Ewing sarcoma is a primary bone tumor initiated by \textit{EWSR1–ETS} gene fusions. To identify secondary genetic lesions that contribute to tumor progression, we performed whole-genome sequencing of 112 Ewing sarcoma samples and matched germline DNA. Overall, Ewing sarcoma tumors had relatively few single-nucleotide variants, indels, structural variants, and copy-number alterations. Apart from whole chromosome arm copy-number changes, the most common somatic mutations were detected in \textit{STAG2} (17%), \textit{CDKN2A} (12%), \textit{TP53} (7%), \textit{EZH2}, \textit{BCOR}, and \textit{ZMYM3} (2.7% each). Strikingly, \textit{STAG2} mutations and \textit{CDKN2A} deletions were mutually exclusive, as confirmed in Ewing sarcoma cell lines. In an expanded cohort of 299 patients with clinical data, we discovered that \textit{STAG2} and \textit{TP53} mutations are often concurrent and are associated with poor outcome. Finally, we detected subclonal \textit{STAG2} mutations in diagnostic tumors and expansion of \textit{STAG2}-immunonegative cells in relapsed tumors as compared with matched diagnostic samples.

**SIGNIFICANCE:** Whole-genome sequencing reveals that the somatic mutation rate in Ewing sarcoma is low. Tumors that harbor \textit{STAG2} and \textit{TP53} mutations have a particularly dismal prognosis with current treatments and require alternative therapies. Novel drugs that target epigenetic regulators may constitute viable therapeutic strategies in a subset of patients with mutations in chromatin modifiers.
INTRODUCTION

Ewing sarcoma is the second most common primary malignant bone tumor in children and adolescents; the mean age at diagnosis is 15 years. Ewing sarcoma can affect any bone, but the most common primary sites are pelvis, femur, and tibia (reviewed in ref. 1). The annual incidence of Ewing sarcoma is approximately 3 per million, with a slight male bias. Histologically, Ewing sarcoma belongs to the group of small blue round-cell tumors, and the tumor cells often have abundant cytoplasmic glycogen and express CD99 on the plasma membrane (1).

Genetically, most Ewing sarcomas are characterized by a specific t(11;22)(q12;q11.2) translocation that fuses the EWSR1 gene on chromosome (chr) 22 with the FLI1 gene on chr 11 (2). In 10% to 15% of cases, EWSR1 is fused to genes encoding other members of the ETS family of transcription factors, including ERG, ETV1, E1AF, or FEV (1). In an even smaller number of cases, TAF15 and TLS/FUS, which encode the two other members of the TET family of RNA-binding proteins, may be fused to ETS family members (1). All fusions juxtapose the N-terminal domain of the TET gene family member to the DNA-binding domain of the ETS gene family member. These TET–ETS fusions are potent oncogenes that can transform NIH3T3 cells (3) by perturbing the expression of genes required for a variety of cellular processes, including cell-cycle regulation, signal transduction, and telomere maintenance (reviewed in ref. 1).

Chromosome or array-based comparative genomic hybridization (CGH) as well as SNP arrays have identified recurrent DNA copy-number alterations (CNA) in Ewing sarcoma (4–10). The most common copy-number gains occur in whole chromosomes 8 and 12 and the q (long) arm of chr 1. The long arm of chr 16 and the CDKN2A locus on chr 9p are the most common copy-number losses in Ewing sarcoma. The adverse prognosis conferred by chr 1q gain and chr 16q or CDKN2A loss has been reported, as has the negative impact of TP53 mutations (11). Finally, somatic STAG2 mutations were recently observed in a significant fraction of Ewing sarcoma cases (21%; ref. 12). Although the role of TET–ETS oncogenes in Ewing sarcoma tumorigenesis and progression has been extensively studied and the relation of copy-number changes to prognosis is emerging, relatively little is known domain of the TET gene family member to the DNA-binding domain of the ETS gene family member. These TET–ETS fusions are potent oncogenes that can transform NIH3T3 cells (3) by perturbing the expression of genes required for a variety of cellular processes, including cell-cycle regulation, signal transduction, and telomere maintenance (reviewed in ref. 1).
about additional secondary genetic lesions in Ewing sarcoma beyond these chromosomal lesions.

To identify secondary genetic lesions that contribute to Ewing sarcoma tumorigenesis after formation of the TET–ETS fusion, we performed whole-genome sequencing (WGS) of 112 tumors and their matched germline DNA. The most frequent point mutations involved the \textit{STAG2} and \textit{TP53} genes, and the prognostic significance of these mutations was further demonstrated in a series of 299 cases. \textit{STAG2} mutations were significantly associated with the occurrence of structural variations and were mutually exclusive with \textit{CDKN2A} deletions. In some cases, we also observed a small number of \textit{STAG2}-deficient tumor cells that survived treatment and comprised the major clone in the recurrent tumors.

**RESULTS**

\textbf{Ewing Sarcoma Has Low Numbers of Single-Nucleotide and Structural Variants}

Our discovery set for WGS comprised 112 Ewing sarcomas with matched germline DNA (Fig. 1 and Supplementary Table S1). All Ewing sarcoma tumors, with the exception of one (SJ001301) that had insufficient tumor sample for analysis, expressed \textit{EWSR1–ETS} fusions: \textit{EWSR1–FLI1} in 101 cases, \textit{EWSR1–ERG} in 9 cases, and \textit{EWSR1–ETV1} in one case. Tumor and germline DNA were sequenced at a median depth of 35× and 25×, respectively. Mapping, detection, and annotation of single-nucleotide variants (SNV), insertions or deletions (indels), and structural variants (SV), functional predictions, and CNAs were computed from the WGS data as previously described (13–15). Eighty percent of the tumors had >70% tumor purity, leading to a 98% power for detecting mutations present in the predominant tumor clones in this cohort (Supplementary Table S2A).

