysis, supporting the notion that other aspects of cell line transcriptional programming may more faithfully model tumor biology.

**Clonal Evolution and Tumor Heterogeneity**

For 2 patients, we performed WES on tumor DNA from both a primary tumor at diagnosis and a metastatic tumor at relapse (Supplementary Fig. S7A and Supplementary Table S17).
To explore clonal evolution, we estimated the cancer cell fraction (CCF) of these mutations by a central dot surrounded by a cloud. Light dots indicate single mutations. Black dots indicate multiple mutations with the same CCFs in both tumors. "Cloud plots" indicate the 90% CI of each CCF estimate (smaller clouds represent higher certainty).

**Figure 6.** Clonal evolution in Ewing sarcoma. **A** and **B**, Venn diagrams indicate the number of SNVs and indels detected in paired tumor samples by WES for mutations occurring at bases with adequate coverage to power mutation detection in both tumors. Purple circles include mutations identified in the tumor sequenced from the 26 tumor–normal cohort, and the blue circles include mutations identified in the subsequent relapse. "Cloud plots" indicate the cancer cell fraction (CCF) of these mutations by a central dot surrounded by a cloud. Light dots indicate single mutations. Black dots indicate multiple mutations with the same CCFs in both tumors. Clouds depict the 90% CI of each CCF estimate (smaller clouds represent higher certainty). **A**, left and middle, cloud plots demonstrate the overlap of mutations identified in 2 patients from whom we sequenced a diagnostic tumor and relapsed metastatic tumor. Right top, MRI and CT images of primary pelvic tumor at diagnosis and lung metastasis at relapse for patient SJDES004. Right bottom, a model depicting the proposed evolution of a metastatic relapsed tumor from a primary tumor. Blue circles represent Ewing sarcoma cells. Symbols within cells represent clusters of mutations (circles) or specific mutations. Time moves from left to right with tumor sampling indicated by vertical lines. The large teardrop indicates the evolving relapsed cell population. Smaller, multicolored teardrops indicate subclonal populations of cells. The absence of some mutations at relapse that are estimated to be clonal at diagnosis indicates that metastatic spread likely occurs after tumor initiation but before the emergence of the dominant clone in the primary tumor. **B**, left and middle, cloud plots demonstrate mutations identified in two separate locally relapsed tumors from patients with Ewing sarcoma. Right top, MRI and PET images (left to right) from diagnosis, relapse 1 (R1), and relapse 2 (R2) from patient SJDES007 with Ewing sarcoma involving the scapula. Right bottom, schematic depicts the proposed mechanism of evolution from diagnosis to each relapse for tumors from patient SJDES007. At diagnosis, tumors likely contain a mutation in STAG2 (extrapolated from IHC demonstrating loss of STAG2 expression) that cannot be 100% clonal due to the presence of different STAG2 mutations at R1 and R2. The presence of clonal mutations in R1 that are not present in R2 (and vice versa) suggest that local relapses emerge from cells that have diverged from the primary tumor.
mutations at diagnosis became clonal at relapse, and recurrent metastatic tumors acquired new mutations. For two additional patients, we were unable to obtain diagnostic samples for sequencing but were able to sequence tumors from separate relapses (Supplementary Fig. S7B and Supplementary Tables S1D and S18). In both cases, relapses were local to the site of the diagnostic tumor, but there was remarkably little overlap in the mutations identified at the first relapse (R1) compared with the subsequent relapse (R2). For patient SJDES007, we confirmed that the same genomic EWS–FLI rearrangement presented by WGS in R1 was also present in R2, suggesting that these were not independent occurrences of Ewing sarcoma in the same patient. These findings indicate that tumor heterogeneity in the primary tumor influences the mutational profile of tumors at relapse (Fig. 6B).

One striking observation from patient SJDES007 was the finding that the TP53 and STAG2 mutations identified in tumor R1 were not detected in R2. Instead, mutations at different sites were detected in both of these genes at R2. This convergent evolution is consistent with the finding that mutations in these genes may co-occur more often than expected by chance and suggests that they play an important role in relapse (Fig. 6B and Supplementary Table S11A–S11D). Although copy-number loss in R2 could account for the loss of the TP53 mutation initially seen in R1, copy-number loss is unlikely to explain the presence of different STAG2 mutations in tumors R1 and R2 because the STAG2 gene is located on the X chromosome and this is a male patient (Supplementary Fig. S7B). In fact, the absence of STAG2 expression in the diagnostic sample (determined by IHC staining of a formalin-fixed paraffin-embedded sample; Supplementary Table S12) suggests that the tumor populations that gave rise to each relapse had diverged before diagnosis, further supporting the notion that tumor heterogeneity at diagnosis influences tumor evolution at relapse. Finally, sequencing of tumors sampled at R1 and R2 from patient SJDES014, an infant with multiply relapsed chest wall Ewing sarcoma, demonstrated that the EWS–FLI rearrangement was the only somatic mutation in common to both relapses (Fig. 6D and Supplementary Tables S2B, S4C, and S18). Although analysis of this patient’s normal exome did not reveal a mutation associated with a known cancer predisposition, the patient’s unusually young age at diagnosis and the absence of other clonal, tumor-initiating events leads us to speculate that germline variants or lineage-specific epigenetic changes may play an important role in Ewing sarcoma oncogenesis for some patients.

