Published in final edited form as:
Cancer Discov. 2014 November ; 4(11): 1342–1353. doi:10.1158/2159-8290.CD-14-0622.

Genomic landscape of Ewing sarcoma defines an aggressive subtype with co-association of *STAG2* and *TP53* mutations

Franck Tirode[1,2,*], **Didier Surdez**[1,2,*], **Xiaotu Ma**[3], **Matthew Parker**[3], **Marie Cécile Le Deley**[4], **Armita Bahrami**[5], **Zhaojie Zhang**[3], **Eve Lapouble**[6], **Sandrine Grossetête-Lalami**[1,2], **Michael Rusch**[3], **Stéphanie Reynaud**[6], **Thomas Rio-Frio**[2], **Erin Hedlund**[3], **Gang Wu**[3], **Xiang Chen**[3], **Gaelle Pierron**[6], **Odile Oberlin**[7], **Sakina Zaidi**[1,2], **Gordon Lemmon**[3], **Pankaj Gupta**[3], **Bhavin Vadodaria**[8], **John Easton**[8], **Marta Gut**[9], **Li Ding**[10,11,12], **Elaine R. Mardis**[10,11,12], **Richard K. Wilson**[10,11,12], **Sheila Shurtleff**[5], **Valérie Laurence**[13], **Jean Michon**[14], **Perrine Marec-Bérard**[15], **Ivo Gut**[9], **James Downing**[8], **Michael Dyer**[16,17], **Jinghui Zhang**[3,**], and **Olivier Delattre**[1,2,**] **for the St. Jude Children's Research Hospital – Washington University Pediatric Cancer Genome Project and the International Cancer Genome Consortium**

[1]INSERM U830, Laboratory of Genetics and Cancer Biology, Institut Curie, 26 rue d'Ulm, 75248, Paris cedex 05

[2]Centre de Recherche, Institut Curie, 26 rue d'Ulm, 75248, Paris cedex 05

[3]Department of Computational Biology, St. Jude Children's Research Hospital, Memphis, Tennessee 38105, USA

[4]Departement d'Epidémiologie et de Biostatistiques, Gustave Roussy, 114 rue Edouard Vaillant, 94805 Villejuif, France

[5]Department of Pathology, St. Jude Children's Research Hospital, Memphis, Tennessee 38105, USA

[6]Unité de Génétique Somatique, Adolescents et Jeunes Adultes, Centre Hospitalier, Institut Curie, 26 rue d'Ulm, 75248, Paris cedex 05

[7]Departement de Pédiatrie, Gustave Roussy, 114 rue Edouard Vaillant, 94805 Villejuif, France

[8]The Pediatric Cancer Genome Laboratory, St. Jude Children's Research Hospital, Memphis, Tennessee 38105, USA

[9]Centro Nacional de Análisis Genómico (CNAG), Barcelona, 080028, Spain

[10]The Genome Institute, Department of Genetics, Washington University School of Medicine in St. Louis, St. Louis, Missouri 63108, USA

[11]Department of Medicine, Washington University School of Medicine in St. Louis, St. Louis, Missouri 63108, USA

[**]Co-corresponding authors: Olivier Delattre, Inserm U830 Laboratory of Cancer Genetics and Biology, Institut Curie - Centre de Recherche, 26 rue d'Ulm, 75248 Paris cedex 05, Phone: (33) 1 56 24 66 79, olivier.delattre@curie.fr. Jinghui Zhang, Department of Computational Biology, MS 1160, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, TN, 38105-3678, USA, Phone: (901) 595-6829, jinghui.zhang@stjude.org.
[*]These two authors contributed equally

Conflict of interest: The authors disclose no potential conflicts of interest

[12]Siteman Cancer Center, Washington University School of Medicine in St. Louis, St. Louis, Missouri 63108, USA

[13]Département d'Oncologie Medicale, Adolescents et Jeunes Adultes, Centre Hospitalier, Institut Curie, 26 rue d'Ulm, 75248, Paris cedex 05

[14]Département d'Oncologie Pediatrique, Adolescents et Jeunes Adultes, Centre Hospitalier, Institut Curie, 26 rue d'Ulm, 75248, Paris cedex 05

[15]Institute for Paediatric Haematology and Oncology, Leon Bérard Cancer Centre, University of Lyon, 1 Place Joseph Renaut, 69008 Lyon, France

[16]Developmental Neurobiology, St. Jude Children's Research Hospital, Memphis, Tennessee 38105, USA

[17]Howard Hughes Medical Institute, Chevy Chase, Maryland 20815, USA

Abstract

Ewing sarcoma is a primary bone tumor initiated by *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 *STAG2* (17%), *CDKN2A* (12%), *TP53* (7%), *EZH2*, *BCOR*, and *ZMYM3* (2.7% each). Strikingly, *STAG2* mutations and *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 *STAG2* and *TP53* mutations are often concurrent and are associated with poor outcome. Finally, we detected subclonal *STAG2* mutations in diagnostic tumors and expansion of STAG2 immuno-negative cells in relapsed tumors as compared with matched diagnostic samples.

Keywords

Ewing sarcoma; genomics; mutations; whole genome sequencing; prognostic

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 reference 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%–15% of cases, *EWSR1* is fused to 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*, 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 reference 1).

Chromosome or array-based comparative genomic hybridization (CGH) as well as single-nucleotide polymorphism (SNP) arrays have identified recurrent DNA copy-number alterations 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%) (12). While 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 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 of 112 tumors and their matched germline DNA. The most frequent point mutations involved the *STAG2* and *TP53* genes, and the prognostic significance of these mutations was further demonstrated in a series of 299 cases. *STAG2* mutations were significantly associated with the occurrence of structural variations and were mutually exclusive with *CDKN2A* deletions. In some cases, we also observed a small number of *STAG2*-deficient tumor cells that survived treatment and comprised the major clone in the recurrent tumors.