The median number of somatic SVs was 7 (range, 0–66) per tumor (Supplementary Table S2B). In most cases (106/112; 95%), WGS detected SVs within the previously described \textit{EWSR1} and \textit{ETS} chromosome breakpoint regions (ref. 16; Fig. 1; Supplementary Table S2C; and Supplementary Fig. S1A–S1D). Five cases (SJ001303, SJ001320, IC198, IC273, and IC086) exhibited chromothripsis, including three cases with chromothripsis on chr 21 and 22 associated with \textit{EWSR1–ERG} fusions (SJ001303, IC198, and IC273) and one case involving chr 22 associated with an \textit{EWSR1–FLI1} fusion (SJ001320). CNAs could be reliably analyzed from WGS data in 103 cases. Nine cases were excluded from CNA analysis due to low tumor purity or uneven sequencing coverage. The most frequent CNAs were gain of whole chr 8 (49/103; 47%), gain of whole chr 12 (22/103; 21%), gain of the long arm of chr 16 (19/103; 18%), deletion of the long arm of chr 16 (15/103; 14%), and deletion of the short arm of chr 14 (11/103; 11%).
Comprehensive Genomic Analysis of Ewing Sarcoma

(18/103; 17%), and deletion of the CDKN2A locus on the short arm of chr 9 (12/103; 12%; Fig. 1; Supplementary Fig. S2; and Supplementary Table S2D). Chr 1q gain and chr 16q loss were correlated with shorter survival (P = 2 × 10^{-5} and P = 0.0037, respectively, log-rank test; Fig. 2A). As chr 16q deletion and chr 1q gain were highly significantly co-associated (P = 10^{-8}, Fisher exact test), their combination did not show an additional effect on overall survival (Fig. 2A). Chr 8 and chr 12 gains also showed a significant, although less pronounced, co-association (P = 1.63 × 10^{-3}, Fisher exact test), but neither chr 8 nor chr 12 gains, nor their combination, were correlated with shorter survival (data not shown).

Experimental validation by custom capture and Illumina sequencing of all WGS-predicted SNVs and indels in 19 cases showed a 95.6% verification rate. Across the entire cohort, the median number of SNVs was 319 (range, 13–1,747) per genome (Supplementary Table S2E and S2F). The background mutation rate ranged from 8.0 × 10^{-9} to 1.4 × 10^{-6} (median, 2.4 × 10^{-7}) per base. The predominant changes were C(G)>T(A) transitions (Supplementary Fig. S3). No bias in SNV distribution was observed in cases with the highest number of SNVs. In particular, rainfall plots for 31 samples that had at least 400 somatic SNVs across the genome showed no patterns of kataegis (17). On average, there were 10 (range, 1–39) coding variants per tumor, and the ratio of missense to silent mutations was 2.4. We observed a positive correlation between age at diagnosis and the number of SNVs (r^2 = 0.42; P = 2.7 × 10^{-5}, Pearson correlation). Patients older than 20 years at diagnosis had significantly more SNVs than did younger patients (P = 0.001, Mann–Whitney U test). Survival analysis showed a negative correlation with the number (tertile-based) of SNVs/indels, i.e., a greater number of SNVs/indels was associated with shorter survival time (Fig. 2B; P = 0.04, log-rank test). Tertile-based survival did not differ significantly according to the number of SVs. However, the 8 patients whose tumors had a large number of SVs (outliers in the box plot distribution of SVs shown in Fig. 2C) had very poor outcomes (Fig. 2C; P = 0.003, log-rank test).

The Most Frequent Coding Variants Occur in STAG2, TP53, and Epigenetic Regulators

The gene most frequently carrying a somatic mutation in our cohort was STAG2 (17%, 19/112). We identified 6 nonsense mutations, 10 indels leading to frameshifts, 1 missense mutation, 1 splice-site mutation, and 1 duplication of exon 22 (Figs. 1 and 3A). As the STAG2 protein is an integral member of the cohesin complex (18) and was found to be associated with aneuploidy (12), we also investigated the relation of
Figure 3. STAG2 mutations and their prognostic significance in Ewing sarcomas. A, schematic of the STAG2 protein and mutations. Mutations found in tumor samples are indicated above the protein, and those observed in cell lines are indicated below. Mutation nomenclature is based on the NM_001042749 reference sequence. Exon and amino-acid numbering is indicated below the protein. The recurrent R216* mutation was observed in 7 cases. One tumor (case IC871) had two mutations (indicated in bold). SCD, stromalin conservative domain; GR, glutamine-rich region. Box plots show comparison of the number of SVs (B) and SNVs/indels (C) in wild-type (WT) and STAG2-mutated tumor samples. Samples with a fractured genome or low tumor cell content (see Fig. 1) were excluded from analysis, leaving 17 STAG2-mutated cases and 86 wild-type cases. Box represents the central 50% of data points (interquartile range). Upper and lower whiskers represent the largest and smallest observed values within 1.5 times the interquartile range from the ends of the box. Circles, individual values. P values were determined by using the Mann–Whitney U test. D, overall survival among 299 patients according to STAG2 mutation status. The number of patients in the different groups is indicated in brackets. E, overall survival of the 299 patients according to their STAG2 and/or TP53 mutation status.
STAG2 mutations to the number of SVs across the discovery cohort. A significantly greater number of SVs was observed in STAG2-mutated cases (Fig. 3B; \( P = 0.006 \), Mann–Whitney U test). In contrast, STAG2 status was not associated with the number of SNVs or indels (Fig. 3C).