**DISCUSSION**

One ongoing debate in the pediatric oncology community has focused on the utility of clinical-grade deep sequencing of pediatric tumors to enable molecularly informed therapy, matching genetic lesions with targeted drugs. In this regard, our study reveals that Ewing sarcoma genomes are among the most simple reported to date and demonstrates a marked paucity of mutations in immediately targetable pathways, such as mutations in the MAPK or PI3K–AKT–mTOR pathways. In the case of MAPK pathway mutations, a similar observation was made in prostate cancer in which mutations in the MAPK pathway were rare and observed only in ETS fusion-negative tumors (43). Indeed, it has been suggested that oncogenic ETS proteins can replace the need for MAPK pathway mutations by the upregulation of a MAPK transcriptional program (50). At relapse, the frequency of genetic aberrancies is significantly higher, although the paucity of targetable lesions remains a challenge. This does raise the point, however, that diagnostic and relapsed tumors are quite different, supporting the notion that relapsed tissue warrants biopsy and genomic characterization for optimal precision-based therapy in the future.

On the one hand, the simplicity of the Ewing sarcoma genome is disquieting in that there is not an immediately translatable intervention. In addition, in the two cases with paired relapse samples from the same patient, only the minority of mutations were present in both relapses. These results, however, do speak to the critical importance of understanding and either directly targeting the EWS–ETS fusion event in this disease or identifying synthetic lethal vulnerabilities in the presence of these fusion proteins. Toward this end, by integrating multiple genomic approaches, we addressed several previously outstanding questions regarding the EWS–ETS fusions in Ewing sarcoma tumors. First, in Ewing sarcoma, the reciprocal fusion, ETS–EWS, is generally not expressed and thus unlikely to be relevant to disease pathogenesis. This is in contrast to other fusions, such as PML–RARa in acute promyelocytic leukemia, in which the reciprocal fusion is thought to be critical to leukemogenesis (51). Second, although wild-type EWS is highly expressed in Ewing sarcoma tumors, both wild-type FLI and wild-type ERG are poorly expressed. Third, our data highlight the need to better understand the role of nonrearranged ETS genes in Ewing sarcoma even beyond the expressed EWS–ETS fusion proteins.

One of the most striking recurrent events found in our sequencing effort was mutations or rearrangements of the STAG2 gene. In further support of a functional relevance of STAG2 to Ewing sarcoma, it is the finding of unique STAG2 mutations at different relapses in 1 patient. Moreover, the co-occurrence of these mutations with unique mutations in TP53 at each relapse supports a likely functional interaction of these two events in promoting disease recurrence. Indeed, co-occurrence of STAG2 and TP53 mutations was observed more often than expected by chance in Ewing sarcoma tumors. All frameshift and nonsense mutations in STAG2 resulted in the loss of protein expression in Ewing sarcoma tumors with sufficient tissue for evaluation by IHC. Moreover, we observed additional samples with loss of STAG2 expression despite normal coding sequences. STAG2 is a member of the cohesin complex, a multimeric complex consisting of at least four core proteins: a heterodimer of SMC1 and SMC3 forming a hinged ring-like structure and two “clasp” proteins, STAG and RAD21 (52, 53). Although mutations in multiple members of the cohesin complex have been reported in cancer, in this collection of Ewing sarcoma tumors, we observed recurrent mutations only in STAG2, consistent with recent reports (4, 16, 54, 55). The cohesin complex is known to play a role in the regulation of sister chromatid exchange during mitosis and meiosis, and in some reports, mutations in STAG2 are associated with aneuploidy (15, 36). In Ewing sarcoma tumors, however, there was no difference in tumor ploidy...
among tumors expressing STAG2 or those with STAG2 loss. Furthermore, we have identified a STAG2 mutation in one tumor sample with no copy-number alterations, suggesting that the loss of STAG2 expression likely contributes to disease pathogenesis or progression through other mechanisms. One alternative hypothesis is that mutations in STAG2 lead to altered chromatin architecture and transcription, a putative mechanism in the developmental disorders characterized by mutations in the cohesin complex, such as Cornelia de Lange Syndrome and Roberts Syndrome (53, 56–59). Moreover, despite a small sample size, it is notable that 88% of patients with STAG2 loss at diagnosis also had metastatic disease, in comparison to only 27% of patients with tumors expressing STAG2. Because of a lack of available follow-up data for the majority of patients whose tumors were included in this study, we were unable to definitively demonstrate that these patients had a worse outcome. Although metastatic disease is currently the most predictive marker of poor prognosis in Ewing sarcoma, this hypothesis will need verification in a larger cohort. Thus, although loss-of-function mutations of STAG2 in Ewing sarcoma may portend a poor prognosis, it may also impart a new, targetable vulnerability for the development of alternative treatment strategies.