RESULTS

Ewing sarcoma has low numbers of single-nucleotide and structural variants

Our discovery set for whole-genome sequencing (WGS) comprised 112 Ewing sarcomas with matched germline DNA (Fig. 1, Supplementary Table S1). All Ewing sarcoma tumors, with the exception of one (SJ001301) that had insufficient tumor sample for analysis, expressed *EWSR1–ETS* fusions: *EWSR1–FLI1* in 101 cases, *EWSR1–ERG* in 9 cases and *EWSR1–ETV1* in one case. Tumor and germline DNA were sequenced at a median depth of 35X and 25X, respectively. Mapping, detection, and annotation of single-nucleotide variants (SNV), indels, and structural variants (SV), functional predictions, and copy-number alterations (CNA) were computed from the whole-genome sequencing 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 to 66) per tumor (Supplementary Table S2b). In most cases (106/112; 95%), WGS detected SVs within the previously described *EWSR1* and *ETS* chromosome breakpoint regions (16) (Fig. 1, Supplementary Table S2c, Supplementary Fig. S1A–D). Five cases (SJ001303, SJ001320, IC198, IC273, IC086) exhibited chromothripsis, including three cases with chromothripsis on chr 21 and 22 associated with *EWSR1–ERG* fusions (SJ001303, IC198 and IC273) and one case involving chr 22 associated with an *EWSR1–FLI1* fusion (SJ001320). Copy-number alterations (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 1 (19/103; 18%), deletion of the long arm of chr 16 (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 ($P = 2\times10^{-5}$ 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 ($p=1.63\times10^{-3}$, Fisher exact test), but neither chr 8 nor chr 12 gains, nor their combination, was correlated with shorter survival (data not shown).

Experimental validation by custom capture and Illumina sequencing of all WGS predicted SNVs and indels on 19 cases showed a 95.6% verification rate. Across the entire cohort, the median number of SNVs was 319 (range, 13 to 1747) per genome (Supplementary Table S2e and 2f). 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 to 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 ($r^2= 0.42$, $P=2.7\times10^{-5}$, Pearson correlation). Patients older than 20 years at diagnosis had significantly more SNVs than did younger patients ($P = 0.001$, Mann-Whitney 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 eight 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 cohesion complex (18) and was found to be associated with aneuploidy (12), we also investigated the relation of *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 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 (*TTN, CSMD1, MACF1, RYR2)* (19), the third most frequently mutated genes were *EZH2*, *BCOR*, and *ZMYM3*, which each presented with 3 mutations (3/112, 2.7%,) (Fig. 1 and Supplementary Table S2). All three *EZH2* mutations were missense mutations within the SET domain (Y646F, Y646H, and A682G according to NM_004456). *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, Supplementary Table S2). *ZMYM3* exhibited two indels (L82fs according to NM_201599) and one 17-kb intragenic deletion (Fig. 1, 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 *EZH2*, *BCOR*, and *ZMYM3*, we identified novel somatic mutations in *SETD2*, *MLL2*, *MLL3*, and *PRDM9* (Fig. 1, Supplementary Table S2). Of note, two novel missense mutations were observed in *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).

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). The exclusive pattern of *STAG2* and *CDKN2A* alterations shown in primary tumors (Fig. 1) was fully replicated in the cell lines (Table 1). Across the 15 cell lines that could be investigated by western blot, all cases with *STAG2* mutations but one (MHH-ES-1) expressed p16. Reciprocally, all cases with *CDKN2A* deletion expressed STAG2 (Supplementary Fig. S4A–B). 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 Ewing sarcoma patients. 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 ($p=4\times10^{-4}$, Logrank test). In total, forty-one 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 was 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 3E). 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 ($P = 2.4\times10^{-4}$, 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's test: 0.0076, *STAG2/CDKN2A*, 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–C).

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 (SJEWS001303) or detected at a subclonal level at diagnosis (SJEWS014721) (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, pathological, and molecular findings. The background mutation rate of Ewing sarcoma was relatively low (2.4×10^{-7}), 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 array-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 gain have strong negative prognostic significance.

The most frequently mutated gene in Ewing sarcoma is *STAG2*. *STAG1* and *STAG2*, the human orthologs of yeast *Scc3p*, are components of the cohesin multiprotein complex that plays an essential role in sister chromatin 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 copy-number alterations 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 considered, the positive correlation between *STAG2* mutation and the number of SVs is no longer significant. The analysis of survival data must also take into account the association between *STAG2* and *TP53* mutations. Indeed, in our extended series of patients, the prognostic significance of *STAG2* mutation appears to be strongly dependent on the co-existence of a *TP53* mutation. The prognosis of cases with both *STAG2* and *TP53* mutations appears particularly unfavorable (Fig. 2E). Together, these data suggest that *STAG2* and *TP53* mutation may cooperate to increase genetic instability in a particularly aggressive subtype of Ewing sarcoma. Consistent with this hypothesis, it is noteworthy that *STAG2* and *TP53* mutations are much more frequent in cell lines derived mainly from aggressive cases. Finally, our results suggest that *STAG2*-mutated Ewing sarcoma subclones at diagnosis may evolve and become the major clone at recurrence. Further investigation of the relation of clonal expansion to tumor progression or response to therapy will be of great interest.