TP53 was mutated in 8 cases (Fig. 1). All mutations were missense, with the exception of one nonsense mutation (p.R317* according to NM_000546), and were described in the COSMIC database. After excluding the very large genes that are recurrently mutated in most cancer genome studies (\( \text{TTN, CSMD1, MACF1, and RYR2; ref. 19} \)), the third most frequently mutated genes were \( \text{EZH2, BCOR, and ZMYM3} \), which each presented with 3 mutations (3/112, 2.7%; Fig. 1 and Supplementary Table S2). All three \( \text{EZH2} \) mutations were missense mutations within the SET domain (Y646F, Y646H, and A682G according to NM_004456). \( \text{BCOR} \) exhibited one missense mutation (S1083I according to NM_017745), one indel leading to a frameshift (M1259fs), and one 116-kb intragenic deletion (Fig. 1 and Supplementary Table S2).

ZMYM3 exhibited two indels (L82fs according to NM_201599) and one 17-kb intragenic deletion (Fig. 1 and Supplementary Table S2).

All other somatic gene mutations were observed in less than three cases. Mutations affecting epigenetic regulators have been found to be significantly associated with some pediatric cancers (20). In addition to the mutations in \( \text{EZH2, BCOR, and ZMYM3} \), we identified novel somatic mutations in \( \text{SETD2, MLL2, MLL3, and PRDM9} \) (Fig. 1 and Supplementary Table S2). Of note, two novel missense mutations were observed in \( \text{EWSR1} \). Finally, we used the significantly mutated gene (SMG) test in the mutational significance in cancer (MuSiC) suite (21) to identify genes that are significantly enriched in somatic SNVs and indels. Only \( \text{STAG2, TP53, and EZH2} \) were found to be significantly enriched (Supplementary Table S3).

**STAG2 and CDKN2A Genetic Lesions Are Mutually Exclusive**

When investigating the relationships between gene mutations, SVs, and CNAs, we found a mutually exclusive pattern of STAG2 and CDKN2A genetic alterations (Fig. 1). To confirm this mutually exclusive profile, we investigated STAG2 and CDKN2A in a panel of 19 Ewing sarcoma cell lines. STAG2 mutations and CDKN2A deletions were observed in 9 and 6 of the 19 cell lines, respectively (Table 1). Of note, two novel missense mutations were observed in \( \text{EWSR1} \). Finally, we used the significantly mutated gene (SMG) test in the mutational significance in cancer (MuSiC) suite (21) to identify genes that are significantly enriched in somatic SNVs and indels. Only STAG2, TP53, and EZH2 were found to be significantly enriched (Supplementary Table S3).

<table>
<thead>
<tr>
<th>Cell line</th>
<th>STAG2*</th>
<th>CDKN2A*</th>
<th>TP53*</th>
</tr>
</thead>
<tbody>
<tr>
<td>EW-3</td>
<td>p.R216*</td>
<td>WT</td>
<td>WT</td>
</tr>
<tr>
<td>MHH-ES1</td>
<td>p.Q735fs</td>
<td>WT</td>
<td>p.S215del</td>
</tr>
<tr>
<td>ORS</td>
<td>p.D625fs</td>
<td>WT</td>
<td>p.C176F</td>
</tr>
<tr>
<td>POE</td>
<td>p.F667fs</td>
<td>WT</td>
<td>p.L194R</td>
</tr>
<tr>
<td>SK-ES-1</td>
<td>p.Q735*</td>
<td>WT</td>
<td>C176F</td>
</tr>
<tr>
<td>SK-NM-C</td>
<td>p.M1, R546Del</td>
<td>WT</td>
<td>p.M1, T125Del</td>
</tr>
<tr>
<td>A673</td>
<td>WT</td>
<td>del[1a,1b,2,3]</td>
<td>p.A119fs</td>
</tr>
<tr>
<td>EW-1</td>
<td>WT</td>
<td>del[1a,1b,2,3]</td>
<td>p.R273C</td>
</tr>
<tr>
<td>EW-7</td>
<td>WT</td>
<td>del[1a]b</td>
<td>WT</td>
</tr>
<tr>
<td>EW-16</td>
<td>WT</td>
<td>del[1a,1b,2,3]</td>
<td>p.K120fs</td>
</tr>
<tr>
<td>STA-ET-1</td>
<td>WT</td>
<td>del[1a,1b,2,3]</td>
<td>WT</td>
</tr>
<tr>
<td>STA-ET-3</td>
<td>WT</td>
<td>del[1b,2,3]</td>
<td>p.R213*</td>
</tr>
<tr>
<td>STA-ET-8</td>
<td>WT</td>
<td>WT</td>
<td>p.C176F</td>
</tr>
<tr>
<td>RD-ES</td>
<td>WT</td>
<td>WT</td>
<td>p.R273C</td>
</tr>
<tr>
<td>STA-ET-9</td>
<td>WT</td>
<td>WT</td>
<td>p.P152T</td>
</tr>
</tbody>
</table>

*STAG2 and TP53 mutations are annotated with respect to reference sequences NM_001042749 and NM_000546. For CDKN2A, numbers indicate the corresponding homozygous deleted exons (del) at this locus (exon 1a is specific for CDKN2Aink4a, exon 1b is specific for CDKN2Aarf, and exons 2 and 3 are common to both).

bIndicates a G>A polymorphism identified in EW-7 and EW-22 cell lines (rs3731249).

cThe STA-ET-3 cell line has a C to T heterozygous mutation (het) at position chr9:21,971,120 (hg19), leading to nonsense (p.R80* for p16ink4a based on NM_000077) and missense (p.P94L for p14arf based on NM_058195) mutations.
Fig. S4A and S4B). When tumor and cell line results were combined, this mutually exclusive pattern of alteration was highly significant ($P = 0.0079$, Fisher exact test). The frequency of TP53 mutations was extremely high in the cell lines (Table 1). Altogether, all tested cell lines harbored at least one STAG2, TP53, or CDKN2A lesion.