In this study, we also identified rare, recurrent mutations in genes such as BORL1, MLL2, and ERF, which did not reach statistical significance in this dataset. Indeed, one of the shortcomings of comprehensive sequencing efforts for rare pediatric tumors, such as this study, is the lack of adequate samples to power the discovery of rare events. Assuming a stable mutational frequency of 0.62 mutations per Mb, we would need to sequence more than 200 tumor/normal pairs to have 90% power to identify mutations present in 5% of patients and approximately 500 samples to detect significant events in 3% of patients (19). With an incidence of approximately 250 cases in the United States per year, such an analysis will require a large multi-institutional effort or meta-analysis as data from other Ewing sarcoma sequencing projects become available. However, we present here a novel approach to utilizing tumors acquired from pathologic discards that do not have paired normal tissue. Although we expect to see an increase in false positives using this approach, this analysis increased our confidence that we were not “missing” significantly mutated, well-validated cancer-associated genes expected to occur in 15% or more of patients with Ewing sarcoma. It is yet to be seen whether this method will be as informative for other cancer types, particularly those with a much higher rate of somatic mutations. A second shortcoming of our study is the lack of deep clinical information for many of the samples. Prospective studies to definitively address the relevance of STAG2 aberrancy on long-term survival, for example, will be of future interest.

In summary, we report a massively parallel sequencing approach, this analysis increased our confidence that we were not “missing” significantly mutated, well-validated cancer-associated genes expected to occur in 15% or more of patients with Ewing sarcoma. It is yet to be seen whether this method will be as informative for other cancer types, particularly those with a much higher rate of somatic mutations. A second shortcoming of our study is the lack of deep clinical information for many of the samples. Prospective studies to definitively address the relevance of STAG2 aberrancy on long-term survival, for example, will be of future interest.

**METHODS**

**Samples**

Tumor tissue and matched blood from patients with Ewing sarcoma were collected with informed consent under a Hospital Sant Joan de Déu (Barcelona, Spain), Institutional Review Board (IRB)-approved protocol. Additional tumor tissue samples were collected on an IRB-approved tissue discard protocol from Boston Children’s Hospital (Boston, MA) and acquired through the Cooperative Human Tissue Network, Southern Division (Birmingham, AL). Protocols were reviewed and approved by the Broad Institute’s (Cambridge, MA) IRB. DNA and RNA were extracted from tumors and cell lines using standard techniques.

**Massively Parallel Sequencing**

**Whole-exome sequencing**

Library construction followed procedures as previously reported (60). Barcoded exon capture libraries were pooled into batches of 96 samples for sequencing.

**Whole-genome sequencing**

Creation of adaptor-ligated barcoded libraries was performed as for hybrid capture, except that DNA was sheared to a larger fragment size (adapter-ligated fragment; range, 340–510 bp). To minimize molecular duplication and facilitate deep and diverse sequence coverage, two libraries were prepared from each sample.

**Transcriptome sequencing**

RNA for each sample was converted into a library of template molecules for sequencing according to the protocol for the Illumina TruSeq unstranded RNA Sample Protocol (61). All libraries were sequenced on the Illumina HiSeq 2000 instrument set to generate 76-bp paired-end reads for WES and RNASeq and 101-bp paired-end reads for WGS.

**Genotyping arrays**

Tumor-normal matched samples were processed and hybridized to Affymetrix SNP 6.0 arrays for genotyping and copy-number analysis (62).

**Contamination estimates**

SNP fingerprinting was used to cross-check each sequenced lane from a tumor-normal pair against all other lanes from that pair and to ensure matching between WES, WGS, SNP array, and RNASeq. Cross-contamination was estimated by using ContEst (63). All samples were found to have less than 4% contamination.