We observed a previously unreported, mutually exclusive pattern of *STAG2* and *CDKN2A* mutation 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 mechanisms 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 17/112 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 manuscript, Brohl and colleagues published an article (46) describing the genomic landscape of Ewing sarcoma based mostly on exome sequencing and RNA-seq. The observed frequency of *STAG2*, *TP53*, and *CDKN2A* is similar to the findings reported in the present manuscript. 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 is not reported in the Brohl et al 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 suggest that these alterations may be, at least partially, redundant.

MATERIALS AND METHODS

Patients and tumors

Our discovery cohort comprised 112 patients with Ewing sarcoma; both tumor and germline samples underwent whole-genome sequencing. 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 *EWSR1–ETS* fusion assessed by RT-QPCR. 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 (PCGP). 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 Institutional Review Board of the Institut Curie (Paris, France) and by the Comité de Protection de Personnes Ile-de-France I (regional ethics committee; GenEwing n° 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 biological 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 American Type Culture Collection; 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, Lyon, France; STA-ET-1, STA-ET-3, and STA-ET-8 from Prof. Heinrich Kovar, Children's Cancer Research Institute–Childhood, Vienna, Austria; and EW-22, EW-23, MIC, ORS, and POE from the Institut Curie, Paris, France. 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 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 by using the QuickGene610L kit (FujiFILM; Singapore) according to the manufacturer's protocol. RT-PCR of tumor RNA using specific oligonucleotide primers and probe was performed as previously described (48).

Whole-genome sequencing

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; MA, USA) and size-selected using AMPure XP beads (Agencourt, Beckman Coulter) to obtain fragments of ~450bp. 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 35x for the tumor samples and 25x for the germline samples. Across the entire cohort, 96% of the genomic regions had ≥20x coverage.

Data are available in the European Genome-phenome Archive with the study accession number: 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; single-nucleotide variants (SNVs); detection of small insertions or deletions (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 vs. normal tissue and were analyzed by using the CONSERTING 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 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 locus that are caused by physiological 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 (LOH), copy number change, and mutant allele fraction (MAF) of SNVs, as previously described (25).

The background mutation rate (BMR) was calculated by dividing the number of high-quality tier-3 SNVs by the total length of tier-3 regions covered at least 10X (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 SJ001301) that could not be investigated by RT-PCR was shown by WGS to harbor a *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). Patterns of oscillating pattern 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's 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 Nimbelgen 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 (BWA) prior to pipeline analysis. Of the 6659 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 TrueSeqAmplicon (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*, *SPTBN5*, *PCDH10* and *CREBRF*). Sequencing libraries were prepared following manufacturer's protocol and barcoded amplicons were multiplexed and sequenced on Illumina2500 HiSeqFast flow cells; mean target coverage was 98% and 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 Sanger sequencing of the amplicons. Across the follow-up 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: GGCTGGCTGGTCACCAGA, CDKN2A-ex1A_probe: FAM 5′-ATGGAGCCTTCGGCTGACTGGCT-3′ BHQ1, CDKN2A-ex1A_R: CGCCCGCACCTCCTCTAC; CDKN2A-ex2_F: GGCTCTACACAAGCTTCCTTTCC, CDKN2A-ex2_probe: FAM 5′-CATGCCGGCCCCCACCCT-3′ BHQ1, CDKN2A-ex2_R: CCTGCCAGAGAGAACAGAATGG and were respectively normalized to *TGFBR2* genomic levels (located on chr 3, the most stable chromosome across the Ewing sarcoma WGS cohort): TGFBR2_F: GCAAATCTGGTTGCCCTAGCAAGA, TGFBR2_probe: 5′Yakima-Yellow-CCCGTTTGCACATGAGAGGGTAAGT-3′BHQ1, TGFBR2_R:

AAAGTGGGTTGGGAGTCACCTGAA. Duplex PCR (*CDKN2A*/*TGFBR2*) using TaqMan Universal PCR Master Mix (Life technologies) was performed and the mean of normalized *CDKN2A* Exon1a and Exon2 were calculated ($CDKN2A^{EX1A-2}$). Ten nontumorigenic germline DNAs were used as controls (Average $CDKN2A^{EX1A-2}$ 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 $CDKN2A^{EX1A-2}$ 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's test), and likelihood to combine the category-specific binomials and obtain overall *P* values. Genes with FDR (false discovery rate) < 0.1 in 2 of the 3 statistical tests were considered significantly mutated.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Financial support: This work was supported in part by a Cancer Center Support grant (CA21765) and grants to M.A.D (EY014867, EY018599, CA168875) from the U.S. National Institutes of Health, and by the American Lebanese Syrian Associated Charities (ALSAC). M.A.D. is a Howard Hughes Medical Institute Investigator. Whole-genome sequencing was supported as part of the St. Jude Children's Research Hospital –Washington University Pediatric Cancer Genome Project.

The whole genome sequencing of French cases was performed by the CNAG in Barcelona and supported by grants from the French National Cancer Institute INCa within the framework of the ICGC 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 Canceropôle Ile-de-France.

This work was also supported by grants from the Ligue Nationale Contre Le Cancer (Equipe labellisée), by European PROVABES [ERA-NET TRANSCAN JTC-2011], ASSET [FP7-HEALTH-2010-259348] and EEC [HEALTH-F2-2013-602856] grants. We also thank the following associations for their invaluable support: the Société Française des Cancers de l'Enfant, Courir pour Mathieu, Dans les pas du Géant, Olivier Chape, Les Bagouzamanon, Enfants et Santé and les Amis de Claire. Didier Surdez is supported by the Institut Curie –(SIRIC) (Site de Recherche Intégrée en Cancérologie) program. We are indebted to UNICANCER for providing access to the clinical databases.

We thank Fabien Calvo for continuous support. We thank 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 and Alban Lermine, Nicolas Servant, Philippe Hupé, and Emmanuel Barillot for their help in processing the NGS data.

