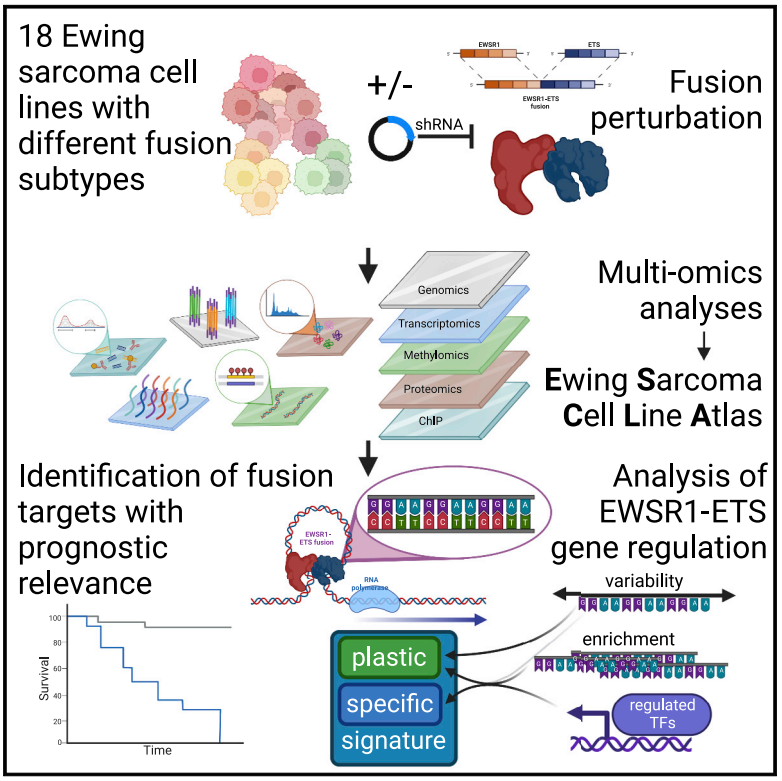
Cell Reports

Resource

# Systematic multi-omics cell line profiling uncovers principles of Ewing sarcoma fusion oncogene-mediated gene regulation

## Graphical abstract




## Highlights

- A comprehensive multi-omics dataset of 18 Ewing sarcoma (EwS) cell lines is presented
- DNA binding of EWSR1-ETS depends on GGAA mSat enrichment and their architecture
- EwS exhibits a specific gene signature despite strong inter-patient plasticity
- Certain heterogeneously EWSR1-ETS-regulated genes may serve as prognostic biomarkers

## Authors

Martin F. Orth, Didier Surdez, Tobias Faehling, ..., Jan Koster, Olivier Delattre, Thomas G.P. Grünewald

## Correspondence

t.gruenewald@dkfz-heidelberg.de

## In brief

Orth et al. leverage multi-omics analyses to study heterogeneity in the pediatric cancer Ewing sarcoma. They present a comprehensive dataset of 18 cell lines to illustrate heterogeneous binding of pathognomonic fusion oncoproteins to polymorphic regulatory DNA elements, which may contribute to the variable expression of prognostically relevant genes.

Orth et al., 2022, Cell Reports *41*, 111761
December 6, 2022 
https://doi.org/10.1016/j.celrep.2022.111761

CellPress

Cell Reports

CellPress
OPEN ACCESS

Resource

# Systematic multi-omics cell line profiling uncovers principles of Ewing sarcoma fusion oncogene-mediated gene regulation

Martin F. Orth,[1] Didier Surdez,[2,3] Tobias Faehling,[4,5] Anna C. Ehlers,[4,5] Aruna Marchetto,[1] Sandrine Grossetête,[2] Richard Volckmann,[6] Danny A. Zwijnenburg,[6] Julia S. Gerke,[1] Sakina Zaidi,[2] Javier Alonso,[7,8] Ana Sastre,[9] Sylvain Baulande,[10] Martin Sill,[4,11] Florencia Cidre-Aranaz,[4,5] Shunya Ohmura,[4,5] Thomas Kirchner,[12,13,14] Stefanie M. Hauck,[15] Eva Reischl,[16] Melissa Gymrek,[17,18] Stefan M. Pfister,[4,11,19] Konstantin Strauch,[20,21,22] Jan Koster,[6] Olivier Delattre,[2] and Thomas G.P. Grünewald[1,4,5,23,24,*]

[1]Max-Eder Research Group for Pediatric Sarcoma Biology, Institute of Pathology, Faculty of Medicine, LMU Munich, 80337 Munich, Germany
[2]INSERM Unit 830 "Genetics and Biology of Cancers," Institut Curie Research Center, 75005 Paris, France
[3]Balgrist University Hospital, Faculty of Medicine, University of Zürich, 8008 Zürich, Switzerland
[4]Hopp Children's Cancer Center (KiTZ), 69120 Heidelberg, Germany
[5]Division of Translational Pediatric Sarcoma Research, German Cancer Research Center (DKFZ), German Cancer Consortium (DKTK), 69120 Heidelberg, Germany
[6]Department of Oncogenomics, Amsterdam University Medical Centers (AUMC), 1105 Amsterdam, the Netherlands
[7]Unidad de Tumores Sólidos Infantiles, Instituto de Investigación de Enfermedades Raras, Instituto de Salud Carlos III, 28029 Madrid, Spain
[8]Centro de Investigación Biomédica en Red de Enfermedades Raras (CB06/07/1009, CIBERER-ISCIII), 28029 Madrid, Spain
[9]Unidad Hemato-oncología Pediátrica, Hospital Infantil Universitario La Paz, 28029 Madrid, Spain
[10]Institut Curie Genomics of Excellence (ICGex) Platform, Institut Curie Research Center, 75005 Paris, France
[11]Division of Pediatric Neuro-Oncology, German Cancer Research Center (DKFZ), German Cancer Consortium (DKTK), 69120 Heidelberg, Germany
[12]Institute of Pathology, Faculty of Medicine, LMU Munich, 80337 Munich, Germany
[13]German Cancer Consortium (DKTK), Partner Site Munich, 80337 Munich, Germany
[14]German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany
[15]Research Unit Protein Science and Metabolomics and Proteomics Core, Helmholtz Zentrum München – German Research Center for Environmental Health, 85764 Neuherberg, Germany
[16]Helmholtz Zentrum München – German Research Center for Environmental Health, 85764 Neuherberg, Germany
[17]Division of Genetics, Department of Medicine, University of California, San Diego, San Diego, CA 92093, USA
[18]Department of Computer Science and Engineering, University of California, San Diego, San Diego, CA 92093, USA
[19]Department of Pediatric Hematology & Oncology, Heidelberg University Hospital, 69120 Heidelberg, Germany
[20]Institute of Medical Biometry, Epidemiology, and Informatics (IMBEI), University Medical Center, Johannes Gutenberg University, 55131 Mainz, Germany
[21]Institute of Genetic Epidemiology, Helmholtz Zentrum München – German Research Center for Environmental Health, 85764 Neuherberg, Germany
[22]Institute for Medical Information Processing, Biometry, and Epidemiology (IBE), Faculty of Medicine, LMU Munich, 81377 Munich, Germany
[23]Institute of Pathology, Heidelberg University Hospital, 69120 Heidelberg, Germany
[24]Lead contact
*Correspondence: t.gruenewald@dkfz-heidelberg.de
https://doi.org/10.1016/j.celrep.2022.111761

## SUMMARY

Ewing sarcoma (EwS) is characterized by EWSR1-ETS fusion transcription factors converting polymorphic GGAA microsatellites (mSats) into potent neo-enhancers. Although the paucity of additional mutations makes EwS a genuine model to study principles of cooperation between dominant fusion oncogenes and neo-enhancers, this is impeded by the limited number of well-characterized models. Here we present the Ewing Sarcoma Cell Line Atlas (ESCLA), comprising whole-genome, DNA methylation, transcriptome, proteome, and chromatin immunoprecipitation sequencing (ChIP-seq) data of 18 cell lines with inducible EWSR1-ETS knockdown. The ESCLA shows hundreds of EWSR1-ETS-targets, the nature of EWSR1-ETS-preferred GGAA mSats, and putative indirect modes of EWSR1-ETS-mediated gene regulation, converging in the duality of a specific but plastic EwS signature. We identify heterogeneously regulated EWSR1-ETS-targets as potential prognostic EwS biomarkers. Our freely available ESCLA (http://r2platform.com/escla/) is a rich resource for EwS research and highlights the power of comprehensive datasets to unravel principles of heterogeneous gene regulation by chimeric transcription factors.

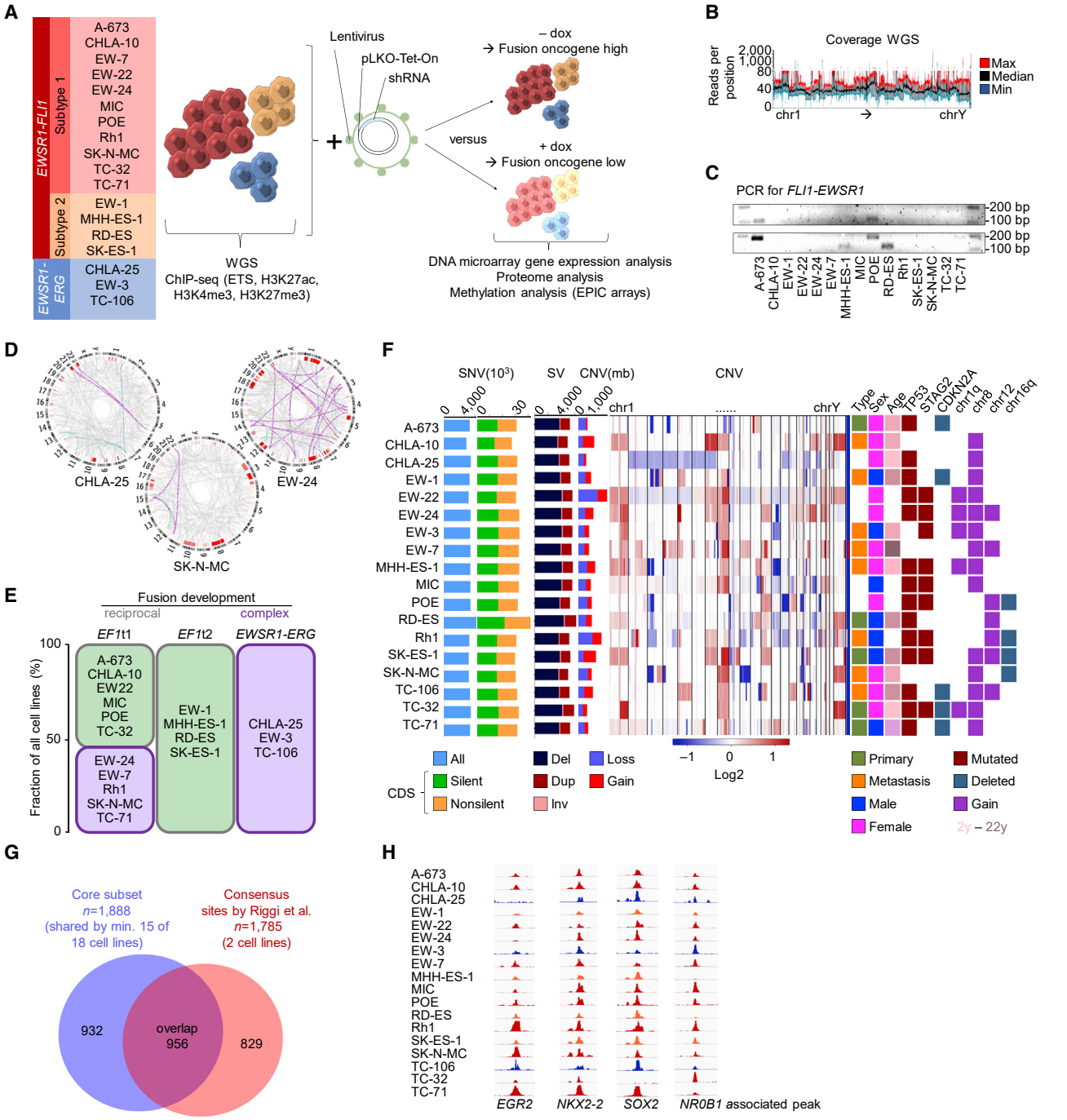



**Figure 1. The Ewing Sarcoma Cell Line Atlas (ESCLA)**

(A) Schematic of the ESCLA. dox, doxycycline; WGS, whole-genome sequencing; ChIP-seq, chromatin immunoprecipitation sequencing.

(B) Coverage of WGS data over the entire human genome, counted per 90-kb window, maximum (max), median, and minimum (min) coverage across cell lines; range is shown in gray.

(C) Gel electrophoresis of the PCR-amplified *FLI1-EWSR1* transcript as an indicator for reciprocal *EWSR1-FLI1* translocation; 100-bp ladder.

(D) Circos plots indicating the most significant genomic rearrangements per cell line in purple.

(E) ESCLA cell lines and the respective type of fusion development. *EF1*t1/2, *EF1* subtype 1/2.

*(legend continued on next page)*

## INTRODUCTION

Ewing sarcoma (EwS) is an aggressive bone and soft tissue cancer, mostly affecting children, adolescents, and young adults.[1] It is composed of monomorphic, undifferentiated, small, round cells[2] and characterized by chromosomal translocations generating in-frame fusions of *EWSR1* and variable members of the *ETS* family of transcription factors (TFs) (in ~85% *FLI1*, in ~10% *ERG*).[3–6] Distinct *EWSR1-FLI1* fusions exist, differing in the number of retained exons.[4–6]

*EWSR1-ETS* encode aberrant TFs with neo-morphic features[4,5,7] that massively rewire the cellular transcriptome, epigenome (DNA methylation and histone modifications), and spliceome.[8] These global changes are mediated partially through EWSR1-ETS-binding to GGAA microsatellites (GGAA mSats), which are frequently located at condensed chromatin regions and converted into potent neo-enhancers.[9,10] Apart from *EWSR1-ETS*, EwS features a striking paucity of somatic mutations and epigenetic alterations (DNA methylation).[11–14]

Despite its monomorphic histology and rather simple genetic architecture, EwS is clinically heterogeneous.[15–17] Prior studies based on a few cell lines suggested that interaction of EWSR1-FLI1 with specific enhancer-like GGAA mSats may contribute to EwS tumorigenesis[18] and clinical heterogeneity,[19] depending on the structure or "length" of the respective GGAA mSat. Cell lines have been crucial for many discoveries in cancer[20,21] including EwS.[22] Because EwS does not occur in animals, and because no *bona fide* genetically engineered mammalian models have been generated, possibly because of the specific nature and localization of GGAA mSats in the human genome,[8,23] patient-derived cell lines remain crucial for EwS research. However, the limited number of comprehensively characterized cell lines may have impeded full discovery of the principles of EWSR1-ETS-mediated gene regulation and representation of the spectrum of (clinical) heterogeneity, also regarding different *EWSR1-ETS* (sub)types.