**Sequencing Analysis**

**DNA sequence alignment and variant detection**

To prepare read alignments for analysis, we processed all sequence data through the Broad Institute's data-processing pipeline, "Picard." For each sample, this pipeline combines data from multiple libraries and flow cell runs into a single BAM file. This file contains reads aligned to the human genome with quality scores recalibrated using the TableRecalibration tool from the Genome Analysis Toolkit (GATK; ref. 64). Reads were aligned to the Human Genome Reference Consortium build 37 (GRCh37) using the Burrows-Wheeler Aligner. All BAM files were deposited in the database of Genotypes and Phenotypes (dbGaP) upon publication. Variant detection and analysis of the BAM files were performed using the Broad Institute’s Cancer Genome Analysis infrastructure program "Firehose." Firehose facilitates comparison of BAM files from matched tumor–normal pairs and coordinates execution of specific modules including quality control, local realignment, mutation calling, small insertion and deletion identification, rearrangement
detection, variant annotation, computation of mutation rates, and calculation of sequencing metrics. Module versioning and logging of the specific analytic parameters is also tracked. Somatic SNVs and indels were identified using MuTect (Firehose CallSomaticMutations v123; ref. 65), and Indelocator (Firehose CallIndelPipeline v65; ref. 66), respectively. Each mutation was annotated by Oncotator (Firehose OncotorPipeline v81). We manually reviewed all somatic mutations and indels identified in the 26 tumor-normal pairs by the automated pipeline using the Integrated Genomic Viewer (IGV; ref. 67). Manual review removed approximately 20% of the somatic mutations from subsequent analysis, consistent with a very low false-positive rate (~0.1/Mb). Genomic rearrangements were detected from WGS and WES data using dRanger (Firehose dRanger v146) or manual review using IGV. Somatic copy-number variants were detected from SNP-array data using the Broad Institute’s copy-number pipeline and from WES data using CapSeg (Firehose CapSegModule v77), which uses read depth in the exome to estimate somatic copy-number changes across the genome. SNP-array data were analyzed using the tool ABSOLUTE to infer the tumor purity and ploidy (49, 68).

Germline filtering for tumor-only and cell line samples We devised a filtering approach to reduce the nonsomatic component of mutations detected in tumor samples and cell lines without matched normal samples. The filter removed recurring germline mutations with population minor allele frequencies exceeding 0.5% and removed any candidate mutation occurring at sites of recurring artifacts identified in more than 15% of a panel of 2,913 normal exomes. The processing for tumor-only and cell line data consisted of three steps.

1. Candidate mutation detection and raw filtering. After aligning tumor and cell line data to the reference genome (GRCh37), we generated candidate mutation lists with MuTect (Firehose CallSomatic-Mutations v131; ref. 65) and Indelocator (Firehose CallIndelPipeline v77; ref. 66) by pairing tumor samples with a normal sample (SJDE001). Variants with less than 3 alternate alleles (alt) reads were removed. Also at this step, mutations outside the coding regions were eliminated from consideration because non-coding coverage is expected to vary between WES platforms (69).

2. Common population variant filter. Tumor and cell line variants were filtered if the allele frequencies exceeded 0.5% in 1000 Genomes Project phase I data (any major population group) or in African American or European American populations in the ESP (20, 21).

3. Panel of normals filter (PoN). For remaining variants, corresponding loci were examined in data from 2,913 normal samples. Each site in each normal sample was classified into one of the following categories for each sample:

   - PoN(1): total read depth at loci < 8x (i.e., insufficient coverage)
   - PoN(2): alt count ≥ 10 and alt fraction ≥ 20%
   - PoN(3): alt count ≥ 3 and alt fraction > 20%
   - PoN(4): alt count ≥ 3 and alt fraction ≥ 3%
   - PoN(5): alt count ≥ 3 and alt fraction ≥ 1%
   - PoN(6): alt count ≥ 2 and alt fraction ≥ 0.3%
   - PoN(7): alt count ≥ 2 and alt fraction ≥ 0.1%
   - PoN(8): total depth ≥ 8 and none of previous criteria.

The fraction of the 2,913 normal samples in each of the eight categories was calculated for each candidate mutation site. Candidate variants at loci that did not meet the following criteria were removed from subsequent analysis:

\[
\text{Fraction samples PoN(1)<0.15 & PoN(2)<0.001 & Sum(PoN(4-7)<0.15 & PoN(8)<0.7) }
\]

Although approximately 150 rare germline variants per tumor sample pass this filter, we found that the nonsomatic somatic ratio of mutations was reduced to approximately 12:1 with this method (based on the comparison of filtering of 26 tumors with and without pairing of normal samples), which is sufficient for MutSig 2CV to identify significantly recurring mutations above the background germline mutation rate (19). Following the “panel of normal” filter, SNVs were also subjected to “oxo-G” filtering to remove C>A artifact from library preparation–associated oxidative damage (70).