We thank the following clinicians and pathologists for providing samples used in this work:

I. Aerts, P. Anract, C. Bergeron, L. Boccon-Gibod, F. Boman, F. Bourdeaut, C. Bouvier, R. Bouvier, L. Brugières, E. Cassagnau, J. Champigneulle, C. Cordonnier, J. M. Coindre, N. Corradini, A. Coulomb-Lhermine, A. De Muret, G. De Pinieux, A.S. Defachelles, A. Deville, F. Dijoud, F. Doz, C. Dufour, K. Fernandez, N. Gaspard, L. Galmiche-Rolland, C. Glorion, A. Gomez-Brouchet, J.M. Guinebretière, H. Jouan, C. Jeanne-Pasquier, B. Kantelip, F. Labrousse, V. Laithier, F. Larousserie, G. Leverger, C. Linassier, P. Mary, G. Margueritte, E. Mascard, A. Moreau, J. Michon, C. Michot, F. Millot, Y. Musizzano, M. Munzer, B. Narciso, O. Oberlin, D. Orbach, H. Pacquement, Y. Perel, B. Petit, M. Peuchmaur, J.Y. Pierga, C. Piguet, S. Piperno-Neumann, E. Plouvier, D. Ranchere-Vince, J. Rivel, C. Rouleau, H. Rubie, H. Sartelet, G. Schleiermacher, C. Schmitt, N. Sirvent, D. Sommelet, P. Terrier, R. Tichit, J. Vannier, J. M. Vignaud, V. Verkarre.

References

1. Lessnick SL, Ladanyi M. Molecular pathogenesis of Ewing sarcoma: new therapeutic and transcriptional targets. Annu Rev Pathol. 2012; 7:145–59. [PubMed: 21942527]
2. Delattre O, Zucman J, Plougastel B, Desmaze C, Melot T, Peter M, et al. Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours. Nature. 1992; 359:162–5. [PubMed: 1522903]
3. May WA, Lessnick SL, Braun BS, Klemsz M, Lewis BC, Lunsford LB, et al. The Ewing's sarcoma EWS/FLI-1 fusion gene encodes a more potent transcriptional activator and is a more powerful transforming gene than FLI-1. Mol Cell Biol. 1993; 13:7393–8. [PubMed: 8246959]
4. Armengol G, Tarkkanen M, Virolainen M, Forus A, Valle J, Bohling T, et al. Recurrent gains of 1q, 8 and 12 in the Ewing family of tumours by comparative genomic hybridization. Br J Cancer. 1997; 75:1403–9. [PubMed: 9166930]
5. Brisset S, Schleiermacher G, Peter M, Mairal A, Oberlin O, Delattre O, et al. CGH analysis of secondary genetic changes in Ewing tumors: correlation with metastatic disease in a series of 43 cases. Cancer Genet Cytogenet. 2001; 130:57–61. [PubMed: 11672775]
6. Brownhill SC, Taylor C, Burchill SA. Chromosome 9p21 gene copy number and prognostic significance of p16 in ESFT. Br J Cancer. 2007; 96:1914–23. [PubMed: 17533400]
7. Hattinger CM, Pötschger U, Tarkkanen M, Squire J, Zielenska M, Kiuru-Kuhlefelt S, et al. Prognostic impact of chromosomal aberrations in Ewing tumours. Br J Cancer. 2002; 86:1763–9. [PubMed: 12087464]
8. Kovar H, Jug G, Aryee DN, Zoubek A, Ambros P, Gruber B, et al. Among genes involved in the RB dependent cell cycle regulatory cascade, the p16 tumor suppressor gene is frequently lost in the Ewing family of tumors. Oncogene. 1997; 15:2225–32. [PubMed: 9393981]
9. Mackintosh C, Ordóñez JL, García-Domínguez DJ, Sevillano V, Llombart-Bosch A, Szuhai K, et al. 1q gain and CDT2 overexpression underlie an aggressive and highly proliferative form of Ewing sarcoma. Oncogene. 2012; 31:1287–98. [PubMed: 21822310]
10. Ozaki T, Paulussen M, Poremba C, Brinkschmidt C, Rerin J, Ahrens S, et al. Genetic imbalances revealed by comparative genomic hybridization in Ewing tumors. Genes Chromosomes Cancer. 2001; 32:164–71. [PubMed: 11550284]
11. Huang H-Y, Illei PB, Zhao Z, Mazumdar M, Huvos AG, Healey JH, et al. Ewing sarcomas with p53 mutation or p16/p14ARF homozygous deletion: a highly lethal subset associated with poor chemoresponse. J Clin Oncol. 2005; 23:548–58. [PubMed: 15659501]
12. Solomon DA, Kim T, Diaz-Martinez LA, Fair J, Elkahloun AG, Harris BT, et al. Mutational inactivation of STAG2 causes aneuploidy in human cancer. Science. 2011; 333:1039–43. [PubMed: 21852505]
13. Wang J, Mullighan CG, Easton J, Roberts S, Heatley SL, Ma J, et al. CREST maps somatic structural variation in cancer genomes with base-pair resolution. Nat Methods. 2011; 8:652–4. [PubMed: 21666668]
14. Zhang J, Benavente CA, McEvoy J, Flores-Otero J, Ding L, Chen X, et al. A novel retinoblastoma therapy from genomic and epigenetic analyses. Nature. 2012; 481:329–34. [PubMed: 22237022]
15. Zhang J, Ding L, Holmfeldt L, Wu G, Heatley SL, Payne-Turner D, et al. The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. Nature. 2012; 481:157–63. [PubMed: 22237106]