Thus, we generated the Ewing Sarcoma Cell Line Atlas (ESCLA), comprising 18 molecularly defined EwS cell lines with doxycycline (dox)-inducible short hairpin RNA (shRNA)-mediated knockdown of the respective oncofusion. These cell lines were multi-dimensionally molecularly profiled, which identified hundreds of potential EWSR1-ETS-targets and yielded insights into their heterogenous regulation.

## RESULTS

### The ESCLA

The ESCLA includes 11 EwS cell lines harboring *EWSR1-FLI1* subtype 1 (*EWSR1* exon 7 to *FLI1* exon 6),[24] four subtype 2 (*EWSR1* exon 7 to *FLI1* exon 5), and three with *EWSR1-ERG* (*EWSR1* exon 7 to *ERG* exon 6) (Figure 1A). Whole-genome sequencing (WGS) was performed on all parental cell lines on an Illumina platform with a relatively constant coverage of ~38× (Figure 1B). In keeping with a prior study,[3] all *EWSR1-ERG*-positive cell lines showed evidence of complex chromosomal rearrangements (i.e., chromoplexy). 5 of 11 (45.5%) *EWSR1-FLI1* subtype 1 cell lines exhibited chromoplexy-typical rearrangements, and the remaining *EWSR1-FLI1* cell lines showed characteristics of a balanced chromosomal translocation, including an inverse *FLI1-EWSR1* fusion, as determined by breakpoint analysis of WGS data and PCR amplification (Figures 1C–1E).[25,26] The WGS data mirrored previously described common copy number variations (CNVs), such as gains at chromosome 1q (chr1q), chr8, and chr12 (in 28%, 78%, and 39% of cell lines, respectively),[14,27,28] non-silent mutations in *STAG2* and *TP53* (in 50% and 78%, respectively), and deletions of *CDKN2A* and chr16q (in 28% and 22%, respectively)[14] (Figure 1F). Except for focal copy number gains and losses, all ESCLA cell lines were rather euploid, as indicated by a test adapted from ploidyNGS.[29] Although the cell line CHLA-10 was initially described as *TP53* non-functional,[30] we identified only a monoallelic missense mutation (P33R) predicted to be functionally irrelevant and classified as clinically benign (NCBI ClinVAr: SCV000033391).

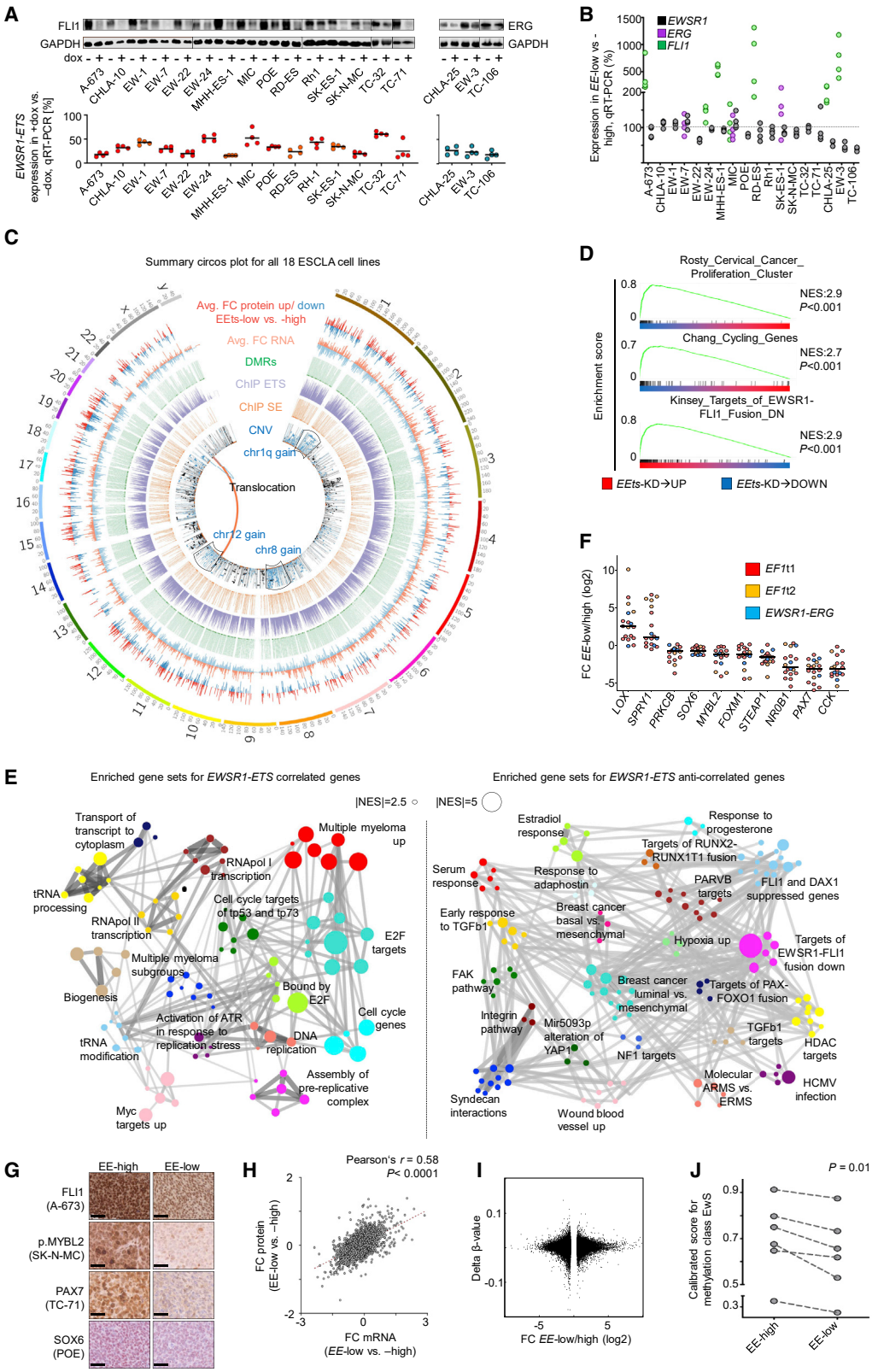
Our WGS approach using 150-bp paired-end reads and a PCR-free protocol enabled genotyping of the mainly intergenic (61.7%) GGAA mSats with haplotype inference and phasing for short tandem repeats (HipSTR).[32] Previous studies have described efficient EWSR1-FLI1-binding to GGAA mSats with ~13–17 consecutive repeats[8,19,33] and 4 consecutive repeats as the minimum number for EWSR1-FLI1-binding.[9] Of the 5,742 GGAA mSats identified across the reference genome (hg19) with 4–17 consecutive GGAA repeats, 3,647 (63.5%) were genotyped for both alleles using HipSTR and high-quality filters (STAR Methods; Table S1).

Chromatin immunoprecipitation sequencing (ChIP-seq) was performed with specific antibodies against FLI1, ERG, and the histone marks H3K27ac, H3K4me3, and H3K27me3 in parental EwS cell lines. ChIP-seq identified, in total, 156,092 EWSR1-ETS binding sites, of which 91,945 were cell line specific. Although the FLI1 and ERG antibodies did not differ in the number of ChIP-seq peaks or in enrichment of an expected EWSR1-ETS binding site in ChIP-PCR (1.9%–8.5% versus 1.6%–3.5% of input CCND1-EF1 immunoprecipitated DNA, respectively), we noted variability of detected EWSR1-ETS peaks across our 18 ESCLA cell lines (median, 13,416 peaks; interquartile range, 24,279). Considering that especially cell lines with peak numbers below the median may limit identification of shared EWSR1-ETS binding sites, we avoided subclassification of peaks in "weak," "intermediate," and "strong" using arbitrary cutoffs. By jointly

(F) Bar plots of counts for single-nucleotide variants (SNVs), structural variants (SVs), and copy number variants (CNVs; called using CNVkit) per cell line; additional CNV heatmap across the genome; and tile plot indicating clinical characteristics and commonly described variants (arm-level chromosomal gains/losses as defined by CNVnator). CDS, coding sequence; del, deletion; dup, duplication; inv, inversion. Sex refers to the genetic sex determined by analysis of sex chromosomes and STR profiling.

(G) Venn diagram indicating the overlap of EWSR1-FLI1 binding sites described by Riggi et al.[31] and the ESCLA core subset (EWSR1-ETS-bound sites in a minimum of 15 of 18 cell lines; Table S2A.

(H) EWSR1-ETS-ChIP-seq at GGAA mSats regulating known EWSR1-ETS-targets.




*(legend on next page)*

analyzing all peaks, we identified 280 EWSR1-ETS binding sites shared by all 18 ESCLA cell lines. 1,888 EWSR1-ETS binding sites were shared by more than 80% of EwS cell lines (at least any 15 of the 18 cell lines per locus) that were hereafter designated as the "core subset" (Table S2A). This core subset included 956 EWSR1-FLI1 binding sites identified previously in two EwS cell lines (A-673 and SK-N-MC);[31] for example, the binding sites close to genes being critical for the EWSR1-ETS-mediated transformation capacity, such as *EGR2*, *NKX2-2*, *NPY1R*, *PPP1R1A*, and *SOX2*[7,18,34–37] (Figures 1G and 1H). We identified 932 core EWSR1-ETS binding sites adjacent to genes that were highly significantly enriched in Gene Ontology (GO) terms involved in developmental and differentiation programs or adjacent to other known drivers of EwS malignancy, such as *NR0B1* and *SOX6*[38–40] (Table S2B).

Genome-wide H3K27ac-based enhancer stitching uncovered 4,339 super-enhancers (SEs) being present in at least one cell line. Strikingly, 99.3% ($p < 0.0001$) of these SEs exhibited co-localization with EWSR1-ETS binding sites (Table S2C), which supports earlier studies demonstrating a strong effect of EWSR1-FLI1 on the EwS epigenome.[31,41] 58 SEs that co-localized with 378 genes were shared by all 18 cell lines (Tables S2C and S2D). GO enrichment analysis revealed that these 378 SE-associated genes were primarily involved in cellular metabolism and regulation of transcription (Table S2D). They also comprised known EWSR1-FLI1-targets and potential diagnostic markers, such as *PRKCB*, *CCND1*, and *ATP1A1*.[42–45] They encompassed 16 of 39 developmental *HOX* family genes ($p < 0.0001$), mainly the *HOXB* and *HOXD* cluster, and an antisense exon, which constitutes a long noncoding RNA.[46] Consistent with reports documenting that especially posterior *HOXD* genes may be critical for EwS,[47–50] only three were anterior, two central (of 15 and 8 anterior/central *HOX* genes), but 11 were posterior (of 16 posterior *HOX* genes, 69%).[51]

As expected, peaks of the activating histone mark H3K4me3 rather decorated the transcriptional start sites (TSSs) of highly expressed genes (median expression percentile, 70th–76th per cell line), whereas peaks of the repressive histone mark H3K27me3 were found at genes with lower expression (median expression percentile, 30th–37th). Genes marked by H3K4me3 were highly significantly enriched for EWSR1-ETS-induced genes ($p < 0.0001$) and those by H3K27me3 for EWSR1-ETS-repressed genes ($p < 0.0001$). 11,090 genes were decorated by H3K4me3 in at least 15 cell lines and significantly enriched for GO terms referring to proliferation ($p < 0.0001$), whereas 2,335 genes were marked by H3K27me3 and enriched for neuronal biological processes ($p < 0.005$) (Table S2E).

To investigate the effect of *EWSR1-ETS*, all ESCLA cell lines were stably transduced with the dox-inducible pLKO-Tet-On all-in-one system[52] harboring a fusion transcript-specific shRNA for *EWSR1-FLI1* subtype 1 and *EWSR1-ERG*-positive cell lines and *FLI1*-specific shRNA for *EWSR1-FLI1* subtype 2 cell lines. Knockdown of the respective fusion was validated at the mRNA and protein level. As shown in Figure 2A, the remaining *EWSR1-ETS* expression upon shRNA induction varied from 15.5%–60.3% at the mRNA level and was below 30% in 10 cell lines (median, 27.7% across all cell lines). Knockdown of the oncofusion was accompanied by increased expression of the wild-type (i.e., non-fused) *FLI1* gene in some *EWSR1-FLI1* cell lines, despite *FLI1* being targeted by the shRNA in some instances. The same effect was observed in 2 of 3 *EWSR1-ERG* cell lines (Figure 2B).

Transcriptome analyses were carried out in all cell lines under EWSR1-ETS-high and EWSR1-ETS-low conditions (triplicates per group) using Affymetrix Clariom D microarrays (Table S3A). Similarly, proteome profiling was carried out by mass spectrometry, and genome-wide DNA methylation was assessed by Infinium MethylationEPIC BeadChips (Illumina) under EWSR1-ETS-high and -low conditions. These analyses clearly demonstrated that modulation of EWSR1-ETS expression had global and massive effects on the transcriptome, proteome, and methylome (Figure 2C), which underscores the definition of *EWSR1-ETS* as "dominant" or "master" oncogenes.[19,36,53,54]

Gene set enrichment analysis (GSEA) and weighted correlation network analysis (WGCNA) revealed that EWSR1-ETS-modulated transcripts were most significantly enriched in EwS-specific and growth-promoting signatures[8,53] (Figures 2D and 2E; Table S3B) comprising known EWSR1-FLI1-targets, including *FOXM1*, *LOX*, and *MYBL2*.[19,55,56] Other known EWSR1-ETS-targets, such as *CCK*, *NR0B1*, *PAX7*, *PRKCB*,

**Figure 2. Effects of EWSR1-ETS knockdown on the transcriptome, proteome, and methylome of EwS cell lines**

(A) Representative western blot and dot plot for qRT-PCR (biological replicates, n = 4) for EWSR1-ETS without and with shRNA induction by dox addition. A bar indicates mean; GAPDH was used as a housekeeping protein.[19]

(B) Dot plot indicating differential expression of non-fused *EWSR1*, *ERG*, and *FLI1* under *EWSR1-ETS*-low versus -high (*EE*-low/high) conditions (n = 4). Only genes with expression of at least 1‰ of the fusion in the *EE*-high state were considered as expressed and depicted.