Determining DNA somatic variant significance GISTIC2.0 (Firehose Gistic2_Pipeline v99) was used to identify significantly recurring SCNAs from both SNP-array and WES data (33). Identification of gene mutations most likely to be driving tumorigenesis (significantly mutated genes) was performed by the MutSig 2CV algorithm (19). Estimating CCF Tumor purity and ploidy values were estimated using the ABSOLUTE Toolkit (68) based on SCNA and SNV allele fractions. In samples without sufficient aneuploidy for ABSOLUTE to converge on a solution, a custom MATLAB script estimated purity based on evidence from allele fraction distribution of mutations that were consistent with clonal heterozygous somatic SNVs. The CCF of each mutation was then estimated by ABSOLUTE or by the custom MATLAB script using purity and ploidy estimates (49).

Transcriptome sequencing alignment and analysis The PRADA pipeline was used for alignment of RNA sequenced reads (Firehose BWATranscipe v32) to a reference genome consisting of GRCh37 and the human transcriptome. Mutations detected from the WES and WGS were compared with RNAseq data from the same samples to estimate the mutation validation rate (Supplementary Fig. S2, Firehose CrossvalidateMutations v15; ref. 66). RNA fusions were identified from the PRADA pipeline (Firehole RNAFusion v19) based on reads containing mates mapping to different coding regions. Exon-level transcript expression from RNASeq data was evaluated using RNA-SeQC (Firehose RNASeqMetrics v35; ref. 71). Reads were then converted to gene-level expression by determining the reads mapped to the genome per kilobase per million mapped reads (RPKM) for each gene (72). We also calculated gene-fragment RPKMs for the EWS, FLI, and ERG genes involved in EWS-ETS translocations. RPKM calculations for pre-fusion exon regions and for postfusion exon regions were performed separately for EWS, FLI, and ERG genes in each tumor sample.

Cell Culture

Ewing sarcoma cell lines were provided by the laboratory of Dr. Todd Golub (Broad Institute) except for the EWS502 and EWS834 cells, both of which were a gift from Dr. Jonathan Fletcher (Brigham and Women’s Hospital, Boston, MA) and cultured as previously described (32). CHLA-258 cells were grown in Iscove’s Modified Dulbecco’s Medium (Invitrogen) with 20% heat-inactivated FBS (Gemini Bio-Products) and supplemented with 10 mg/L insulin, 5.5 mg/L transferrin, and 6.7 μg/L sodium selenite (ITS; Thermo Fisher Scientific IST) and 0.7 mmol/L l-glutamine (Thermo Fisher Scientific). Cell lines were authenticated with WES and RNASeq as described in the article. Please see the Supplementary Data for additional methods.

Disclosure of Potential Conflicts of Interest

C. Rodriguez-Galindo is a consultant/advisory board member for Novimmune. No potential conflicts of interest were disclosed by the other authors.

Authors’ Contributions

Conception and design: T.R. Golub, J. Mora, K. Stegmaier Development of methodology: C. Stewart, A. Taylor-Weiner, K.C. Kurek, S.L. Carter, M. Suñol, M.S. Lawrence, M.D. Fleming, J. Mora Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): B.D. Crompton, K.C. Kurek,
The Genomics of Ewing Sarcoma


Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): B.D. Crompton, C. de Torres, L. Ambrogio, D. Auclair, S. Seepo, J. Mora

Study supervision: M.D. Fleming, J. Mora, K. Stegmaier

Other (variant validation): B. Blumenstiel

Other (sequencing follow-up assay): M. DeFelice

Acknowledgments

The authors thank the members of the Broad Institute Biological Samples Platform, Genetic Analysis Platform, and Genome Sequencing Platform for their assistance. The authors also thank Elizabeth Hwang for technical support for this project.

Grant Support

This work was conducted as part of the Slim Initiative for Genomic Medicine (SIGMA), a joint U.S.-Mexico project founded by the Carlos Slim Health Institute. This work was also supported by a Stand Up Medicine (SIGMA), a joint U.S.–Mexico project founded by the Carlos Slim Health Institute. This work was also supported by a Stand Up Medicine (SIGMA), a joint U.S.–Mexico project founded by the Carlos Slim Health Institute.

REFERENCES


RESEARCH ARTICLE

Crompton et al.


The Genomics of Ewing Sarcoma