**STAG2 and TP53 Mutations Are Co-Associated in Highly Aggressive Tumors**

To determine whether STAG2 and/or TP53 mutations are associated with outcome in Ewing sarcoma, we analyzed these genes by targeted capture sequencing in an additional 199 French patients with Ewing sarcoma. Across the whole series, 30% of patients had metastatic spread at diagnosis. The presence of a metastasis was associated with a shorter overall survival time ($P = 4 \times 10^{-4}$, log-rank test). In total, 41 patients (13.2%) had STAG2 mutations (Fig. 3A and Supplementary Table S4) and 16 patients (5.2%) had TP53 mutations. The STAG2 mutations included 15 nonsense, 4 missense, 17 frameshift, and 4 splice-site mutations, 2 in-frame deletions, and 1 exon duplication (Figs. 1 and 3A and Supplementary Table S4). One tumor (IC871) had two distinct STAG2 mutations. Overall survival data were available for 299 patients. The presence of a STAG2 mutation was not significantly associated with dismal prognostic factors, including tumor size, response to chemotherapy, resection quality, or tumor spread. However, patients with STAG2 mutations demonstrated a significantly lower probability of survival, similar to patients with TP53-mutated tumors (Fig. 3D and E). Patients with neither STAG2 nor TP53 mutations had the highest probability of survival, and patients whose tumors carried mutations in both genes had the worst outcome (Fig. 3E). A significant decrease in overall survival of patients with either STAG2 or TP53 mutation alone was not observed. In our cohort, STAG2 and TP53 mutations were significantly co-associated ($P = 2.4 \times 10^{-4}$, Fisher exact test).

We also explored the CDKN2A status across these additional tumors. Expanding the CDKN2A cohort confirmed the exclusion pattern with STAG2 mutations. Indeed, we identified only 2 tumors with both STAG2 mutations and CDKN2A deletions. When compiling all our data (299 tumors and 19 cell lines), the overlap between STAG2 and CDKN2A genetic lesions was much lower than expected by chance [Fisher test: 0.0076, STAG2/CDKN2A, wild-type (WT)/WT: 221, WT/Mut: 49, Mut/WT: 46, Mut/Mut: 2]. CDKN2A status was not significantly associated with overall survival across the whole series (Supplementary Fig. S5A–S5C).

**Subclonal STAG2 Mutations May Expand at Relapse**

Finally, we investigated whether STAG2 mutation occurs in subclones within some tumors and whether it evolves during the course of the disease. We first took advantage of the high coverage obtained in the capture-based sequencing experiments to investigate the ratios of mutated/wild-type alleles. Seven diagnostic samples showed evidence of subclonal mutations, i.e., a mutant allele frequency <0.25 despite high tumor purity (Supplementary Table S4; example in Fig. 4A). In 21 cases, STAG2 immunostaining could be investigated in paired primary/relapse or pre-/post-therapy samples. In 18 cases, STAG2 immunostaining in paired primary/relapse or pre-/post-therapy was unaltered, of which 16 were positive and two were negative at both time points. However, in three cases, STAG2 staining at relapse revealed a reduction in STAG2-immunopositive cells (Fig. 4B). Consistent with the immunostaining result, loss-of-function STAG2 mutations were detected at relapse with high allelic fractions but were either not detected (SJEW001303) or detected at a subclonal level at diagnosis (SJEW014721; Supplementary Table S5).

**DISCUSSION**

To our knowledge, the work reported here is the most comprehensive genomic analysis of Ewing sarcoma performed to date. The cases we studied met all of the criteria defining *bona fide* Ewing sarcoma, including clinical, pathologic, and molecular findings. The background mutation rate of Ewing sarcoma was relatively low (2.4 × 10^{-4}), with a median of 10 coding somatic mutations per tumor. The Ewing sarcoma mutation rate is much lower than that usually observed in adult cancers and in the upper range of what is described in other pediatric solid malignancies and brain tumors, including neuroblastoma (22–24), retinoblastoma (14), rhabdomyosarcoma (25–26), medulloblastoma (27, 28), pilocytic astrocytoma (29), pediatric glioblastoma (30), and osteosarcoma (31). We also observed a positive correlation between age at diagnosis and the number of SNVs.

The CNA most frequently detected in the present study was gain of chr 8, which was observed in close to 50% of cases, in agreement with previous CGH or SNP-array data (4, 5, 7). Loss of chr 16q and gain of chr 1q were strongly co-associated, which is fully consistent with the presence in these tumors of a derivative chr 16 resulting from an unbalanced t(1;16) translocation previously identified by cytogenetics in Ewing sarcoma (32). However, no SVs specific for this translocation were detected, consistent with the hypothesis that this t(1;16) translocation occurs within repeated elements of centromeric regions that cannot be reliably detected by WGS. In this cohort, we also report that 16q and/or 1q gains have strong negative prognostic significance.