16. Zucman J, Melot T, Desmaze C, Ghysdael J, Plougastel B, Peter M, et al. Combinatorial generation of variable fusion proteins in the Ewing family of tumours. EMBO J. 1993; 12:4481–7. [PubMed: 8223458]
17. Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, et al. Mutational processes molding the genomes of 21 breast cancers. Cell. 2012; 149:979–93. [PubMed: 22608084]
18. Remeseiro S, Cuadrado A, Losada A. Cohesin in development and disease. Dev Camb Engl. 2013; 140:3715–8.
19. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature. 2013; 499:214–8. [PubMed: 23770567]
20. Huether R, Dong L, Chen X, Wu G, Parker M, Wei L, et al. The landscape of somatic mutations in epigenetic regulators across 1,000 paediatric cancer genomes. Nat Commun. 2014; 5:3630. [PubMed: 24710217]
21. Dees ND, Zhang Q, Kandoth C, Wendl MC, Schierding W, Koboldt DC, et al. MuSiC: identifying mutational significance in cancer genomes. Genome Res. 2012; 22:1589–98. [PubMed: 22759861]
22. Molenaar JJ, Koster J, Zwijnenburg DA, van Sluis P, Valentijn LJ, van der Ploeg I, et al. Sequencing of neuroblastoma identifies chromothripsis and defects in neuritogenesis genes. Nature. 2012; 483:589–93. [PubMed: 22367537]
23. Pugh TJ, Morozova O, Attiyeh EF, Asgharzadeh S, Wei JS, Auclair D, et al. The genetic landscape of high-risk neuroblastoma. Nat Genet. 2013; 45:279–84. [PubMed: 23334666]
24. Sausen M, Leary RJ, Jones S, Wu J, Reynolds CP, Liu X, et al. Integrated genomic analyses identify ARID1A and ARID1B alterations in the childhood cancer neuroblastoma. Nat Genet. 2013; 45:12–7. [PubMed: 23202128]
25. Chen X, Stewart E, Shelat AA, Qu C, Bahrami A, Hatley M, et al. Targeting oxidative stress in embryonal rhabdomyosarcoma. Cancer Cell. 2013; 24:710–24. [PubMed: 24332040]
26. Shern JF, Chen L, Chmielecki J, Wei JS, Patidar R, Rosenberg M, et al. Comprehensive genomic analysis of rhabdomyosarcoma reveals a landscape of alterations affecting a common genetic axis in fusion-positive and fusion-negative tumors. Cancer Discov. 2014; 4:216–31. [PubMed: 24436047]
27. Pugh TJ, Weeraratne SD, Archer TC, Pomeranz Krummel DA, Auclair D, Bochicchio J, et al. Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations. Nature. 2012; 488:106–10. [PubMed: 22820256]
28. Robinson G, Parker M, Kranenburg TA, Lu C, Chen X, Ding L, et al. Novel mutations target distinct subgroups of medulloblastoma. Nature. 2012; 488:43–8. [PubMed: 22722829]
29. Jones DTW, Hutter B, Jäger N, Korshunov A, Kool M, Warnatz H-J, et al. Recurrent somatic alterations of FGFR1 and NTRK2 in pilocytic astrocytoma. Nat Genet. 2013; 45:927–32. [PubMed: 23817572]
30. Schwartzentruber J, Korshunov A, Liu X-Y, Jones DTW, Pfaff E, Jacob K, et al. Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. Nature. 2012; 482:226–31. [PubMed: 22286061]
31. Chen X, Bahrami A, Pappo A, Easton J, Dalton J, Hedlund E, et al. Recurrent somatic structural variations contribute to tumorigenesis in pediatric osteosarcoma. Cell Rep. 2014; 7:104–12. [PubMed: 24703847]
32. Mugneret F, Lizard S, Aurias A, Turc-Carel C. Chromosomes in Ewing's sarcoma. II. Nonrandom additional changes, trisomy 8 and der(16)t(1;16). Cancer Genet Cytogenet. 1988; 32:239–45. [PubMed: 3163262]
33. Balbás-Martínez C, Sagrera A, Carrillo-de-Santa-Pau E, Earl J, Márquez M, Vazquez M, et al. Recurrent inactivation of STAG2 in bladder cancer is not associated with aneuploidy. Nat Genet. 2013; 45:1464–9. [PubMed: 24121791]
34. Guo G, Sun X, Chen C, Wu S, Huang P, Li Z, et al. Whole-genome and whole-exome sequencing of bladder cancer identifies frequent alterations in genes involved in sister chromatid cohesion and segregation. Nat Genet. 2013; 45:1459–63. [PubMed: 24121792]