(C) Circos plot visualizing genome-wide EWSR1-ETS-binding and effects of EWSR1-ETS knockdown on the transcriptome, proteome, and DNA methylation. Fold changes (FCs) of protein and RNA levels are mean values for all cell lines. Differentially methylated regions (DMRs), ChIP peak sites, and CNVs are displayed stacked for each cell line. Avg., average; EETS, EWSR1-ETS; SE, super-enhancer.

(D) Representative gene set enrichment analysis (GSEA) for EETS co- and anti-regulated genes. NES, normalized enrichment score. See also Tables S3A and S3B.

(E) Weighted correlation network analysis (WGCNA) on the GSEA of *EWSR1-ETS*-correlated and -anticorrelated genes using average rank of expression FC across each cell line. The depicted network analyses were based on significant gene sets with an NES of a minimum of 2.5 (Table S3B).

(F) Dot plot indicating FCs of the depicted EWSR1-ETS-target mRNA level upon EWSR1-ETS knockdown; lines medians.

(G) Representative micrographs of immunohistochemically stained FLI1, p.MYBL2, PAX7 and SOX6 EwS cell line xenografts. Scale bars, 50 μm.

(H) Scatterplot indicating average expression FC per gene across cell lines at the mRNA and protein level; a red line indicates linear regression (Table S4A).

(I) Scatterplot indicating genes with absolute FC greater than 0.5 at the RNA level in *EE*-low versus -high condition in more than 6 cell lines and the average delta of the β values of CpG sites, which were uniquely annotated for the promoters of the respective gene.

(J) Before-and-after plot indicating a consistent decrease of the calibrated score for the methylation class EwS (mean value for three replicates each) in the six ESCLA cell lines with highest scores; paired t test.

*SPRY1*, *SOX6*, and *STEAP1*,[39,40,56–60] showed strong deregulation upon EWSR1-ETS knockdown (Figure 2F). *In vivo* differential expression of MYBL2, PAX7, and SOX6 was validated at the protein level in tissue microarrays of xenografts comprising 6 of 18 ESCLA cell lines (Figure 2G). mRNA expression regulation by the oncoproteins correlated significantly positively with the protein regulation (Pearson's $r = 0.58$, $p < 0.001$; Figure 2H; Tables S3A and S4A).

Because the achieved EWSR1-ETS knockdown efficiencies appeared to be variable across cell lines in qRT-PCR (Figure 2A and S1A), a dose dependency of EWSR1-ETS-target regulation on knockdown was likely to be observed in the transcriptome and proteome data. Comparing only genes with an average expression fold change (FC) below −1 (i.e., genes likely driven by EWSR1-ETS) upon EWSR1-ETS knockdown between three groups of cell lines (lowest, intermediate, and highest EWSR1-ETS knockdown in qRT-PCR) indicated a highly significant correlation of the target genes' FC with EWSR1-ETS knockdown efficiency (Figure S1B). The baseline expression levels of the oncofusions across cell lines did not correlate with the baseline expression levels of 10 of the most prominent and best validated EWSR1-ETS-targets (Figure 2F), indicating that dynamics in relative EWSR1-ETS expression may be more relevant for target gene regulation than baseline expression values (Table S3C).

Thus, to avoid arbitrary cutoffs and false negatives in a p value-based approach and to account for variability in the achieved EWSR1-ETS knockdown, we ranked all genes by their FC of RNA or protein levels (EWSR1-ETS low versus high), respectively, and defined genes/proteins as being differentially regulated when their log2 FC exceeded a cell-line-specific threshold (STAR Methods; Figure S1C). This approach minimized potential biases conferred by variability in the achieved EWSR1-ETS knockdown (Figure S1D). As expected, the calculated thresholds for the RNA level (ranging from −0.478 to −1.585 for downregulated and from 0.468–1.614 for upregulated genes upon *EWSR1-ETS* knockdown) correlated significantly with the magnitude of the respective *EWSR1-ETS* knockdown (Pearson's $r = -0.55$ and 0.66, $p = 0.028$ and 0.003). Similar observations were made at the protein level (thresholds ranging from −0.357 to −0.915 and from 0.335–1.043; Pearson's $r = -0.68$ and 0.72, $p = 0.002$ and $p < 0.001$). When comparing 252 randomly picked genes from the transcriptome dataset between the ESCLA cell lines, this approach obviously minimized the effect of EWSR1-ETS knockdown variability because it harmonized the number of identified DEGs per cell line by cell-line-specific cutoff calculation (Figure S1D; for better visualization, only 1% of the genes represented on the employed array platform are displayed).

We applied this approach to all 18 ESCLA cell lines separately, for which 4,392 genes with mRNA and protein quantification were available, and we found 7,597 gene regulation events (i.e., regulation of a specific gene in a particular cell line) at the RNA level (2,428 up- and 5,169 downregulations). At the protein level, 6,895 protein regulation events were observed (3,586 up- and 3,309 downregulations). The overlap of regulation events in the same direction was significantly higher than expected by chance (observed, 1,169; expected, 326; $p < 0.0001$), supporting our previous notion of a strong correlation between RNA and protein level. However, in 134 instances, the direction of regulation differed, which concerns 106 genes across cell lines (Table S4B). In 90 dissociation events, the dissociated regulation was an upregulation at the protein level but downregulation at the RNA level. GO analysis of the 106 genes for which dissociation events were observed in a subset of cell lines revealed significantly enriched gene annotations for cell cycle processes (Table S4C). Because these genes are strictly temporally regulated, it is tempting to speculate that the observed dissociation might result from delayed decay of the proteins upon EWSR1-ETS knockdown.

Although one may expect causal epigenetic alterations to precede changes in the transcriptome upon *EWSR1-ETS* knockdown, Illumina MethylationEPIC arrays did not identify any shared differentially methylated region across cell lines. The methylation status of single CpG sites in promoters did not significantly correlate with differential gene expression across all ESCLA cell lines (Figure 2I). However, it is possible that the observed methylation patterns may reflect the *in vitro* conditions and do not recapitulate the *in vivo* conditions. Yet, despite the lack of the natural microenvironment and the 2D *in vitro* culture, we noted that the ESCLA cell lines generally clustered under conditions of EWSR1-ETS-high with an established reference EwS methylation dataset from 37 primary EwS tumors.[61] The cell lines with the best-matching scores were CHLA-10, CHLA-25, MHH-ES1, MIC, RD-ES, and TC-32 (average matching score, 0.685; range, 0.326–0.913). They showed a significant decrease of the average matching score upon *EWSR1-ETS* silencing ($p = 0.01$; Figure 2J). A similar observation was made when considering all 18 ESCLA cell lines ($p = 0.002$). These findings may indicate that, although EwS tumors exhibit a unique DNA methylation pattern partially caused by EWSR1-ETS, and although we cannot fully exclude that epigenetic changes may have occurred after a longer observational period, DNA methylation appears to not be the prevailing epigenetic mechanism through which the transcriptional effects of EWSR1-ETS are mediated.[13,41]

Because EWSR1-FLI1 affects alternative splicing,[62–65] we analyzed the microarray data at the exon level. 336,899 probe selection regions (PSRs) were investigated for deviating dynamics compared with the gene level. Because 101,232 PSRs were significantly differentially expressed compared with the respective whole transcripts (without correction for multiple testing, 16,845 significant results were expected; factor 6 enrichment, $p < 0.0001$), alternative splicing appears to be a relevant function of EWSR1-ETS. For *ARID1A*, a known EWSR1-FLI1-target of alternative splicing,[64] all 11 PSRs deviated significantly from the dynamics at the gene level. After correction for multiple testing, 129 genes were still affected (Table S3D). These hits reidentified *TERT* as alternatively spliced, which has been described as one of 10 confirmed EWSR1-FLI1 alternative splicing targets.[63]

Our ESCLA constitutes a high-quality, genome-wide, and multi-dimensional dataset for multiple EwS cell lines representing the major *EWSR1-ETS* fusions found in EwS (95% of cases).

## Differences and commonalities of distinct EWSR1-ETS fusions

To systematically analyze potential differences and commonalities of EwS with different EWSR1-ETS (sub)types, we first

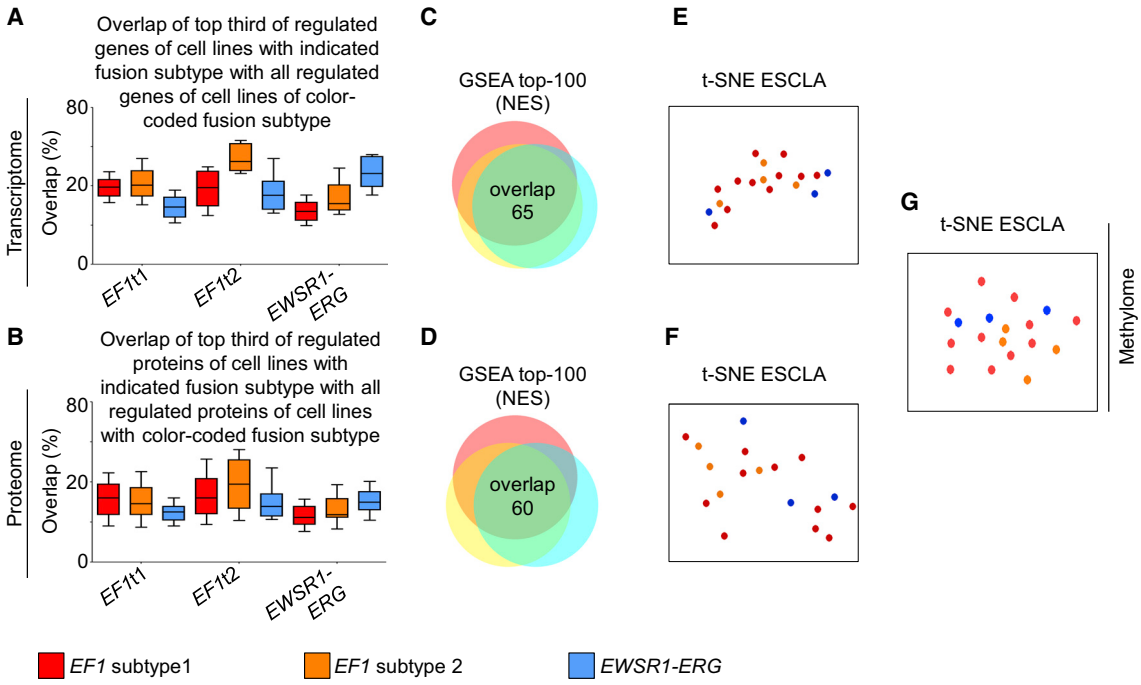



**Figure 3. Differences and commonalities of distinct EWSR1-ETS fusions**

(A) Boxplot indicating overlap of the top third of genes regulated at the mRNA level upon EWSR1-ETS knockdown (only the 924 genes with the highest FCs were considered regulated because this was the minimum number of regulated genes among all cell lines; the top third corresponds to the 208 genes of the 924 with the highest FCs) of any cell line positive for the fusion, indicated on the x axis with the top 924 regulated genes of all other cell lines with the color-coded fusion (sub)type. Center lines represent the median, boxes the interquartile range, and whiskers the 10th–90th percentiles. EF1t1/2, EWSR1-FLI1 subtype 1/2.

(B) Boxplot as in (A) for the top third (72) of regulated proteins of any cell line, with fusion (sub)type indicated on the x axis, versus the top 216 regulated proteins (minimum number of regulated proteins among all cell lines) of all other cell lines.

(C and D) Venn diagram indicating overlap of the top 100 GSEA results per fusion type. GSEA was performed on the gene list ranked by the average FC rank of each gene across all cell lines of a specific fusion (sub)type at the mRNA level (C) and protein level (D).

(E and F) t-distributed stochastic neighbor embedding (t-SNE) plots for the full transcriptome and proteome of all 18 EwS cell lines in the ESCLA.

(G) t-SNE plot for all 18 EwS cell lines in ESCLA based on DNA methylation.

defined EWSR1-ETS-regulated genes and proteins for each cell line separately by a semi-ranked approach in analogy to SE identification with ROSE.[66] Applying this method to all ESCLA cell lines enabled us to refine a previously described EWSR1-ETS-dependent transcriptional "core signature" that was based (besides heterologous *EWSR1-ETS* expression models) on only four different EwS cell lines profiled on different platforms.[67] Defining our core signature as concordantly differentially expressed genes (DEGs) in at least any 15 of all 18 possible cell lines (>80%), we identified 44 EWSR1-ETS-induced and 26 EWSR1-ETS-repressed core signature genes (Table S3E). The EWSR1-ETS-induced core genes comprised *CCK*, *E2F2*, *IL1-RAP*, *LOXHD1*, *PPP1R1A*, and *STEAP1*, all of which were shown to contribute to EwS malignancy.[35,57,60,68–70] Conversely, the EWSR1-ETS-suppressed core genes comprised *LOX* and *FOXO1*, which antagonize EwS malignancy.[55,71]

Second, we compared DEGs across cell lines and fusion subtypes. If the EWSR1-ETS subtypes vastly differed in their DNA-binding preferences, then the rate of random overlap would be only 3.7% and 4.7% (for the mRNA and protein level, respectively) of DEGs across cell lines, even when comparing only the top 33% DEGs of one cell line with all other DEGs of any other cell line. As expected from the similar molecular architecture and conserved C-terminal DNA-binding domains of these EWSR1-ETS fusions,[5,72] we noted a large and highly significant overlap in DEGs and corresponding GSEA outputs at the mRNA and protein levels of 23%–41% ($p < 0.001$ for both; Figures 3A–3D). No subgroups related to the respective fusion could be identified when applying t-distributed stochastic neighbor embedding (t-SNE) or principal-component analysis (PCA) to the mRNA, protein, or CpG methylation data (Figures 3E–3G and S2A).