The most frequently mutated gene in Ewing sarcoma is STAG2. STAG1 and STAG2, the human orthologs of yeast Scc3p, encode components of the cohesin multiprotein complex that plays an essential role in sister chromatid cohesion (18). STAG1 and STAG2 exist in different cohesin complexes that are essential for telomere or centromere cohesion, respectively (18). STAG2 mutations were initially observed in a diverse range of cancers, including glioblastoma, melanoma, and Ewing sarcoma (12). Subsequently, STAG2 mutations were described in a significant proportion of bladder cancers (12, 33, 34) and myeloid neoplasms (35). Although experimental systems have shown that STAG1 and STAG2 inactivation drives aneuploidy (12, 18), STAG2 mutations were not found to be associated with aneuploidy or CNAs in bladder cancer (33, 36). The case may be slightly different in Ewing sarcoma, as we observed a positive correlation between the presence of STAG2 mutation and the number of SVs. However, the interpretation of this correlation must take into account the strong co-association of STAG2 and TP53 mutations in our cohort. When cases with only one of these two mutations are
Comprehensive Genomic Analysis of Ewing Sarcoma

Figure 4. Subclonal presence of STAG2 mutations. A, Integrative Genomics Viewer representation showing the subclonal presence of STAG2 mutations in one sample. B, evolution of STAG2 staining between diagnosis and relapse in two independent cases. Whereas only a small subset of tumor cells lacked STAG2 expression at diagnosis (see insets), the tumor cells were homogeneously negative at relapse. The few STAG2-positive stromal cells serve as an internal positive control.

We observed a previously unreported, mutually exclusive pattern of STAG2 and CDKN2A mutations in Ewing sarcoma. This mutual exclusivity was observed in primary tumors and confirmed in cell lines. In addition to their role in sister chromatin cohesion, STAG2-containing cohesin complexes play an essential role in nuclear chromatin organization, particularly in the epigenetic mechanisms of insulation through direct interaction between STAG2 and CTCF, a multifunctional transcription factor that regulates chromosomal boundaries of gene expression, as recently demonstrated at the H19/Igf2 locus (37). Interestingly, CTCF has also been shown to regulate the CDKN2A locus (38), raising the possibility that STAG2 loss of function alters the epigenetic regulation of CDKN2A in CDKN2A-wild-type cases. However, as previously reported (6, 39–41), methylation is not a common mechanism for CDKN2A inactivation in Ewing sarcoma and is therefore not expected to occur in most STAG2-wild-type cases. The role of STAG2 in chromatin structure, particularly in the distribution of histone marks, and expression of the CDKN2A locus should be further investigated in depth.

Three EZH2 mutations (Y646F, Y646H, and A682G, all in the SET domain) were observed in our cohort of patients. EZH2 encodes a member of the multiprotein polycomb repressive complex 2 (PRC2), which catalyzes trimethylation of histone H3 lysine 27 (H3K27me3). Residues Y646 and to a lesser extent A682 are frequently mutated in B-cell lymphoma, and these mutations have been shown to enhance EZH2 enzymatic activity and promote malignant lymphoid transformation (42–44). Mutations of EZH2 have also been observed in a subset of acute T-cell and myeloid malignancies (15, 45). In addition to EZH2, potentially deleterious mutations in ZMYM3 and BCOR, which also encode epigenetic regulators, were reported in three cases each. In total, we observed recurrent mutations in epigenetic regulators in
17 of 112 Ewing sarcoma cases (15.2%). As described above, recent data strongly suggest that STAG2 plays a major role in epigenetic insulation and may therefore be considered an epigenetic regulator. This finding reinforces the need for studies that clarify how mutations affecting the epigenetic landscape of Ewing sarcoma may cooperate with the EWSR1–ETS fusion to promote the development of overt Ewing sarcoma.

After submission of this article, Brohl and colleagues (46) published an article describing the genomic landscape of Ewing sarcoma based mostly on exome sequencing and RNA sequencing. The observed frequency of STAG2 of Ewing sarcoma based on the WGS dataset is similar to the findings reported in this article. They also observed the association of TP53 and STAG2 mutations. However, significant correlation with clinical outcome could not be demonstrated, possibly due to the smaller size of the patient cohort. Finally, the exclusive pattern of CDKN2A and STAG2 alterations was not reported in the Brohl and colleagues dataset. The different techniques used in the two reports and the different sizes of the patient series may account for this discrepancy, which requires further investigation.

In conclusion, our comprehensive genetic analysis of Ewing sarcoma identified recurrent mutations in STAG2, TP53, and epigenetic regulators. We showed that a STAG2 mutation gains prognostic significance when associated with TP53 mutations and that a STAG2-mutated subclone may expand during the course of the disease. Finally, the mutual exclusion between STAG2 and CDKN2A loss-of-function mutations suggests that these alterations may be, at least partially, redundant.

METHODS

Patients and Tumors

Our discovery cohort comprised 112 patients with Ewing sarcoma; both tumor and germline samples underwent WGS. All tumors selected for WGS were predicted to contain a large proportion of tumor cells based on pathology reports, previous CGH or SNP arrays, and/or a low Ct (cycle threshold) of the EWSR1–ETS fusion assessed by qRT-PCR. Eighteen Ewing sarcomas were obtained from the St. Jude tissue resource core facility for genome sequence analysis with St. Jude Institutional Review Board (IRB) approval for the Pediatric Cancer Genome Project. The remaining cases were those referred to Institut Curie from all over France for molecular diagnosis of Ewing sarcoma. Samples were stored in a tumor bank at the Institut Curie. The genetic study was approved by the IRB of the Institut Curie and by the Comité de Protection de Personnes Ile-de-France I (regional ethics committee; GenEwing no. IC 2009-02); specific informed consent was provided. Most patients were treated according to the EuroEwing protocol (47). An anonymization procedure was performed before compilation of clinical, histologic, and biologic information in a secure database with restricted access. All tumors included in this study were positive for the EWSR1–ETS fusion. Detailed clinicopathologic and sequencing information is provided in Fig. 1 and Supplementary Table S1.