35. Kon A, Shih L-Y, Minamino M, Sanada M, Shiraishi Y, Nagata Y, et al. Recurrent mutations in multiple components of the cohesin complex in myeloid neoplasms. Nat Genet. 2013; 45:1232–7. [PubMed: 23955599]
36. Taylor CF, Platt FM, Hurst CD, Thygesen HH, Knowles MA. Frequent inactivating mutations of STAG2 in bladder cancer are associated with low tumour grade and stage and inversely related to chromosomal copy number changes. Hum Mol Genet. 2014; 23:1964–74. [PubMed: 24270882]
37. Xiao T, Wallace J, Felsenfeld G. Specific sites in the C terminus of CTCF interact with the SA2 subunit of the cohesin complex and are required for cohesin-dependent insulation activity. Mol Cell Biol. 2011; 31:2174–83. [PubMed: 21444719]
38. Witcher 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. [PubMed: 19450526]
39. Lopez-Guerrero JA, Pellin A, Noguera R, Carda C, Llombart-Bosch A. Molecular analysis of the 9p21 locus and p53 genes in Ewing family tumors. Lab Invest. 2001; 81:803–14. [PubMed: 11406642]
40. Tsuchiya T, Sekine K, Hinohara S, Namiki T, Nobori T, Kaneko Y. Analysis of the p16INK4, p14ARF, p15, TP53, and MDM2 genes and their prognostic implications in osteosarcoma and Ewing sarcoma. Cancer Genet Cytogenet. 2000; 120:91–8. [PubMed: 10942797]
41. Patel N, Black J, Chen X, Marcondes AM, Grady WM, Lawlor ER, et al. DNA methylation and gene expression profiling of Ewing sarcoma primary tumors reveal genes that are potential targets of epigenetic inactivation. Sarcoma. 2012; 2012:498472. [PubMed: 23024594]
42. Béguelin W, Popovic R, Teater M, Jiang Y, Bunting KL, Rosen M, et al. EZH2 is required for germinal center formation and somatic EZH2 mutations promote lymphoid transformation. Cancer Cell. 2013; 23:677–92. [PubMed: 23680150]
43. Berg T, Thoene S, Yap D, Wee T, Schoeler N, Rosten P, et al. A transgenic mouse model demonstrating the oncogenic role of mutations in the polycomb-group gene EZH2 in lymphomagenesis. Blood. 2014; 123:3914–24. [PubMed: 24802772]
44. Kridel R, Sehn LH, Gascoyne RD. Pathogenesis of follicular lymphoma. J Clin Invest. 2012; 122:3424–31. [PubMed: 23023713]
45. Bejar R, Stevenson K, Abdel-Wahab O, Galili N, Nilsson B, Garcia-Manero G, et al. Clinical effect of point mutations in myelodysplastic syndromes. N Engl J Med. 2011; 364:2496–506. [PubMed: 21714648]
46. Brohl AS, Solomon DA, Chang W, Wang J, Song Y, Sindiri S, et al. The genomic landscape of the Ewing sarcoma family of tumors reveals recurrent STAG2 mutation. PLoS Genet. 2014; 10:e1004475. [PubMed: 25010205]
47. Ladenstein R, Pötschger U, Le Deley MC, Whelan J, Paulussen M, Oberlin O, et al. Primary disseminated multifocal Ewing sarcoma: results of the Euro-EWING 99 trial. J Clin Oncol. 2010; 28:3284–91. [PubMed: 20547982]
48. Peter M, Gilbert E, Delattre O. A multiplex real-time pcr assay for the detection of gene fusions observed in solid tumors. Lab Invest. 2001; 81:905–12. [PubMed: 11406651]
49. Korbel JO, Campbell PJ. Criteria for inference of chromothripsis in cancer genomes. Cell. 2013; 152:1226–36. [PubMed: 23498933]
50. Louis-Brennetot C, Coindre J-M, Ferreira C, Pérot G, Terrier P, Aurias A. The CDKN2A/CDKN2B/CDK4/CCND1 pathway is pivotal in well-differentiated and dedifferentiated liposarcoma oncogenesis: an analysis of 104 tumors. Genes Chromosomes Cancer. 2011; 50:896–907. [PubMed: 21910158]

Statement of significance

Whole-genome sequencing reveals that the somatic mutation rate in Ewing sarcoma is low. Tumors that harbor *STAG2* and *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.

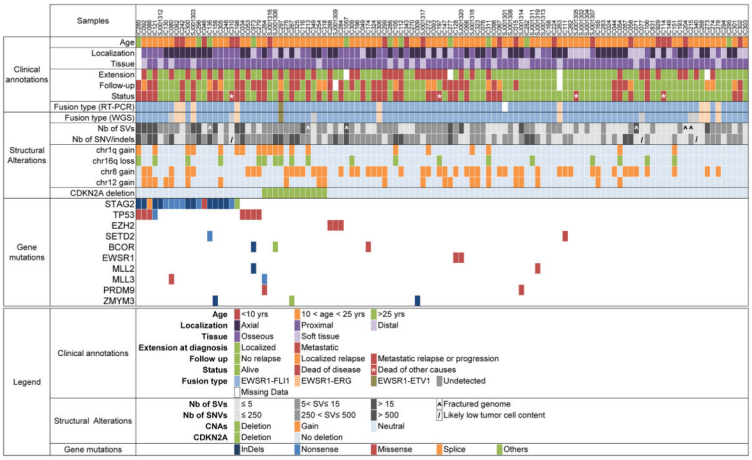

Figure 1. A comprehensive profile of the genetic abnormalities in Ewing sarcoma and associated clinical information

Key clinical characteristics are indicated, including primary site, type of tissue, and metastatic status at diagnosis, follow-up, and last news. Below is the consistency of detection of gene fusions by RT-PCR and WGS. The numbers of structural variants and single-nucleotide variants as well as indels are reported in a greyscale. The presence of the main copy-number changes, chr 1q gain, chr 16 loss, chr 8 gain, chr 12 gain and interstitial *CDKN2A* deletion is indicated. Last are listed the most significant mutations and their types. See Supplementary Table S2 for the complete lists of SNVs/indels, SVs, and CNAs. For gene mutations, others refer to: duplication of exon 22 leading to frameshift (*STAG2*), deletion of exon 2 to 11 (*BCOR*) and deletion of exons 1 to 6 (*ZMYM3*).