The transcriptomic and proteomic overlaps were far from perfect, even when comparing DEGs between cell lines with the same fusion subtype, suggesting that factors other than the fusion subtype may cause expression heterogeneity in EwS. However, PCA and t-SNE based on the transcriptomes of EWSR1-ETS-high cell lines stratified for other factors, like chromoplexy and recurrent mutations, did also not result in any unambiguous subgroup (Figure S2B). Anderson et al.[3] reported that EwS tumors with or without chromoplexy may differ in their transcriptomes. Accordingly, we carried out GSEA on ranked gene lists, comparing complex rearranged EWSR1-FLI1-positive cell lines with cell lines with reciprocal translocations or EWSR1-ERG cell lines (all chromoplexy). Very few nominally significant gene sets with absolute normalized enrichment scores

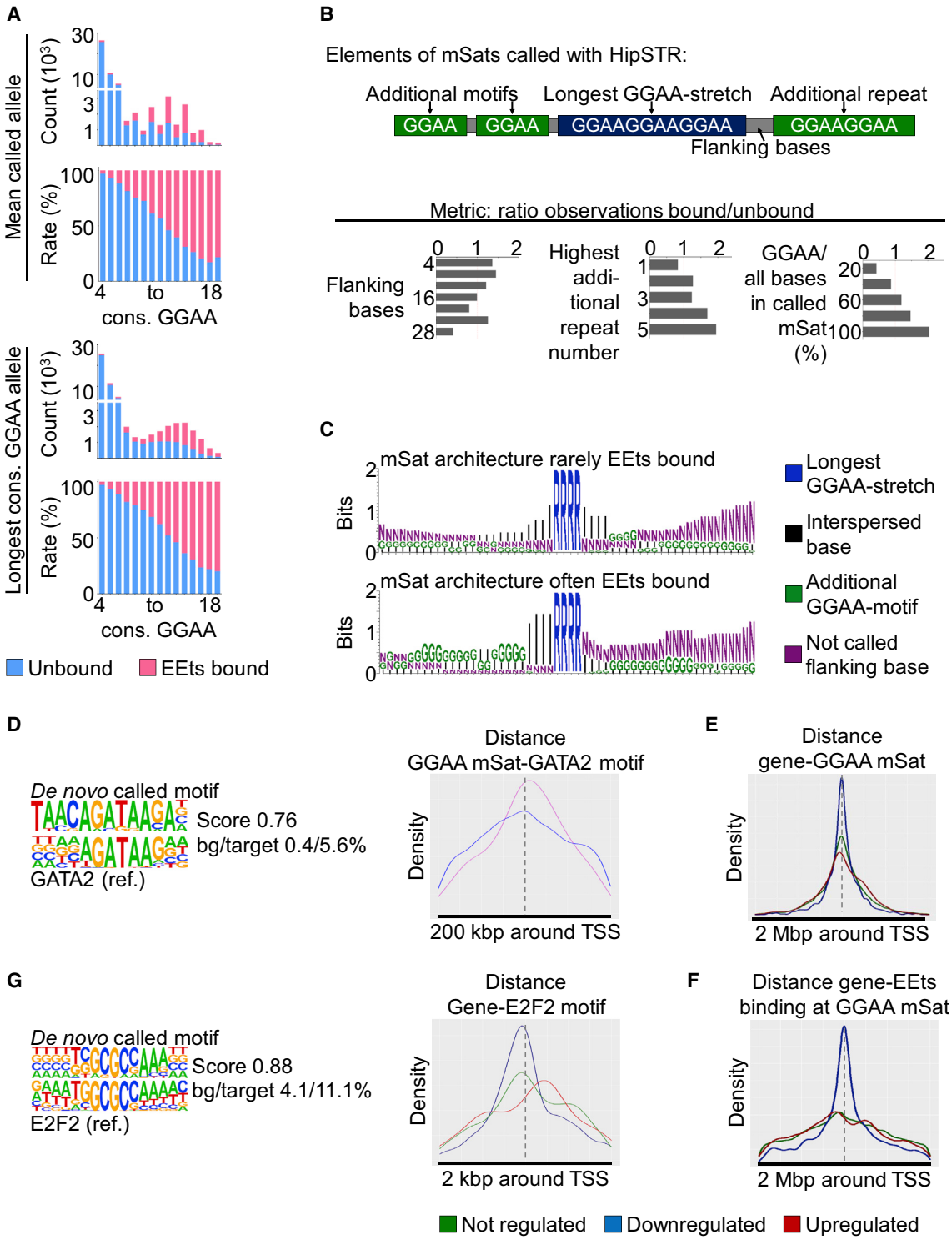



**Figure 4. Characteristics of EWSR1-ETS-bound GGAA mSats in the ESCLA**

(A) Bar plots indicating absolute and average counts of genotyped GGAA microsatellites (mSats) stratified according to average (to panel) or maximum (bottom panel) consecutive (cons.) GGAA repeats for both alleles, EWSR1-ETS (EETS)-binding is indicated in pink.

(B) Top panel: description of the GGAA mSat architecture with the terms additional motif, longest GGAA-stretch, flanking and interspersed bases and additional repeats in GGAA mSats. Bottom panel: bar plots indicating the ratio of GGAA mSats found to be EWSR1-ETS bound/unbound in ChIP, normalized to the expected ratio for the longest GGAA mSat, stratified by different components of the GGAA mSat architecture.

*(legend continued on next page)*

(NES) greater than 2 and a false discovery rate (FDR) of less than 0.05 were identified in the first analysis (complex versus reciprocal translocations). These mainly concerned generic gene sets involved in mitochondrial proteins and DNA packing. Because the same gene sets were enriched in the latter analysis, they are not specific for chromoplexy (complex *EWSR1-FLI*1 versus *EWSR1-ERG* rearrangement).

Crompton et al.,[12] in a joint analysis of 9 EwS cell lines and 21 EwS tumors, identified that EwS cell lines and tumors with loss of *STAG2* were enriched in gene signatures associated with metastasis. However, applying the same algorithms to our transcriptome data (EWSR1-ETS-high, 18 cell lines) did not yield any hits with high signal-to-noise ratios ($\geq$1.5) or enrichment of metastasis signatures among the top hits. Because the sample size might be too small to detect subtle effects of *STAG2* loss of function already on gene sets, we carried out comparative transcriptomic analyses in *STAG2*-mutated versus wild-type ESCLA cell lines. This approach identified 20 genes with an absolute log2 FC greater than 2.0 and $p < 0.01$. Of the 9 genes upregulated in *STAG2*-mutated cell lines, only one was in a promoter-enhancer interaction chain (including at least one EWSR1-FLI1-bound GGAA mSat enhancer) identified in TC-71 cells in recently published high-throughput protein centric chromatin (HiChIP) data.[73] Contrarily, 11 genes were downregulated in *STAG2*-mutated cells, and for 4 and 6 genes, promoter-enhancer interaction chains were identified in A-673 and TC-71 cells, respectively. All of these chain interactions were weaker under *STAG2* knockout conditions. This might indicate that *STAG2* mutations result in altered CCCTC-binding factor (CTCF)-anchored loop extrusion, causing reduced enhancer-promoter interactions and EWSR1-ETS-mediated gene regulation, explaining subtly altered gene expression signatures. Subset analyses for chromosomal aberrations (chr1q, chr8, chr12, and chr16q with restriction on genes localized on the respective chromosome) or based on DEGs (now considering only genes with at least moderate EWSR1-ETS-dependent regulation in at least 50% of cell lines) did not yield any obvious clusters (Figure S2B).

## Characteristics of EWSR1-ETS-bound GGAA mSats in the ESCLA

To confirm that EWSR1-FLI1 and EWSR1-ERG bind to the same DNA motifs, we carried out *de novo* motif analyses with HOMER.[74] These analyses demonstrated that the known EWSR1-FLI1-binding motif, a GGAA mSat, was the top enriched DNA motif among the EWSR1-ERG binding sites compared with background comparison sites (22.93% in target sites versus 0.92% in background; Figure S3A) as well as the single GGAA motif (45.7% in target sites versus 5.68% in background). In line with the identical DNA binding motifs for fusion subtypes, we noted a significant overlap in shared binding sites. Of the 2,639 sites that were bound in each of the three EWSR1-ERG cell lines, 280 were bound by each of the 15 EWSR1-FLI1 cell lines ($p < 0.0001$) and 2,228 by more than 50% of these cell lines. Accordingly, we subsequently analyzed the DNA binding patters of both fusion types jointly.

Although the mechanisms through which EWSR1-ETS fusions suppress transcription of specific genes are poorly understood and likely mediated indirectly, a direct *trans*-activational role of EWSR1-ETS is considered for most EWSR1-ETS-induced genes. For induction of gene transcription, EWSR1-ETS-binding to GGAA mSats has been shown to play a pivotal role.[7,10,18,39,40] We found that EWSR1-ETS-induced genes were strongly and highly significantly enriched for nearby EWSR1-ETS-bound GGAA mSats (defined as a minimum of 4 consecutive GGAA repeats) compared with unregulated genes (on average, 1.41 versus 0.98 EWSR1-ETS-bound GGAA mSats in a window of 2 Mbp around the TSS; $p < 0.0001$). In agreement with indirect regulation of genes suppressed by EWSR1-ETS (i.e., genes being upregulated upon EWSR1-ETS knockdown), such an enrichment of EWSR1-ETS-bound GGAA mSats was less pronounced for those genes (on average, 1.11 bound GGAA mSats).

Because GGAA mSats are highly polymorphic,[40,75] we hypothesized that their polymorphism may affect EWSR1-ETS-binding, resulting in heterogeneous (neo)enhancer function. Thus, we carried out several analyses to assess the "length" and composition of EWSR1-ETS-bound and -unbound GGAA mSats as well as their flanking regions. We first studied the number of consecutive GGAA repeats based on the HipSTR output. As shown in Figure 4A, the average GGAA repeat number of both alleles and the number of GGAA repeats of the "longer" allele per locus were strongly and highly significantly correlated with the rate of observed EWSR1-ETS-binding (Pearson's $r^2 = 0.98$ and 0.99, $p < 0.0001$). In concordance with studies of single cell lines describing a "sweet spot" for EWSR1-ETS-binding,[10,40] we noted a second maximum of genotyped GGAA mSats (for mean and maximum allele length at ~12–14 consecutive GGAA repeats). The increase in EWSR1-ETS-binding rate over the count of maximum consecutive GGAA repeats of both alleles had a slight latency compared with the average count, indicating that long consecutive GGAA mSats on both alleles increase the probability of EWSR1-ETS-binding compared with loci containing only one "long" allele (minimum mean allele length versus

(C) Motifs for the most often EETS-bound and -unbound GGAA mSats.
(D) Left: *de novo* called motif in the flanking region of always EETS-bound GGAA mSats and the matched known reference binding motif for GATA2. Right: density plot indicating the relative position of the GATA2 motif relative to EWSR1-ETS-bound and -unbound GGAA mSats. bg, background (i.e., control sequence) frequency of the *de novo* motif; target, *de novo* motif frequency in investigated target sequences; ref, reference binding motif.
(E) Density histogram of the distance of a GGAA mSat to the transcription start site (TSS; indicated as a dashed gray line) of a downregulated (blue), upregulated (red), and unregulated (green) gene upon EETS knockdown. Genes were considered down- or upregulated when the regulation was observed in at least 33% of cells.
(F) Density histogram indicating the distance of the next EETS-bound GGAA mSat to the TSS (dashed gray line) of a downregulated (blue), upregulated (red), and unregulated (green) gene upon EETS knockdown in the respective cell line.
(G) Left: *de novo* called motifs for TF binding in the promoters of EETS-regulated genes. Right: density plot for the distance of the *de novo* called E2F2 motif to the next TSS of any gene, stratified by the effect of EWSR1-ETS on the respective gene.

maximum allele length with more than 50% of EWSR1-ETS-binding 12 versus 13 GGAA repeats; Figure 4A).

Next we assessed the effect of additional GGAA motifs and "shorter" GGAA mSats and the nature of interspacing bases around the longest consecutive GGAA-stretch on EWSR1-ETS-binding. These additional GGAA motifs and bases were also genotyped by HipSTR within the GGAA mSats' genomic coordinates predicted by Tandem Repeats Finder,[76] and their rates of occurrence at EWSR1-ETS-bound and -unbound GGAA mSats were calculated. To reduce noise, only homozygous alleles were investigated, and the calculated ratios were normalized with the expected ratio for GGAA mSats with the same number of consecutive motif repeats but without further genotyped bases. As shown in Figures 4B and 4C, GGAA mSats that were often bound by EWSR1-ETS typically exhibited, besides relatively high numbers of GGAA repeats at the "longest" stretch, the following two related characteristics: (1) a high number of additional single GGAA motifs and/or additional GGAA repeats in their vicinity and (2) a low number of interspersed bases or flanking bases to the next GGAA motif/repeat.

Finally, we explored whether the flanking regions defined as 1 kbp up- and downstream of EWSR1-ETS-bound GGAA mSats were enriched for TF binding motifs (251 GGAA mSats were bound by EWSR1-ETS in each cell line and examined for TF motifs; 4,934 GGAA mSats were not bound by EWSR1-ETS in any cell line and used as background in the analysis) by *de novo* motif finding with HOMER.[74] Across the ESCLA, the only enriched motifs matching known TFs and fulfilling stringent selection criteria (not indicated as possibly false positive by the motif caller, minimum of ~5% of targets as recommended in the motif caller's description score for known motif at least default threshold 0.6, and minimum 5-fold enrichment) were those for *DUX4*, *GATA2/6*, and *FOXF2* downstream and *RUNX1* and *MIX1* upstream of the GGAA mSats (Figure 4D). These TF motifs were only present in ~5%–7% of the tested target regions, and only *GATA2* appeared to be relatively highly expressed and EWSR1-ETS-driven in the ESCLA, which may suggest that these TFs play a subordinate role in EwS. In accordance, RNAi-mediated knockdown of *GATA2* onto 19% remaining expression (as confirmed by qRT-PCR) did not yield a significant overlap of likely GATA2- and EWSR1-FLI1-driven genes in A-673 cells (Table S5).