The follow-up set comprising 199 tumor DNAs from EWSR1–ETS-positive Ewing sarcomas was distinct from the discovery set and consisted of patients treated according to the EuroEwing99 protocol.

Cell Lines

Ewing sarcoma cell lines were obtained from various sources: A673, RD-ES, SK-ES-1, and SK-NM-C from the ATCC; MHH-ES1 and TC-71 from the German Collection of Microorganisms and Cell Cultures (DSMZ); EW-1, EW-3, EW-7, EW-16, and EW-18 from the International Agency for Research on Cancer; STA-ET-1, STA-ET-3, and STA-ET-8 from Prof. Heinrich Kowar, Children’s Cancer Research Institute–Childhood (Vienna, Austria); and EW-22, EW-23, MIC, ORS, and POE from the Institut Curie. Cell lines were authenticated by their TP53 genotype, which included mutations previously described.

DNA and RNA Extraction

Nucleic acids were isolated from 10 to 25 mg of snap-frozen tumor tissue by standard proteinase K digestion and phenol or Trizol/chloroform extraction for genomic DNA and total RNA, respectively. Germline DNA was extracted from 2 mL of whole blood using the QuickGene610L Kit (FujiFILM) according to the manufacturer’s protocol. RT-PCR of tumor RNA using specific oligonucleotide primers and probe was performed as previously described (48).

WGS

WGS was performed by using the Illumina HiSeq2000 sequencing system (Illumina, Inc.). To prepare short-insert paired-end libraries, the TruSeq Sample Preparation Kit protocol (Illumina) was used with minor modifications. Briefly, 2.0 μg of genomic DNA was sheared on a Covaris E220 ultrasonicator (Covaris) and size-selected using AMPure XP beads (Agencourt; Beckman Coulter) to obtain fragments of approximately 450 bp. The fragmented DNA was end-repaired, adenylated, and ligated to Illumina-specific paired-end adaptors. Each library was sequenced in 2 × 101 bp paired-end mode on a HiSeq2000 flow-cell v3 instrument according to standard Illumina procedures, generating minimal average coverage of 35× for the tumor samples and 25× for the germline samples. Across the entire cohort, 96% of the genomic regions had ≥20× coverage.

Data are available in the European Genome-phenome Archive with the study accession numbers EGAS00001000855 (Institut Curie cohort) and EGAS00001000839 (St. Jude cohort).

Analysis of WGS Data

All samples were processed using the same analysis pipeline. Correspondence between sample and analysis numbers is indicated in Supplementary Table S2B. WGS mapping, coverage, and quality assessment; SNVs; detection of small indels; tier annotation for sequence mutations; and prediction of adverse effects of missense mutations were previously described (14, 15). SVs were analyzed by using the program CREST (13); CNAs were identified by comparing the read depth of matched tumor versus normal tissue and were analyzed by using the CONCERTING algorithm (COpy Number SEgmentation by Regression Tree In Next-Gen sequencing). The reference human genome assembly GRCh37-lite was used to map all samples. We used the program cghMCR to identify recurrent copy-number gain or loss. For this analysis, we excluded 6 cases that showed a highly fragmented copy-number variation (CNV) profile across the genome. These cases have a high number of CNAs across the genome not supported by corresponding SVs. In our experience, these readings are artifacts caused by library construction. The six cases had a total of 182,433 CNV segments, compared with the 16,354 in the remaining 108 cases. We also excluded 3 cases with likely low tumor cell content as suggested by the low number of SNVs and mutation-supporting reads (Supplementary Table S2A). We also excluded from this analysis CNVs identified in the T-cell receptor loci that are caused by physiologic rearrangements of T cells in germline samples. Thus, 16,036 CNV segments from 108 cases were used as input for this analysis. Tumor purity was estimated using loss of heterozygosity, copy-number change, and mutant allele fraction of SNVs, as previously described (25).

The background mutation rate was calculated by dividing the number of high-quality tier-3 SNVs by the total length of tier-3
regions covered at least 10× (Supplementary Table S1). The SVs detected within EWSR1 or ETS genes were consistent with the gene fusions defined by RT-PCR results in all cases. The only case (case S001301) that could not be investigated by RT-PCR was shown by WGS to harbor an EWSR1–FLI1 rearrangement. EWSR1 and ETS genomic rearrangements were undetectable despite positive RT-PCR results in six samples. Three of these six cases had low tumor purity, and one had uneven sequencing coverage.

Chromothripsis was analyzed using the criteria proposed by Korbel and Campbell (49). Oscillating patterns of copy-number states were manually inspected using the Circos plots, and statistical tests were applied to evaluate clustering of breakpoints and randomness of DNA fragment joins. For detecting clustering of breakpoints, we applied the Bartlett goodness-of-fit test for exponential distribution to see if there was a strong departure from the null hypothesis, consistent with the chromothripsis hypothesis. For assessing randomness of DNA fragment joins, we applied the goodness-of-fit tests to evaluate if there was no significant departure from the multinomial distribution with equal probabilities, consistent with the chromothripsis hypothesis. This test was carried out for both intrachromosomal and interchromosomal breakpoints when applicable.