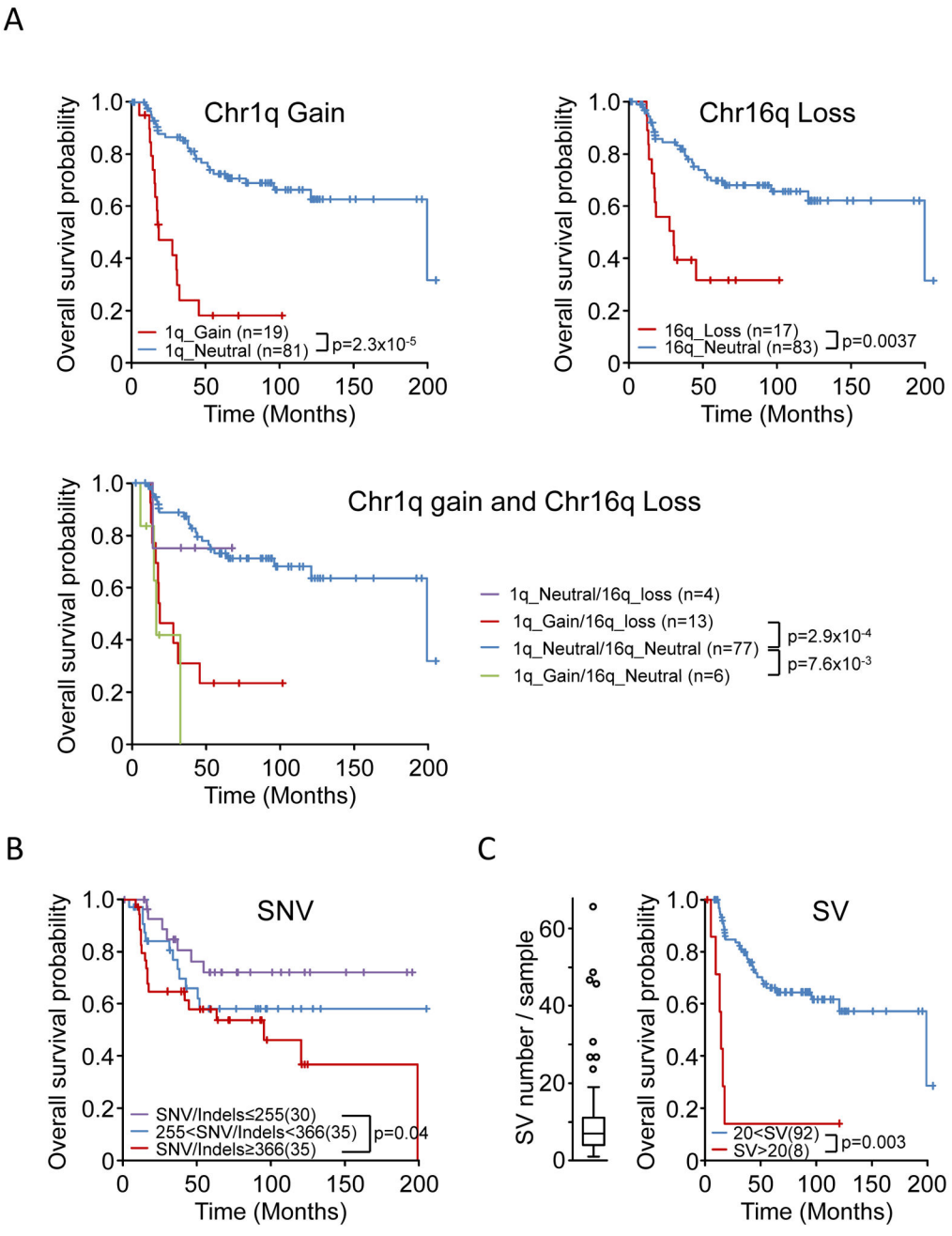

Figure 2. Prognostic significance of CNAs, SVs, and SNV/indels

Kaplan-Meier overall survival estimates according to A) Chromosome 1q gain, chromosome 16q loss; chromosome 1q gain and 16qloss; B) number of SNV/indels. Samples were stratified according to the number of genomic SNVs/indels and split into tertiles; C) a large number of SVs. The overall survival of patients whose tumors harbor an outlier number of SVs (boxplot distribution shown on the left) is compared to that of other patients. Patients with a fractured genome, low tumor purity, or death of causes other than Ewing sarcoma were excluded from the analysis.

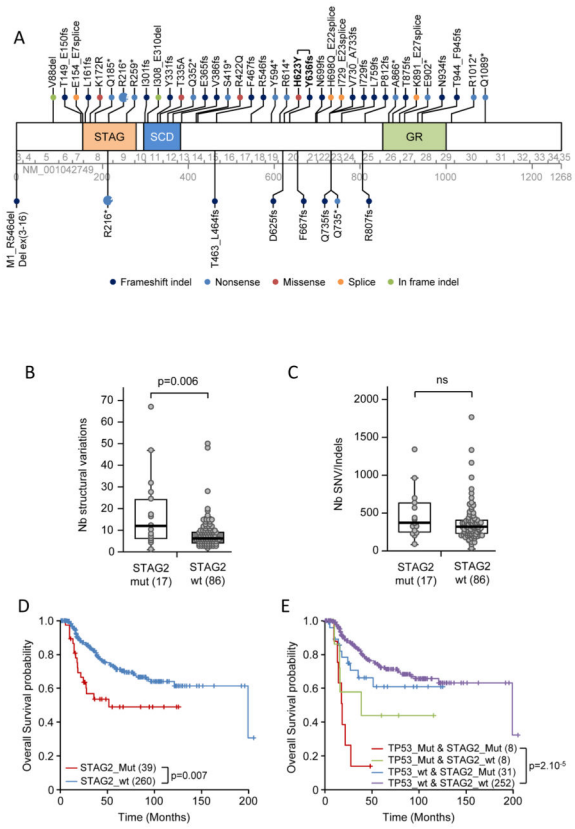

Figure 3. *STAG2* mutations and their prognostic significance in Ewing sarcomas

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 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 represent individual values. *P*-values were determined by using the Mann-Whitney 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.