We next utilized our ESCLA to analyze the distance of EWSR1-ETS-bound and -unbound GGAA mSats to EWSR1-ETS-regulated genes. As expected from the known enhancer-like activity of GGAA mSats in EwS,[31] we noted a significantly ($p < 0.0001$, Wilcoxon rank-sum test) shorter distance of the closest GGAA mSat to the TSS of a given EWSR1-ETS-induced gene, which was much longer in unregulated or suppressed genes (average distance 135 kbp versus 228 kbp and 461 kbp, respectively; Figure 4E). This became more prominent when focusing only on EWSR1-ETS-bound GGAA mSats (average distance 302 kbp versus 420 kbp and 415 kbp, respectively; $p < 0.0001$; Figure 4F). Conversely, the TSSs of EWSR1-ETS-repressed genes were significantly closer to EWSR1-ETS-bound sites located devoid of a GGAA mSat (mostly single GGAA motifs) than those of EWSR1-ETS-induced genes (average distance 363 kbp versus 402 kbp; $p < 0.0001$). EWSR1-ETS-repressed genes were enriched for nearby (2 Mbp around TSS) EWSR1-ETS-binding devoid of a GGAA mSat compared with EWSR1-ETS-induced genes (16.84 versus 13.65 binding events; $p < 0.0001$), suggesting that repressed genes are controlled by EWSR1-ETS-bound single ETS-like motifs.

To explore mechanisms of indirect EWSR1-ETS-mediated gene regulation, we tested the hypothesis that the promoters of such indirectly regulated genes should be enriched for binding motifs of EWSR1-ETS-regulated TFs, using HOMER for 1 kb upstream to 50 bp downstream of TSSs of genes regulated in at least one third of cell lines upon EWSR1-ETS knockdown versus genes never found to be regulated by EWSR1-ETS (880 down-, 643 up-, and 15,967 unregulated genes). Although the promoters of most EWSR1-ETS-regulated genes did not show evidence of enriched TF binding sites, we detected a significant enrichment of NFAT5 motifs in 12.8% of the EWSR1-ETS-suppressed genes (versus 4.3% in unregulated genes). In the promoters of EWSR1-ETS-induced genes, an NFY- or E2F1/2/4-like motif enrichment was observed in 27.7% and 11.1%, respectively (compared with 15% and 4.1% in unregulated genes, respectively) (Figure 4G). Consistent with a potential co-activator function, *NFAT5* was downregulated by EWSR1-ETS in 3 of 18 EwS cell lines, whereas *NFYC* and *E2F1/2/4* were induced by EWSR1-ETS in 6, 12, 15, or 1 of 18 EwS cell lines, respectively. However, although *E2F* gene regulation by EWSR1-ETS is consistent with previous reports and our core signature comprising *E2F2* (Table S3E),[68,77] RNAi-mediated transient *E2F2* knockdown onto 13% remaining expression (as confirmed by qRT-PCR) did not yield a significant overlap of likely E2F2- and EWSR1-FLI1-driven genes in A-673 cells (Table S5).

Riggi et al.[31] identified ELF1 binding near EWSR1-FLI1-repressed genes after fusion knockdown in a single cell line (SK-N-MC). To assess this phenomenon in our ESCLA, we searched genome wide for ELF1 motifs and calculated the distances between EWSR1-ETS down-, up-, and unregulated genes' TSSs and the nearest ELF1 motif. We observed a slightly higher density of ELF1 motifs near the TSSs of EWSR1-ETS-driven genes (downregulated upon EWSR1-ETS knockdown; average distance 2.4 kbp), which indicates a co-regulatory role of ELF1 and EWSR1-ETS in the promoter region. For EWSR1-ETS-suppressed genes, the distance to the next ELF1 motif was greater than for unregulated genes (average distance 3.0 kbp and 2.8 kbp, respectively), hinting toward a potential role of ELF1 in the enhancer rather than the promoter site for those genes (Figure S3B).

## Association of heterogeneously EWSR1-ETS-regulated genes with patient survival

Clinically, EwS is a heterogeneous disease.[8] Given the rather "silent" genome and homogeneous epigenome of EwS[11–14,61] as well as the inherent polymorphic nature of GGAA mSats, it is tempting to speculate that part of the observed clinical heterogeneity may be caused by the inter-individually variable composition of EWSR1-ETS-bound GGAA mSats and, thus, differential EWSR1-ETS-mediated gene regulation. To investigate this possibility, we tested our list of heterogeneously EWSR1-ETS-regulated genes (defined as being regulated strongly and weakly or not in a minimum of 6 of 18 ESCLA cell lines each;

**Table 1. Heterogeneously regulated ESCLA genes associated with overall survival in the entire cohort (n = 196) and patients with localized disease (n = 129)**

| Gene | Description | Adj. p value (entire cohort, n = 196) | Adj. p value (only localized, n = 129) | Surv. corr. | IQR perc. |
|---|---|---|---|---|---|
| *AIF1* | actin binding protein regulating immune cells | 0.0010 | 0.0900 | high | 88 |
| *CDC20* | cell cycle regulator | 0.0004 | 1.0000 | high | 91 |
| *CDC25A* | dual-specific phosphatase, DNA damage responder | 0.0028 | 0.7650 | high | 50 |
| *CDC25C* | cell cycle regulator, triggers mitosis | 0.0194 | 0.3900 | high | 88 |
| *CENPI* | centromere protein, E2F target | <0.0001 | 0.3150 | high | 79 |
| *DUSP26* | dual-specific phosphatase | 0.0161 | 0.1350 | high | 66 |
| *E2F1* | TF, cell cycle regulator | 0.0001 | 1.0000 | high | 34 |
| *EXO1* | exonuclease | <0.0001 | 0.0750 | high | 94 |
| *GPN3* | GTPase | 0.0002 | 0.7500 | high | 68 |
| *GTSE1* | cell cycle-dependent protein | 0.0003 | 0.0300 | high | 80 |
| *KIF14* | microtubule regulator involved in mitosis | 0.0190 | 0.2550 | high | 90 |
| *KIF15* | microtubule regulator involved in mitosis | 0.0075 | 0.0150 | high | 90 |
| *KIF2C* | microtubule regulator involved in mitosis | <0.0001 | 0.3150 | high | 88 |
| *MYBL2* | TF involved in cell cycle | 0.0004 | 0.0900 | high | 87 |
| *NEK2* | protein kinase involved in mitosis | <0.0001 | 0.0088 | high | 79 |
| *NUF2* | centromere-associated protein | 0.0011 | 0.3150 | high | 97 |
| *SPAG5* | associated with mitotic spindle apparatus | 0.0093 | 0.2400 | high | 89 |
| *TPX2* | microtubule regulator involved in mitosis | 0.0014 | 0.1050 | high | 89 |
| *TRIP13* | ATPase involved in cell cycle progression | 0.0084 | 0.0101 | high | 90 |

The p values are Bonferroni corrected. Surv. Corr, expression status correlating with worse survival; IQR perc., percentile of the interquartile range in the expression data of the survival cohort. For genes heterogeneously regulated by EWSR1-ETS, see also Table S3F.

Table S3F) for association with overall survival in a cohort of 196 EwS patients. After Bonferroni correction and filtering for EWSR1-ETS-induced genes, we identified 19 heterogeneously regulated genes that were significantly associated with overall survival (Table 1). These genes comprised *MYBL2*, which was previously identified to be heterogeneously EWSR1-FLI1-regulated because of germline variation in a proximal enhancer-like GGAA mSat.[19]

Because metastasis is the major negative predictor in EwS,[8] we re-analyzed the association of these genes with overall survival, now only taking into account patients with localized disease at diagnosis. As shown in Table 1, the association with overall survival remained significant for 4 of 19 genes, suggesting that their association with survival is independent of metastasis. Whether the other 15 genes are actual drivers of metastasis or simply epiphenomena of the metastatic process would need to be evaluated in future functional studies.

## DISCUSSION

We present the ESCLA, a comprehensive and multidimensional dataset comprising 18 well-curated EwS cell lines. Because all data are freely available via the Gene Expression Omnibus (GEO: GSE176339) or Sequencing Read Archive (SRA: PRJNA610192) and intuitively browsable via the R2 platform (http://r2platform.com/escla/), we believe that this first version of the ESCLA will be an extremely rich scientific resource.

Our targeted analyses shed light on the nature of EWSR1-ETS-preferred GGAA mSats, potential indirect modes of EWSR1-ETS-mediated gene regulation through secondary TFs, and highlighted putative co-regulatory TF binding sites flanking EWSR1-ETS-bound GGAA mSats. However, how to call EWSR1-ETS-bound enhancer-like GGAA mSats has not been consistently defined in the EwS literature. Thus, previous and current conclusions regarding the nature of EWSR1-ETS-bound enhancer-like GGAA mSats may depend on the given definition.

An important finding supported by our ESCLA is the transcriptional "duality" of EwS. On the one hand, EwS is characterized by a highly specific EWSR1-ETS-induced transcriptional program. On the other hand, there is a relatively high degree of plasticity in this program, which may be explained at least partially by genes driven by polymorphic GGAA mSat neo-enhancers, whose affinity for EWSR1-ETS and, thus, enhancer activity can vary substantially because of their variable structure. This duality may explain the difficulties experienced so far in identifying specific and time-sensitive diagnostic marker proteins for EwS[8] and provide the basis for identification of heterogeneous EWSR1-ETS-target gene regulation, which may have implications for the clinical heterogeneity of EwS. However, regional permissive transcription mediated by chromatin configuration could

also contribute to this plasticity, particularly because BAF complexes recruited by EWSR1-ETS are known chromatin remodelers.[31,78,79] Our results support several previously reported aspects of EWSR1-ETS-mediated gene regulation, such as the prominent EWSR1-ETS-binding to GGAA mSats[9] followed by local chromatin opening[31,79] as explained by BAF complex recruitment[78] and indicated by changes in H3K27ac marks,[41] binding-associated transcriptional effects,[10] and coregulatory TFs such as E2F.[41,68,80] However, our findings go beyond these studies, which have analyzed single or just a few, mostly EWSR1-FLI1-positive cell lines. Datasets generated by these studies could not be incorporated seamlessly into the current version of our ESCLA because they were partly performed on different platforms using different protocols and key tools, such as different antibodies, resulting in batch effects, which would render *bona fide* integration in the ESCLA impossible. The duality described here underscores the importance of analyses based on multi-dimensional datasets of a large number of cell lines and may suggest that future functional studies could be more representative of EwS when being carried out in multiple cell lines.

Although only around 5%/20% of primary EwS exhibit *TP53/STAG2* mutations,[11,12,14] respectively, around 80%/50% of our ESCLA cell lines had functionally relevant *TP53/STAG2* mutations, respectively. However, *TP53/STAG2* mutations have been shown to be a negative prognostic parameter,[14] indicating that *TP53/STAG2* mutations may convey a more aggressive phenotype to EwS cells, possibly enabling them to grow better *in vitro* and suggesting that our ESCLA may be especially relevant for this high-risk group of EwS patients.

Knockdown efficiency for EWSR1-ETS varied among the ESCLA cell lines, which complicated comparisons of DEGs and conflicted with a simple definition of DEGs by a uniform FC- and p value-based threshold. Hence, we took the variability of the achieved knockdown into account and defined DEGs by cell-line-specific thresholds using an algorithm adapted from ROSE (STAR Methods; Figure S1). As a proof of concept, this approach identified and refined an EWSR1-ETS-dependent transcriptional core signature (Table S3E) that contained many known and functionally relevant EWSR1-ETS-targets.[57,60,69,71,77] Thus, our approach may prove beneficial for similar comparative analyses of EwS and other cancer types.

EWSR1-ETS-binding to single ETS motifs has been suggested to contribute to EWSR1-ETS-mediated transcriptional repression.[31] The strong enrichment for single ETS-like motifs near EWSR1-ETS-suppressed genes observed in the ESCLA provides more evidence that these binding sites may be involved in transcriptional repression, perhaps in conjunction with altered chromatin configuration and/or accessibility.

To the best of our knowledge, our ESCLA is the largest and most comprehensive EwS cell line atlas that has been established to date. Thus, we believe that our study provides a blueprint for systematic and multi-dimensional analysis of a large array of cancer cell lines of a specific cancer type and can yield insights into the underlying principles of oncogene-mediated gene deregulation governing the complex and variable phenotype of a genetically homogeneous but clinically heterogeneous disease.

## Limitations of the study

The WGS read-length (150 bp paired end) sets a barrier for genotyping very long GGAA mSats, precluding robust conclusions on such potentially important EWSR1-ETS binding sites. Another important aspect is the precise annotation of EWSR1-ETS-regulated genes with the corresponding regulatory element that will be crucial to further dissect the regulatory landscape of EwS. Both aspects will be highly valuable extensions of this first ESCLA version that will be addressed using long-range sequencing technologies and chromosome conformation capture methods, as achieved previously for a few EwS cell lines.[73,81,82]

The results of this study and the ESCLA are based on (immortal) tumor-derived cell lines cultured as monolayers. Additional analyses under more physiological conditions, such as cancer cells grown in three dimensions using recently proposed human plasma-like cell culture media or using primary/patient-derived xenograft (PDX)-derived EwS culture models, might help to uncover additional genes regulated by the EWRS1-ETS fusions. Because the field of multi-omics analysis is evolving rapidly, partly reanalyzing the generated cell lines appears reasonable when advanced technologies meet high-quality criteria to answer more questions; for instance, identification of the non-covered GGAA mSats in long-read sequencing. All omics data were generated at the bulk level, and relevant changes upon EWSR1-ETS knockdown that only affect a minority of cells in a given cell line may remain elusive. Future versions of the ESCLA may comprise single-cell omics technologies to overcome this limitation. Although the study highlighted the genome-wide relevance of the GGAA mSat/EWSR1-ETS interaction, high-resolution chromatin conformation capture is required to disentangle the complex interaction network of the EwS fusions to quantitatively correlate the repeat numbers and RNA transcript regulation (eQTL). Finally, because matched WGS data, RNA sequencing data, and chromatin conformation capture data were not available for large numbers of EwS patients, we could not presently validate, in human patient samples, whether the heterogeneous expression of the identified survival-associated genes is truly based on repeat number variability at associated GGAA mSats. Because EwS is much more common in White than in non-White populations,[83] it will be very interesting to compare germline GGAA repeat number variabilities at survival-associated genes in different human populations.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Provenience of cell lines
  - Culture conditions for cell lines
  - *In vivo* expansion of EwS cell lines

- METHOD DETAILS
  - EWSR1-ETS-knockdown cell lines establishment
  - Transient E2F2 and GATA2 knockdown
  - RNA extraction, RT and qRT-PCR
  - Western blot
  - TMA construction and IHC
  - Whole-genome sequencing (WGS)
  - Genotyping of GGAA mSats
  - Copy number variation and ploidy analysis
  - Variant calling
  - ChIP-seq
  - DNA microarray gene expression analyses
  - Protein quantification
  - Methylation analysis
  - GSEA, WGCNA, gene ontology
  - Survival analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
- ADDITIONAL RESOURCES


## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2022.111761.