Validation Sequencing

For 18 tumor samples (SJEWS001301–1320), the genomic coordinates of putative alterations identified by WGS, including SNVs, SVs, and indels, were used to generate a NimbleGen SeqCap EZ bait set for enrichment of targeted regions (Roche). The baits were hybridized to Truseq sample libraries (Illumina) prepared from amplified genomic DNA (Roche). Pooled samples were sequenced on a HiSeq 2000 by using the paired-end multiplexed 100-cycle protocol. Resulting data were converted to FASTQ files by using CASAVA 1.8.2 (Illumina) and mapped with the Burrows-Wheeler Aligner before pipeline analysis. Of the 6,659 somatic SNVs identified in the 18 cases, we were able to design a validation assay by custom capture for 6,042. Of these, 5,779 were validated as somatic mutations (overall validation rate, 95.6%).

In addition, a custom TrueSeq Amplicon (Illumina) focused on the whole coding sequence was designed for 8 genes that exhibited at least two somatic changes in the discovery cohort (STAG2, TP53, RYR2, MACF1, DIRAS1, CDKN2A, PCDH10, and CREBBP). Sequencing libraries were prepared following the manufacturer’s protocol, and barcoded amplicons were multiplexed and sequenced on Illumina HiSeq platforms. The mean number of mapped reads was 6.1 million. All mutations detected by WGS were confirmed.

CDKN2A status of cell lines was verified by PCR amplification spanning all four CDKN2A exons, as previously described (50), followed by the Sanger sequencing of the amplicons. Across the following cohort, CDKN2A status was determined by real-time PCR on genomic DNA. Two sets of primers (Exon1a and Exon2) were used to detect CDKN2A: CDKN2A-ex1a_F: GCCTGGGCTGGTCACCAAGA, CDKN2A-ex1a_R: AGTGGGCGCCCTGCGCTGATCGCTGCT-3′, CDKN2A-ex2a_F: GCCTTTCACAAAGGTTCTTTC, CDKN2A-ex2a_R: TCCG CCTACAGAAAGCAGATGG, and were respectively normalized to TGFBR2 genomic levels (located on chr 18, the most stable chromosome across the Ewing sarcoma WGS cohort): TGFBR2_F: GCCAA ATCTTGGGTCTCAGCAAGA, TGFBR2_R: 3′Yakima-Yellow CGGCTGTGCACTGAGGAGTAAG-5′BHQ1, TGFBR2_R: AAAGT GCCCTGGGACCTGGCTGATAA. Duplex PCR (CDKN2A/TGFBR2) using TaqMan Universal PCR Master Mix (Life Technologies) was performed, and the mean of normalized CDKN2A Exon1a and Exon2 was calculated (CDKN2A/ TGFBR2). Ten nontumorigenic germ line DNAs were used as controls (average CDKN2A/ TGFBR2 value set to 1). Eleven tumor samples with loss of CDKN2A and 38 CDKN2A wild-type tumor samples from the WGS cohort were used to set CDKN2A/ TGFBR2 heterozygous threshold to 0.7 (no false positive for wild-type CDKN2A samples).

Statistical Analysis

Overall survival was defined as the time from diagnosis to disease-related death or last follow-up. Survival curves were analyzed according to the Kaplan–Meier method and compared using the log-rank test. We used the SMG test in the MuSiC suite (21) to identify genes significantly enriched for somatic SNVs and indel mutations. This test assigns mutations to seven categories (AT transition, AT transversion, CG transition, CG transversion, CpG transition, CpG transversion, and indel) and then uses statistical methods based on convolution, hypergeometric distribution (Fisher test), and likelihood to combine the category-specific binomials and obtain overall P values. Genes with false discovery rate <0.1 in 2 of the 3 statistical tests were considered significantly mutated.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors’ Contributions


Writing, review, and/or revision of the manuscript: F. Tirode, D. Surdez, M.C. Le Deley, A. Bahrami, M. Gut, E.R. Mardis, R.K. Wilson, P. Marec-Bérard, I. Gut, J. Downing, M. Dyer, J. Zhang, O. Delattre

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): D. Surdez, S. Grossetête-Lalami, M. Rusch, E. Hedlund, G. Pierron, B. Vadodaria, M. Dyer, J. Easton, J. Downing, J. Zhang, O. Delattre

Study supervision: R.K. Wilson, J. Downing, M. Dyer, J. Zhang, O. Delattre

Acknowledgments

The authors thank Fabien Calvo for continuous support, Virginie Chêne, Stelly Ballet, Heather Mulder, Panduka Nagahawatte, Donald Yergeau, Yongjin Li, Michael Edmonson, Andrew Thrasher, and Carlo Lucchesi for their invaluable help, Peter Brooks for fruitful discussions, Heinrich Kovar for some of the cell lines, Alban Lermine, Nicolas Servant, Philippe Hupé, and Emmanuel Barillot for their help in processing the Next Generation Sequencing data, and UNICANCER for providing access to the clinical databases.

Grant Support

This work was supported in part by a Cancer Center Support grant (CA21765) and grants to M. Dyer (EY014867, EY018599, and CA168875) from the U.S. National Institutes of Health, and by the American Lebanese Syrian Associated Charities (ALASC). M. Dyer is a Howard Hughes Medical Institute Investigator. WGS was supported as part of the St. Jude Children’s Research Hospital–Washington University Pediatric Cancer Genome Project.

The WGS of French cases was performed by the Centro Nacional de Análisis Genómico (CNAG) in Barcelona and supported by grants from the French National Cancer Institute InCa within the framework of the International Cancer Genome Consortium program. Additional sequencing was performed by the next-generation sequencing platform of the Institut Curie, supported by grants ANR-10-EQPX-03 and ANR10-INBS-09-08 from the Agence Nationale de la Recherche (investissements d’avenir) and by the Cancéropôle Ile-de-France.