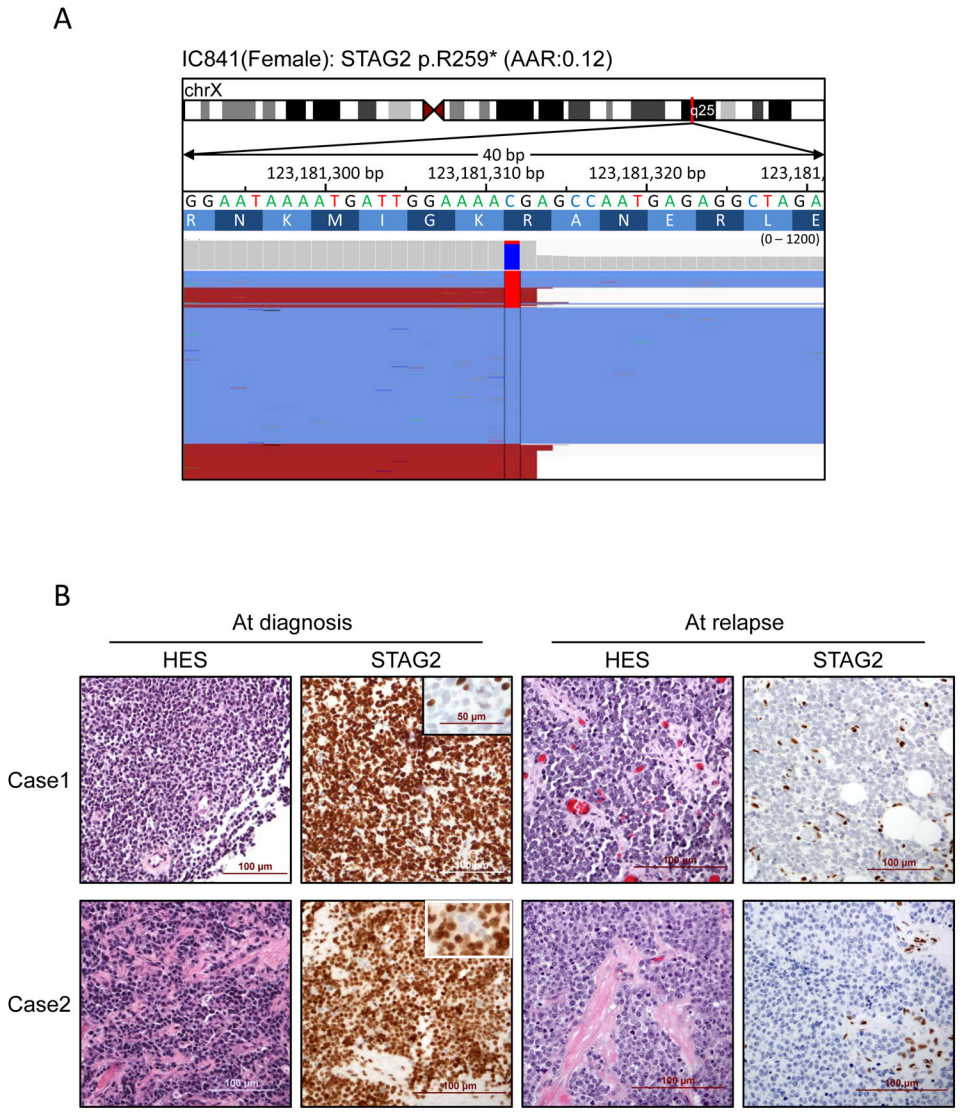

Figure 4. Subclonal presence of *STAG2* mutations

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.

Table 1

Genomic status of *STAG2*, *CDKN2A*, and *TP53* in Ewing sarcoma cell lines

Cell line	*STAG2*[a]	*CDKN2A*[a]	*TP53*[a]
EW-3	p.R216*	wt	wt
EW-22	p.T463_L464fs	wt[b]	p.R175H
EW-23	p.R807fs	wt	p.R273C
MHH-ES1	p.Q735fs	wt	p.S215del
MIC	p.R216*	wt	p.E285K
ORS	p.D625fs	wt	p.C176F
POE	p.F667fs	wt	p.L194R
SK-ES-1	p.Q735*	wt	C176F
SK-NM-C	p.M1_R546Del	wt	p.M1_T125Del
A673	wt	del(1a,1b,2,3)	p.A119fs
EW-1	wt	del(1a,1b,2,3)	p.R273C
EW-7	wt	del(1a)[b]	wt
EW-16	wt	del(1a,1b,2,3)	p.K120fs
STA-ET-1	wt	del(1a,1b,2,3)	wt
TC-71	wt	del(1b,2,3)	p.R213*
STA-ET-3	wt	het[c]	wt
EW-18	wt	wt	p.C176F
RD-ES	wt	wt	p.R273C
STA-ET-8	wt	wt	p.P152T

wt, wild-type.

[a] *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 $CDKNA2^{INK4A}$, exon 1b is specific for $CDKNA2^{ARF}$, exons 2 and 3 are common to both).

[b] Indicates a G->A polymorphism identified in EW-7 and EW-22 cell lines (rs3731249).

[c] The STA-ET-3 cell line has a C to T heterozygous mutation (het) at position chr9:21,971,120 (hgu19), leading to nonsense (p.R80* for $p16^{INK4}$ based on NM_000077) and missense (p.P94L for $p14^{ARF}$ based on NM_058195) mutations.