## ACKNOWLEDGMENTS

We thank Rebeca Alba Rubio, Stefanie Stein, Felina Zahnow, Stefanie Kutschmann, Mario Gipp, Heike Prelle, Andrea Sendelhofert, and Anja Heier for excellent technical assistance and Dr. Fabian Metzger for mass spectrometry readouts. We thank Dr. Paul Northcott, Dr. Carlos Rodriguez-Martin, and Dr. Brian Gudenas for providing R scripts for WGCNA. This project was mainly supported by German Cancer Aid (70112257 to T.G.P.G.). The laboratory of T.G.P.G. was supported by grants from the LMU Munich's Institutional Strategy LMUexcellent within the framework of the German Excellence Initiative, the "Mehr LEBEN für Krebskranke Kinder – Bettina-Bräu-Stiftung," the Wilhelm Sander Foundation (2016.167.1), the Friedrich-Baur Foundation, the Matthias Lackas Foundation, the Dr. Leopold und Carmen Ellinger Foundation, the Deutsche Forschungsgemeinschaft (DFG 458891500), German Cancer Aid (70114111 and 70114278), the SMARCB1 Association, the Federal Ministry of Education and Research (SMART-CARE, HEROES-AYA), the Boehringer Ingelheim Foundation, and the Barbara and Wilfried Mohr Foundation. This research was supported by the Helmholtz Zentrum München – German Research Center for Environmental Health and within the Munich Center of Health Sciences (MC-Health) as part of LMUinnovativ. The research was also supported by the EU Horizon 2020 program (grant 826121, iPC project). High-throughput sequencing of ChIP was performed by the ICGex NGS platform of the Institut Curie, supported by grants ANR-10-EQPX-03 (Equipex) and ANR-10-INBS-09-08 (France Génomique Consortium) from the Agence Nationale de la Recherche ("Investissements d'Avenir" program), Canceropole Île-de-France, and SiRIC-Curie program – SiRIC grant INCa-DGOS-4654. T.F. was supported by a doctoral scholarship of the Heinrich F.C. Behr Foundation and A.E. by a doctoral scholarship of German Cancer Aid. The funding body did not have any role in the design of the study; writing of the manuscript; and collection, analysis, or interpretation of data.

## AUTHOR CONTRIBUTIONS

M.F.O. performed experimental assays. M.F.O., D.S., and S.Z. performed ChIP. M.F.O., A.E., F.C.-A., S.G., S.B., and T.F. performed bioinformatics analyses. A.M. and S.O. contributed to *in vitro* and *in vivo* experiments. J.S.G. assembled the survival dataset and implemented the survival analysis tool. J.A. and A.S. provided additional EwS expression data with clinical annotation. K.S., S.M.H., and E.R. performed MethylationEPIC BeadChip analyses and mass spectrometry. M.S. supported methylation data interpretation. S.M.P. and M.G. supported WGS and HipSTR analyses. J.K., D.A.Z., and R.V. implemented the ESCLA in R2. T.G.P.G., O.D., and T.K. provided lab infrastructure. M.F.O. and T.G.P.G. designed the display items and wrote the paper. T.G.P.G. conceived and supervised the study. All authors read the final version of this manuscript and approved it for submission.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

## INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: December 2, 2021
Revised: August 25, 2022
Accepted: November 8, 2022
Published: December 6, 2022

## REFERENCES

1. Kaatsch, P., Strothotte, J., Becker, C., Bielack, S., Dirksen, U., and Blettner, M. (2016). Pediatric bone tumors in Germany from 1987 to 2011: incidence rates, time trends and survival. Acta Oncol. *55*, 1145–1151. https://doi.org/10.1080/0284186X.2016.1195509.
2. Ewing, J. (2006). The classic: diffuse endothelioma of bone. Clin. Orthop. Relat. Res. *450*, 25–27. https://doi.org/10.1097/01.blo.0000229311.36007.c7.
3. Anderson, N.D., de Borja, R., Young, M.D., Fuligni, F., Rosic, A., Roberts, N.D., Hajjar, S., Layeghifard, M., Novokmet, A., Kowalski, P.E., et al. (2018). Rearrangement bursts generate canonical gene fusions in bone and soft tissue tumors. Science *361*, eaam8419. https://doi.org/10.1126/science.aam8419.
4. Delattre, O., Zucman, J., Plougastel, B., Desmaze, C., Melot, T., Peter, M., Kovar, H., Joubert, I., de Jong, P., and Rouleau, G. (1992). Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours. Nature *359*, 162–165. https://doi.org/10.1038/359162a0.
5. Sorensen, P.H., Lessnick, S.L., Lopez-Terrada, D., Liu, X.F., Triche, T.J., and Denny, C.T. (1994). A second Ewing's sarcoma translocation, t(21;22), fuses the EWS gene to another ETS-family transcription factor, ERG. Nat. Genet. *6*, 146–151. https://doi.org/10.1038/ng0294-146.
6. Zucman, J., Melot, T., Desmaze, C., Ghysdael, J., Plougastel, B., Peter, M., Zucker, J.M., Triche, T.J., Sheer, D., and Turc-Carel, C. (1993). Combinatorial generation of variable fusion proteins in the Ewing family of tumours. EMBO J. *12*, 4481–4487.
7. Boulay, G., Volorio, A., Iyer, S., Broye, L.C., Stamenkovic, I., Riggi, N., and Rivera, M.N. (2018). Epigenome editing of microsatellite repeats defines tumor-specific enhancer functions and dependencies. Genes Dev. *32*, 1008–1019. https://doi.org/10.1101/gad.315192.118.
8. Grünewald, T.G.P., Cidre-Aranaz, F., Surdez, D., Tomazou, E.M., de Álava, E., Kovar, H., Sorensen, P.H., Delattre, O., and Dirksen, U. (2018). Ewing sarcoma. Nat. Rev. Dis. Primers *4*, 5. https://doi.org/10.1038/s41572-018-0003-x.
9. Gangwal, K., Sankar, S., Hollenhorst, P.C., Kinsey, M., Haroldsen, S.C., Shah, A.A., Boucher, K.M., Watkins, W.S., Jorde, L.B., Graves, B.J., and Lessnick, S.L. (2008). Microsatellites as EWS/FLI response elements in Ewing's sarcoma. Proc. Natl. Acad. Sci. USA *105*, 10149–10154. https://doi.org/10.1073/pnas.0801073105.
10. Guillon, N., Tirode, F., Boeva, V., Zynovyev, A., Barillot, E., and Delattre, O. (2009). The oncogenic EWS-FLI1 protein binds in vivo GGAA microsatellite sequences with potential transcriptional activation function. PLoS One *4*, e4932. https://doi.org/10.1371/journal.pone.0004932.
11. Brohl, A.S., Solomon, D.A., Chang, W., Wang, J., Song, Y., Sindiri, S., Patidar, R., Hurd, L., Chen, L., Shern, J.F., et al. (2014). The genomic

landscape of the ewing sarcoma family of tumors reveals recurrent STAG2 mutation. PLoS Genet. *10*, e1004475. https://doi.org/10.1371/journal.pgen.1004475.

12. Crompton, B.D., Stewart, C., Taylor-Weiner, A., Alexe, G., Kurek, K.C., Calicchio, M.L., Kiezun, A., Carter, S.L., Shukla, S.A., Mehta, S.S., et al. (2014). The genomic landscape of pediatric Ewing sarcoma. Cancer Discov. *4*, 1326–1341. https://doi.org/10.1158/2159-8290.CD-13-1037.
13. Sheffield, N.C., Pierron, G., Klughammer, J., Datlinger, P., Schönegger, A., Schuster, M., Hadler, J., Surdez, D., Guillemot, D., Lapouble, E., et al. (2017). DNA methylation heterogeneity defines a disease spectrum in Ewing sarcoma. Nat. Med. *23*, 386–395. https://doi.org/10.1038/nm.4273.
14. Tirode, F., Surdez, D., Ma, X., Parker, M., Le Deley, M.C., Bahrami, A., Zhang, Z., Lapouble, E., Grossetête-Lalami, S., Rusch, M., et al. (2014). Genomic landscape of Ewing sarcoma defines an aggressive subtype with co-association of STAG2 and TP53 mutations. Cancer Discov. *4*, 1342–1353. https://doi.org/10.1158/2159-8290.CD-14-0622.
15. Bosma, S.E., Ayu, O., Fiocco, M., Gelderblom, H., and Dijkstra, P.D.S. (2018). Prognostic factors for survival in Ewing sarcoma: a systematic review. Surg. Oncol. *27*, 603–610. https://doi.org/10.1016/j.suronc.2018.07.016.
16. Gaspar, N., Hawkins, D.S., Dirksen, U., Lewis, I.J., Ferrari, S., Le Deley, M.-C., Kovar, H., Grimer, R., Whelan, J., Claude, L., et al. (2015). Ewing sarcoma: current management and future approaches through collaboration. J. Clin. Oncol. *33*, 3036–3046. https://doi.org/10.1200/JCO.2014.59.5256.
17. Kridis, W.B., Toumi, N., Chaari, H., Khanfir, A., Ayadi, K., Keskes, H., Boudawara, T., Daoud, J., and Frikha, M. (2017). A review of ewing sarcoma treatment: is it still a subject of debate? Rev. Recent Clin. Trials *12*, 19–23. https://doi.org/10.2174/1574887112666170120100147.
18. Grünewald, T.G.P., Bernard, V., Gilardi-Hebenstreit, P., Raynal, V., Surdez, D., Aynaud, M.-M., Mirabeau, O., Cidre-Aranaz, F., Tirode, F., Zaidi, S., et al. (2015). Chimeric EWSR1-FLI1 regulates the Ewing sarcoma susceptibility gene EGR2 via a GGAA microsatellite. Nat. Genet. *47*, 1073–1078. https://doi.org/10.1038/ng.3363.
19. Musa, J., Cidre-Aranaz, F., Aynaud, M.-M., Orth, M.F., Knott, M.M.L., Mirabeau, O., Mazor, G., Varon, M., Hölting, T.L.B., Grossetête, S., et al. (2019). Cooperation of cancer drivers with regulatory germline variants shapes clinical outcomes. Nat. Commun. *10*, 4128. https://doi.org/10.1038/s41467-019-12071-2.
20. Gillet, J.-P., Varma, S., and Gottesman, M.M. (2013). The clinical relevance of cancer cell lines. J. Natl. Cancer Inst. *105*, 452–458. https://doi.org/10.1093/jnci/djt007.
21. Mirabelli, P., Coppola, L., and Salvatore, M. (2019). Cancer Cell Lines Are Useful Model Systems for Medical Research. Cancers *11*. https://doi.org/10.3390/cancers11081098.
22. Garnett, M.J., Edelman, E.J., Heidorn, S.J., Greenman, C.D., Dastur, A., Lau, K.W., Greninger, P., Thompson, I.R., Luo, X., Soares, J., et al. (2012). Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature *483*, 570–575. https://doi.org/10.1038/nature11005.
23. Minas, T.Z., Surdez, D., Javaheri, T., Tanaka, M., Howarth, M., Kang, H.-J., Han, J., Han, Z.-Y., Sax, B., Kream, B.E., et al. (2017). Combined experience of six independent laboratories attempting to create an Ewing sarcoma mouse model. Oncotarget *8*, 34141–34163. https://doi.org/10.18632/oncotarget.9388.
24. Giovannini, M., Biegel, J.A., Serra, M., Wang, J.Y., Wei, Y.H., Nycum, L., Emanuel, B.S., and Evans, G.A. (1994). EWS-erg and EWS-Fli1 fusion transcripts in Ewing's sarcoma and primitive neuroectodermal tumors with variant translocations. J. Clin. Invest. *94*, 489–496. https://doi.org/10.1172/JCI117360.
25. Elzi, D.J., Song, M., Houghton, P.J., Chen, Y., and Shiio, Y. (2015). The role of FLI-1-EWS, a fusion gene reciprocal to EWS-FLI-1, in Ewing sarcoma. Genes Cancer *6*, 452–461. https://doi.org/10.18632/genesandcancer.86.
26. Chen, K., Wallis, J.W., McLellan, M.D., Larson, D.E., Kalicki, J.M., Pohl, C.S., McGrath, S.D., Wendl, M.C., Zhang, Q., Locke, D.P., et al. (2009). BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. Nat. Methods *6*, 677–681. https://doi.org/10.1038/nmeth.1363.
27. Hattinger, C.M., Pötschger, U., Tarkkanen, M., Squire, J., Zielenska, M., Kiuru-Kuhlefelt, S., Kager, L., Thorner, P., Knuutila, S., Niggli, F.K., et al. (2002). Prognostic impact of chromosomal aberrations in Ewing tumours. Br. J. Cancer *86*, 1763–1769. https://doi.org/10.1038/sj.bjc.6600332.
28. Mackintosh, C., Ordóñez, J.L., García-Domínguez, D.J., Sevillano, V., Llombart-Bosch, A., Szuhai, K., Scotlandi, K., Alberghini, M., Sciot, R., Sinnaeve, F., et al. (2012). 1q gain and CDT2 overexpression underlie an aggressive and highly proliferative form of Ewing sarcoma. Oncogene *31*, 1287–1298. https://doi.org/10.1038/onc.2011.317.
29. Augusto Corrêa Dos Santos, R., Goldman, G.H., and Riaño-Pachón, D.M. (2017). ploidyNGS: visually exploring ploidy with Next Generation Sequencing data. Bioinformatics *33*, 2575–2576. https://doi.org/10.1093/bioinformatics/btx204.
30. May, W.A., Grigoryan, R.S., Keshelava, N., Cabral, D.J., Christensen, L.L., Jenabi, J., Ji, L., Triche, T.J., Lawlor, E.R., and Reynolds, C.P. (2013). Characterization and drug resistance patterns of Ewing's sarcoma family tumor cell lines. PLoS One *8*, e80060. https://doi.org/10.1371/journal.pone.0080060.
31. Riggi, N., Knoechel, B., Gillespie, S.M., Rheinbay, E., Boulay, G., Suvà, M.L., Rossetti, N.E., Boonseng, W.E., Oksuz, O., Cook, E.B., et al. (2014). EWS-FLI1 utilizes divergent chromatin remodeling mechanisms to directly activate or repress enhancer elements in Ewing sarcoma. Cancer Cell *26*, 668–681. https://doi.org/10.1016/j.ccell.2014.10.004.
32. Willems, T., Zielinski, D., Yuan, J., Gordon, A., Gymrek, M., and Erlich, Y. (2017). Genome-wide profiling of heritable and *de novo* STR variations. Nat. Methods *14*, 590–592. https://doi.org/10.1038/nmeth.4267.
33. Johnson, K.M., Taslim, C., Saund, R.S., and Lessnick, S.L. (2017). Identification of two types of GGAA-microsatellites and their roles in EWS/FLI binding and gene regulation in Ewing sarcoma. PLoS One *12*, e0186275. https://doi.org/10.1371/journal.pone.0186275.
34. Fadul, J., Bell, R., Hoffman, L.M., Beckerle, M.C., Engel, M.E., and Lessnick, S.L. (2015). EWS/FLI utilizes NKX2-2 to repress mesenchymal features of Ewing sarcoma. Genes Cancer *6*, 129–143. https://doi.org/10.18632/genesandcancer.57.
35. Luo, W., Xu, C., Ayello, J., Dela Cruz, F., Rosenblum, J.M., Lessnick, S.L., and Cairo, M.S. (2018). Protein phosphatase 1 regulatory subunit 1A in ewing sarcoma tumorigenesis and metastasis. Oncogene *37*, 798–809. https://doi.org/10.1038/onc.2017.378.
36. Owen, L.A., Kowalewski, A.A., and Lessnick, S.L. (2008). EWS/FLI mediates transcriptional repression via NKX2.2 during oncogenic transformation in Ewing's sarcoma. PLoS One *3*, e1965. https://doi.org/10.1371/journal.pone.0001965.
37. Riggi, N., Suvà, M.L., Suvà, D., Cironi, L., Provero, P., Tercier, S., Joseph, J.-M., Stehle, J.-C., Baumer, K., Kindler, V., and Stamenkovic, I. (2008). EWS-FLI-1 expression triggers a ewing's sarcoma initiation program in primary human mesenchymal stem cells. Cancer Res. *68*, 2176–2185. https://doi.org/10.1158/0008-5472.CAN-07-1761.
38. García-Aragoncillo, E., Carrillo, J., Lalli, E., Agra, N., Gómez-López, G., Pestaña, A., and Alonso, J. (2008). DAX1, a direct target of EWS/FLI1 oncoprotein, is a principal regulator of cell-cycle progression in Ewing's tumor cells. Oncogene *27*, 6034–6043. https://doi.org/10.1038/onc.2008.203.
39. Marchetto, A., Ohmura, S., Orth, M.F., Knott, M.M.L., Colombo, M.V., Arrigoni, C., Bardinet, V., Saucier, D., Wehweck, F.S., Li, J., et al. (2020). Oncogenic hijacking of a developmental transcription factor evokes vulnerability toward oxidative stress in Ewing sarcoma. Nat. Commun. *11*, 2423. https://doi.org/10.1038/s41467-020-16244-2.
40. Monument, M.J., Johnson, K.M., McIlvaine, E., Abegglen, L., Watkins, W.S., Jorde, L.B., Womer, R.B., Beeler, N., Monovich, L., Lawlor, E.R.,