This work was also supported by grants from the Ligue Nationale Contre Le Cancer (Equipe labellisée), and by European PROVABES (ERA-NET TRANSCAN JTC-2011), ASSET (FP7-HEALTH-2010-259348), and EEC (HEALTH-F2-2013-602856) grants. D. Surdez is supported by the Institut Curie–SIRIC (Site de Recherche Intégrée en Cancérologie) program.

The authors also thank the following associations for their invaluable support: the Société Française des Cancers de l’Enfant, Courir pour Courir; the Société Française des Cancers de l’Enfant, les Amis de Claire; Mathieu, Dans les pas du Géant, Olivier Chape, Les Bagouzamanon; and EEC (HEALTH-F2-2013-602856) grants. D. Surdez is supported by grants ANR-10-EQPX-03 and by the Canceropôle Ile-de-France.

During this study, a French sequencing facility was implemented to allow support: the Société Française des Cancers de l’Enfant, Courir pour Courir; the Société Française des Cancers de l’Enfant, les Amis de Claire; Mathieu, Dans les pas du Géant, Olivier Chape, Les Bagouzamanon; and EEC (HEALTH-F2-2013-602856) grants. D. Surdez is supported by grants ANR-10-EQPX-03 and by the Canceropôle Ile-de-France.

The authors also thank the following associations for their invaluable support: the Société Française des Cancers de l’Enfant, Courir pour Courir; the Société Française des Cancers de l’Enfant, les Amis de Claire; Mathieu, Dans les pas du Géant, Olivier Chape, Les Bagouzamanon; and EEC (HEALTH-F2-2013-602856) grants. D. Surdez is supported by grants ANR-10-EQPX-03 and by the Canceropôle Ile-de-France.

The authors also thank the following associations for their invaluable support: the Société Française des Cancers de l’Enfant, Courir pour Courir; the Société Française des Cancers de l’Enfant, les Amis de Claire; Mathieu, Dans les pas du Géant, Olivier Chape, Les Bagouzamanon; and EEC (HEALTH-F2-2013-602856) grants. D. Surdez is supported by grants ANR-10-EQPX-03 and by the Canceropôle Ile-de-France.

The authors also thank the following associations for their invaluable support: the Société Française des Cancers de l’Enfant, Courir pour Courir; the Société Française des Cancers de l’Enfant, les Amis de Claire; Mathieu, Dans les pas du Géant, Olivier Chape, Les Bagouzamanon; and EEC (HEALTH-F2-2013-602856) grants. D. Surdez is supported by grants ANR-10-EQPX-03 and by the Canceropôle Ile-de-France.

The authors also thank the following associations for their invaluable support: the Société Française des Cancers de l’Enfant, Courir pour Courir; the Société Française des Cancers de l’Enfant, les Amis de Claire; Mathieu, Dans les pas du Géant, Olivier Chape, Les Bagouzamanon; and EEC (HEALTH-F2-2013-602856) grants. D. Surdez is supported by grants ANR-10-EQPX-03 and by the Canceropôle Ile-de-France.

The authors also thank the following associations for their invaluable support: the Société Française des Cancers de l’Enfant, Courir pour Courir; the Société Française des Cancers de l’Enfant, les Amis de Claire; Mathieu, Dans les pas du Géant, Olivier Chape, Les Bagouzamanon; and EEC (HEALTH-F2-2013-602856) grants. D. Surdez is supported by grants ANR-10-EQPX-03 and by the Canceropôle Ile-de-France.

The authors also thank the following associations for their invaluable support: the Société Française des Cancers de l’Enfant, Courir pour Courir; the Société Française des Cancers de l’Enfant, les Amis de Claire; Mathieu, Dans les pas du Géant, Olivier Chape, Les Bagouzamanon; and EEC (HEALTH-F2-2013-602856) grants. D. Surdez is supported by grants ANR-10-EQPX-03 and by the Canceropôle Ile-de-France.

The authors also thank the following associations for their invaluable support: the Société Française des Cancers de l’Enfant, Courir pour Courir; the Société Française des Cancers de l’Enfant, les Amis de Claire; Mathieu, Dans les pas du Géant, Olivier Chape, Les Bagouzamanon; and EEC (HEALTH-F2-2013-602856) grants. D. Surdez is supported by grants ANR-10-EQPX-03 and by the Canceropôle Ile-de-France.
Comprehensive Genomic Analysis of Ewing Sarcoma


38.itcher M, Emerson BM. Epigenetic silencing of the p16(INK4a) tumor suppressor is associated with loss of CTCF binding and a chromatin boundary. Mol Cell 2009;34:271–84.


Genomic Landscape of Ewing Sarcoma Defines an Aggressive Subtype with Co-Association of \textit{STAG2} and \textit{TP53} Mutations

Franck Tirode, Didier Surdez, Xiaotu Ma, et al.

\textit{Cancer Discovery} Published OnlineFirst September 15, 2014.

**Updated version**
Access the most recent version of this article at:
doi:10.1158/2159-8290.CD-14-0622

**Supplementary Material**
Access the most recent supplemental material at:
http://cancerdiscovery.aacrjournals.org/content/suppl/2014/09/13/2159-8290.CD-14-0622.DC1

**E-mail alerts**
Sign up to receive free email-alerts related to this article or journal.

**Reprints and Subscriptions**
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

**Permissions**
To request permission to re-use all or part of this article, use this link http://cancerdiscovery.aacrjournals.org/content/early/2014/10/15/2159-8290.CD-14-0622.
Click on "Request Permissions" which will take you to the Copyright Clearance Center’s (CCC) Rightslink site.