et al. (2014). Clinical and biochemical function of polymorphic NR0B1 GGAA-microsatellites in Ewing sarcoma: a report from the Children's Oncology Group. PLoS One *9*, e104378. https://doi.org/10.1371/journal.pone.0104378.

41. Tomazou, E.M., Sheffield, N.C., Schmidl, C., Schuster, M., Schönegger, A., Datlinger, P., Kubicek, S., Bock, C., and Kovar, H. (2015). Epigenome mapping reveals distinct modes of gene regulation and widespread enhancer reprogramming by the oncogenic fusion protein EWS-FLI1. Cell Rep. *10*, 1082–1095. https://doi.org/10.1016/j.celrep.2015.01.042.

42. Baldauf, M.C., Orth, M.F., Dallmayer, M., Marchetto, A., Gerke, J.S., Rubio, R.A., Kiran, M.M., Musa, J., Knott, M.M.L., Ohmura, S., et al. (2018). Robust diagnosis of Ewing sarcoma by immunohistochemical detection of super-enhancer-driven EWSR1-ETS targets. Oncotarget *9*, 1587–1601. https://doi.org/10.18632/oncotarget.20098.

43. Dauphinot, L., De Oliveira, C., Melot, T., Sevenet, N., Thomas, V., Weissman, B.E., and Delattre, O. (2001). Analysis of the expression of cell cycle regulators in Ewing cell lines: EWS-FLI-1 modulates p57KIP2 and c-Myc expression. Oncogene *20*, 3258–3265. https://doi.org/10.1038/sj.onc.1204437.

44. Magro, G., Brancato, F., Musumeci, G., Alaggio, R., Parenti, R., and Salvatorelli, L. (2015). Cyclin D1 is a useful marker for soft tissue Ewing's sarcoma/peripheral Primitive Neuroectodermal Tumor in children and adolescents: a comparative immunohistochemical study with rhabdomyosarcoma. Acta Histochem. *117*, 460–467. https://doi.org/10.1016/j.acthis.2015.01.005.

45. Surdez, D., Benetkiewicz, M., Perrin, V., Han, Z.-Y., Pierron, G., Ballet, S., Lamoureux, F., Rédini, F., Decouvelaere, A.-V., Daudigeos-Dubus, E., et al. (2012). Targeting the EWSR1-FLI1 oncogene-induced protein kinase PKC-β abolishes ewing sarcoma growth. Cancer Res. *72*, 4494–4503. https://doi.org/10.1158/0008-5472.CAN-12-0371.

46. Fang, J., Wang, J., Yu, L., and Xu, W. (2021). Role of HOXC10 in cancer. Front. Oncol. *11*, 684021. https://doi.org/10.3389/fonc.2021.684021.

47. von Heyking, K., Roth, L., Ertl, M., Schmidt, O., Calzada-Wack, J., Neff, F., Lawlor, E.R., Burdach, S., and Richter, G.H. (2016). The posterior HOXD locus: its contribution to phenotype and malignancy of Ewing sarcoma. Oncotarget *7*, 41767–41780. https://doi.org/10.18632/oncotarget.9702.

48. Pfaltzgraff, E.R., Apfelbaum, A., Kassa, A.P., Song, J.Y., Jiang, W., Suhan, T.K., Wellik, D.M., and Lawlor, E.R. (2019). Anatomic origin of osteochondrogenic progenitors impacts sensitivity to EWS-FLI1-induced transformation. Cancers *11*, E313. https://doi.org/10.3390/cancers11030313.

49. Svoboda, L.K., Harris, A., Bailey, N.J., Schwentner, R., Tomazou, E., von Levetzow, C., Magnuson, B., Ljungman, M., Kovar, H., and Lawlor, E.R. (2014). Overexpression of HOX genes is prevalent in Ewing sarcoma and is associated with altered epigenetic regulation of developmental transcription programs. Epigenetics *9*, 1613–1625. https://doi.org/10.4161/15592294.2014.988048.

50. Svoboda, L.K., Bailey, N., Van Noord, R.A., Krook, M.A., Harris, A., Cramer, C., Jasman, B., Patel, R.M., Thomas, D., Borkin, D., et al. (2017). Tumorigenicity of Ewing sarcoma is critically dependent on the trithorax proteins MLL1 and menin. Oncotarget *8*, 458–471. https://doi.org/10.18632/oncotarget.13444.

51. Hueber, S.D., Weiller, G.F., Djordjevic, M.A., and Frickey, T. (2010). Improving Hox protein classification across the major model organisms. PLoS One *5*, e10820. https://doi.org/10.1371/journal.pone.0010820.

52. Wiederschain, D., Wee, S., Chen, L., Loo, A., Yang, G., Huang, A., Chen, Y., Caponigro, G., Yao, Y.-M., Lengauer, C., et al. (2009). Single-vector inducible lentiviral RNAi system for oncology target validation. Cell Cycle *8*, 498–504. https://doi.org/10.4161/cc.8.3.7701.

53. Franzetti, G.-A., Laud-Duval, K., van der Ent, W., Brisac, A., Irondelle, M., Aubert, S., Dirksen, U., Bouvier, C., de Pinieux, G., Snaar-Jagalska, E., et al. (2017). Cell-to-cell heterogeneity of EWSR1-FLI1 activity determines proliferation/migration choices in Ewing sarcoma cells. Oncogene *36*, 3505–3514. https://doi.org/10.1038/onc.2016.498.

54. Tanner, J.M., Bensard, C., Wei, P., Krah, N.M., Schell, J.C., Gardiner, J., Schiffman, J., Lessnick, S.L., and Rutter, J. (2017). EWS/FLI is a master regulator of metabolic reprogramming in ewing sarcoma. Mol. Cancer Res. *15*, 1517–1530. https://doi.org/10.1158/1541-7786.MCR-17-0182.

55. Agra, N., Cidre, F., García-García, L., de la Parra, J., and Alonso, J. (2013). Lysyl oxidase is downregulated by the EWS/FLI1 oncoprotein and its propeptide domain displays tumor supressor activities in ewing sarcoma cells. PLoS One *8*, e66281. https://doi.org/10.1371/journal.pone.0066281.

56. Cidre-Aranaz, F., and Alonso, J. (2015). EWS/FLI1 target genes and therapeutic opportunities in ewing sarcoma. Front. Oncol. *5*, 162. https://doi.org/10.3389/fonc.2015.00162.

57. Carrillo, J., García-Aragoncillo, E., Azorín, D., Agra, N., Sastre, A., González-Mediero, I., García-Miguel, P., Pestaña, A., Gallego, S., Segura, D., and Alonso, J. (2007). Cholecystokinin down-regulation by RNA interference impairs ewing tumor growth. Clin. Cancer Res. *13*, 2429–2440. https://doi.org/10.1158/1078-0432.CCR-06-1762.

58. Charville, G.W., Wang, W.-L., Ingram, D.R., Roy, A., Thomas, D., Patel, R.M., Hornick, J.L., van de Rijn, M., and Lazar, A.J. (2017). EWSR1 fusion proteins mediate PAX7 expression in Ewing sarcoma. Mod. Pathol. *30*, 1312–1320. https://doi.org/10.1038/modpathol.2017.49.

59. Cidre-Aranaz, F., Grünewald, T.G.P., Surdez, D., García-García, L., Carlos Lázaro, J., Kirchner, T., González-González, L., Sastre, A., García-Miguel, P., López-Pérez, S.E., et al. (2017). EWS-FLI1-mediated suppression of the RAS-antagonist Sprouty 1 (SPRY1) confers aggressiveness to Ewing sarcoma. Oncogene *36*, 766–776. https://doi.org/10.1038/onc.2016.244.

60. Grunewald, T.G.P., Diebold, I., Esposito, I., Plehm, S., Hauer, K., Thiel, U., da Silva-Buttkus, P., Neff, F., Unland, R., Müller-Tidow, C., et al. (2012). STEAP1 is associated with the invasive and oxidative stress phenotype of Ewing tumors. Mol. Cancer Res. *10*, 52–65. https://doi.org/10.1158/1541-7786.MCR-11-0524.

61. Koelsche, C., Kriegsmann, M., Kommoss, F.K.F., Stichel, D., Kriegsmann, K., Vokuhl, C., Grünewald, T.G.P., Romero-Pérez, L., Kirchner, T., de Alava, E., et al. (2019). DNA methylation profiling distinguishes Ewing-like sarcoma with EWSR1-NFATc2 fusion from Ewing sarcoma. J. Cancer Res. Clin. Oncol. *145*, 1273–1281. https://doi.org/10.1007/s00432-019-02895-2.

62. Graham, G.T., Selvanathan, S.P., Zöllner, S.K., Stahl, E., Shlien, A., Caplen, N.J., Üren, A., and Toretsky, J.A. (2022). Comprehensive profiling of mRNA splicing indicates that GC content signals altered cassette exon inclusion in Ewing sarcoma. NAR Cancer *4*, zcab052. https://doi.org/10.1093/narcan/zcab052.

63. Selvanathan, S.P., Graham, G.T., Erkizan, H.V., Dirksen, U., Natarajan, T.G., Dakic, A., Yu, S., Liu, X., Paulsen, M.T., Ljungman, M.E., et al. (2015). Oncogenic fusion protein EWS-FLI1 is a network hub that regulates alternative splicing. Proc. Natl. Acad. Sci. USA *112*, E1307–E1316. https://doi.org/10.1073/pnas.1500536112.

64. Selvanathan, S.P., Graham, G.T., Grego, A.R., Baker, T.M., Hogg, J.R., Simpson, M., Batish, M., Crompton, B., Stegmaier, K., Tomazou, E.M., et al. (2019). EWS–FLI1 modulated alternative splicing of ARID1A reveals novel oncogenic function through the BAF complex. Nucleic Acids Res. *47*, 9619–9636. https://doi.org/10.1093/nar/gkz699.

65. Zöllner, S.K., Selvanathan, S.P., Graham, G.T., Commins, R.M.T., Hong, S.H., Moseley, E., Parks, S., Haladyna, J.N., Erkizan, H.V., Dirksen, U., et al. (2017). Inhibition of the oncogenic fusion protein EWS-FLI1 causes G2-M cell cycle arrest and enhanced vincristine sensitivity in Ewing's sarcoma. Sci. Signal. *10*, eaam8429. https://doi.org/10.1126/scisignal.aam8429.

66. Lovén, J., Hoke, H.A., Lin, C.Y., Lau, A., Orlando, D.A., Vakoc, C.R., Bradner, J.E., Lee, T.I., and Young, R.A. (2013). Selective inhibition of

tumor oncogenes by disruption of super-enhancers. Cell *153*, 320–334. https://doi.org/10.1016/j.cell.2013.03.036.

67. Hancock, J.D., and Lessnick, S.L. (2008). A transcriptional profiling meta-analysis reveals a core EWS-FLI gene expression signature. Cell Cycle *7*, 250–256. https://doi.org/10.4161/cc.7.2.5229.
68. Bilke, S., Schwentner, R., Yang, F., Kauer, M., Jug, G., Walker, R.L., Davis, S., Zhu, Y.J., Pineda, M., Meltzer, P.S., and Kovar, H. (2013). Oncogenic ETS fusions deregulate E2F3 target genes in Ewing sarcoma and prostate cancer. Genome Res. *23*, 1797–1809. https://doi.org/10.1101/gr.151340.112.
69. Deng, Q., Natesan, R., Cidre-Aranaz, F., Arif, S., Liu, Y., Rasool, R.U., Wang, P., Mitchell-Velasquez, E., Das, C.K., Vinca, E., et al. (2022). Oncofusion-driven de novo enhancer assembly promotes malignancy in Ewing sarcoma via aberrant expression of the stereociliary protein LOXHD1. Cell Rep. *39*, 110971. https://doi.org/10.1016/j.celrep.2022.110971.
70. Zhang, H.-F., Hughes, C.S., Li, W., He, J.-Z., Surdez, D., El-Naggar, A.M., Cheng, H., Prudova, A., Delaidelli, A., Negri, G.L., et al. (2021). Proteomic screens for suppressors of anoikis identify IL1RAP as a promising surface target in Ewing sarcoma. Cancer Discov. *11*, 2884–2903. https://doi.org/10.1158/2159-8290.CD-20-1690.
71. Yang, L., Hu, H.-M., Zielinska-Kwiatkowska, A., and Chansky, H.A. (2010). FOXO1 is a direct target of EWS-Fli1 oncogenic fusion protein in Ewing's sarcoma cells. Biochem. Biophys. Res. Commun. *402*, 129–134. https://doi.org/10.1016/j.bbrc.2010.09.129.
72. Sharrocks, A.D. (2001). The ETS-domain transcription factor family. Nat. Rev. Mol. Cell Biol. *2*, 827–837. https://doi.org/10.1038/35099076.
73. Surdez, D., Zaidi, S., Grossetête, S., Laud-Duval, K., Ferre, A.S., Mous, L., Vourc'h, T., Tirode, F., Pierron, G., Raynal, V., et al. (2021). STAG2 mutations alter CTCF-anchored loop extrusion, reduce cis-regulatory interactions and EWSR1-FLI1 activity in Ewing sarcoma. Cancer Cell *39*, 810–826.e9. https://doi.org/10.1016/j.ccell.2021.04.001.
74. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589. https://doi.org/10.1016/j.molcel.2010.05.004.
75. Beck, R., Monument, M.J., Watkins, W.S., Smith, R., Boucher, K.M., Schiffman, J.D., Jorde, L.B., Randall, R.L., and Lessnick, S.L. (2012). EWS/FLI-responsive GGAA microsatellites exhibit polymorphic differences between European and African populations. Cancer Genet. *205*, 304–312. https://doi.org/10.1016/j.cancergen.2012.04.004.
76. Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. *27*, 573–580. https://doi.org/10.1093/nar/27.2.573.
77. Schwentner, R., Papamarkou, T., Kauer, M.O., Stathopoulos, V., Yang, F., Bilke, S., Meltzer, P.S., Girolami, M., and Kovar, H. (2015). EWS-FLI1 employs an E2F switch to drive target gene expression. Nucleic Acids Res. *43*, 2780–2789. https://doi.org/10.1093/nar/gkv123.
78. Boulay, G., Sandoval, G.J., Riggi, N., Iyer, S., Buisson, R., Naigles, B., Awad, M.E., Rengarajan, S., Volorio, A., McBride, M.J., et al. (2017). Cancer-specific retargeting of BAF complexes by a prion-like domain. Cell *171*, 163–178.e19. https://doi.org/10.1016/j.cell.2017.07.036.
79. Patel, M., Simon, J.M., Iglesia, M.D., Wu, S.B., McFadden, A.W., Lieb, J.D., and Davis, I.J. (2012). Tumor-specific retargeting of an oncogenic transcription factor chimera results in dysregulation of chromatin and transcription. Genome Res. *22*, 259–270. https://doi.org/10.1101/gr.125666.111.
80. Sole, A., Grossetête, S., Heintzé, M., Babin, L., Zaïdi, S., Revy, P., Renouf, B., De Cian, A., Giovannangeli, C., Pierre-Eugène, C., et al. (2021). Unraveling ewing sarcoma tumorigenesis originating from patient-derived mesenchymal stem cells. Cancer Res. *81*, 4994–5006. https://doi.org/10.1158/0008-5472.CAN-20-3837.
81. Adane, B., Alexe, G., Seong, B.K.A., Lu, D., Hwang, E.E., Hnisz, D., Lareau, C.A., Ross, L., Lin, S., Dela Cruz, F.S., et al. (2021). STAG2 loss rewires oncogenic and developmental programs to promote metastasis in Ewing sarcoma. Cancer Cell *39*, 827–844.e10. https://doi.org/10.1016/j.ccell.2021.05.007.
82. Vibert, J., Saulnier, O., Collin, C., Petit, F., Borgman, K.J.E., Vigneau, J., Gautier, M., Zaidi, S., Pierron, G., Watson, S., et al. (2022). Oncogenic chimeric transcription factors drive tumor-specific transcription, processing, and translation of silent genomic regions. Mol. Cell *82*, 2458–2471.e9. https://doi.org/10.1016/j.molcel.2022.04.019.
83. Worch, J., Cyrus, J., Goldsby, R., Matthay, K.K., Neuhaus, J., and DuBois, S.G. (2011). Racial differences in the incidence of mesenchymal tumors associated with EWSR1 translocation. Cancer Epidemiol. Biomarkers Prev. *20*, 449–453. https://doi.org/10.1158/1055-9965.EPI-10-1170.
84. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at arXiv. https://doi.org/10.48550/arXiv.1303.3997.
85. Andrews, S. (2019). FastQC A Quality Control Tool for High Throughput Sequence Data (abraham Institute of Bioinformatics). http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
86. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
87. Talevich, E., Shain, A.H., Botton, T., and Bastian, B.C. (2016). CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. PLoS Comput. Biol. *12*, e1004873. https://doi.org/10.1371/journal.pcbi.1004873.
88. Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics *27*, 2987–2993. https://doi.org/10.1093/bioinformatics/btr509.
89. Cingolani, P., Patel, V.M., Coon, M., Nguyen, T., Land, S.J., Ruden, D.M., and Lu, X. (2012). Using Drosophila melanogaster as a model for genotoxic chemical mutational studies with a new program. Front. Genet. *3*, 35. https://doi.org/10.3389/fgene.2012.00035.
90. Layer, R.M., Chiang, C., Quinlan, A.R., and Hall, I.M. (2014). LUMPY: a probabilistic framework for structural variant discovery. Genome Biol. *15*, R84. https://doi.org/10.1186/gb-2014-15-6-r84.
91. Cameron, D.L., Baber, J., Shale, C., Valle-Inclan, J.E., Besselink, N., Cuppen, E., Priestley, P., and Papenfuss, A.T. (2020). GRIDSS2: harnessing the power of phasing and single breakends in somatic structural variant detection. Bioinformatics. https://doi.org/10.1101/2020.07.09.196527.
92. Baca, S.C., Prandi, D., Lawrence, M.S., Mosquera, J.M., Romanel, A., Drier, Y., Park, K., Kitabayashi, N., MacDonald, T.Y., Ghandi, M., et al. (2013). Punctuated evolution of prostate cancer genomes. Cell *153*, 666–677. https://doi.org/10.1016/j.cell.2013.03.021.
93. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. https://doi.org/10.1038/nmeth.1923.
94. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-seq (MACS). Genome Biol. *9*, R137. https://doi.org/10.1186/gb-2008-9-9-r137.
95. Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform. *14*, 178–192. https://doi.org/10.1093/bib/bbs017.
96. Polit, L., Kerdivel, G., Gregoricchio, S., Esposito, M., Guillouf, C., and Boeva, V. (2021). CHIPIN: ChIP-seq inter-sample normalization based

on signal invariance across transcriptionally constant genes. BMC Bioinf. *22*, 407. https://doi.org/10.1186/s12859-021-04320-3.

97. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.
98. Aryee, M.J., Jaffe, A.E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A.P., Hansen, K.D., and Irizarry, R.A. (2014). Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics *30*, 1363–1369. https://doi.org/10.1093/bioinformatics/btu049.
99. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550. https://doi.org/10.1073/pnas.0506580102.
100. Capper, D., Jones, D.T.W., Sill, M., Hovestadt, V., Schrimpf, D., Sturm, D., Koelsche, C., Sahm, F., Chavez, L., Reuss, D.E., et al. (2018). DNA methylation-based classification of central nervous system tumours. Nature *555*, 469–474. https://doi.org/10.1038/nature26000.
101. Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinf. *9*, 559. https://doi.org/10.1186/1471-2105-9-559.
102. Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: an information aesthetic for comparative genomics. Genome Res. *19*, 1639–1645. https://doi.org/10.1101/gr.092759.109.
103. Whang-Peng, J., Triche, T.J., Knutsen, T., Miser, J., Kao-Shan, S., Tsai, S., and Israel, M.A. (1986). Cytogenetic characterization of selected small round cell tumors of childhood. Cancer Genet. Cytogenet. *21*, 185–208. https://doi.org/10.1016/0165-4608(86)90001-4.
104. Cidre-Aranaz, F., and Ohmura, S. (2021). Tumor growth analysis of ewing sarcoma cell lines using subcutaneous xenografts in mice. Methods Mol. Biol. *2226*, 191–199. https://doi.org/10.1007/978-1-0716-1020-6_15.
105. Marchetto, A., and Romero-Pérez, L. (2021). Western blot analysis in ewing sarcoma. Methods Mol. Biol. *2226*, 15–25. https://doi.org/10.1007/978-1-0716-1020-6_2.
106. Marcilla, D., Machado, I., Grünewald, T.G.P., Llombart-Bosch, A., and de Álava, E. (2021). (Immuno)histological analysis of ewing sarcoma. Methods Mol. Biol. *2226*, 49–64. https://doi.org/10.1007/978-1-0716-1020-6_5.
107. Remmele, W., and Stegner, H.E. (1987). [Recommendation for uniform definition of an immunoreactive score (IRS) for immunohistochemical estrogen receptor detection (ER-ICA) in breast cancer tissue]. Pathologe *8*, 138–140.
108. Abyzov, A., Urban, A.E., Snyder, M., and Gerstein, M. (2011). CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. Genome Res. *21*, 974–984. https://doi.org/10.1101/gr.114876.110.
109. Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. *38*, e164. https://doi.org/10.1093/nar/gkq603.
110. Cameron, D.L., Schröder, J., Penington, J.S., Do, H., Molania, R., Dobrovic, A., Speed, T.P., and Papenfuss, A.T. (2017). GRIDSS: sensitive and specific genomic rearrangement detection using positional de Bruijn graph assembly. Genome Res. *27*, 2050–2060. https://doi.org/10.1101/gr.222109.117.
111. Kerdivel, G., and Boeva, V. (2021). Chromatin immunoprecipitation followed by next-generation sequencing (ChIP-Seq) analysis in ewing sarcoma. Methods Mol. Biol. *2226*, 265–284. https://doi.org/10.1007/978-1-0716-1020-6_21.
112. Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U., and Speed, T.P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics *4*, 249–264. https://doi.org/10.1093/biostatistics/4.2.249.
113. Lepper, M.F., Ohmayer, U., von Toerne, C., Maison, N., Ziegler, A.-G., and Hauck, S.M. (2018). Proteomic landscape of patient-derived CD4+ T cells in recent-onset type 1 diabetes. J. Proteome Res. *17*, 618–634. https://doi.org/10.1021/acs.jproteome.7b00712.
114. Waszak, S.M., Robinson, G.W., Gudenas, B.L., Smith, K.S., Forget, A., Kojic, M., Garcia-Lopez, J., Hadley, J., Hamilton, K.V., Indersie, E., et al. (2020). Germline elongator mutations in sonic hedgehog medulloblastoma. Nature *580*, 396–401. https://doi.org/10.1038/s41586-020-2164-5.
115. Mi, H., Muruganujan, A., Ebert, D., Huang, X., and Thomas, P.D. (2019). PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Res. *47*, D419–D426. https://doi.org/10.1093/nar/gky1038.
116. Sannino, G., Marchetto, A., Ranft, A., Jabar, S., Zacherl, C., Alba-Rubio, R., Stein, S., Wehweck, F.S., Kiran, M.M., Hölting, T.L.B., et al. (2019). Gene expression and immunohistochemical analyses identify SOX2 as major risk factor for overall survival and relapse in Ewing sarcoma patients. EBioMedicine *47*, 156–162. https://doi.org/10.1016/j.ebiom.2019.08.002.
117. Dai, M., Wang, P., Boyd, A.D., Kostov, G., Athey, B., Jones, E.G., Bunney, W.E., Myers, R.M., Speed, T.P., Akil, H., et al. (2005). Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. Nucleic Acids Res. *33*, e175. https://doi.org/10.1093/nar/gni179.
118. Johnson, W.E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics *8*, 118–127. https://doi.org/10.1093/biostatistics/kxj037.
119. Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., Treviño, V., Shen, H., Laird, P.W., Levine, D.A., et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. Nat. Commun. *4*, 2612. https://doi.org/10.1038/ncomms3612.

## STAR★METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Antibodies | | |
| Rabbit monoclonal anti-ERG (EPR3864) | Abcam, Cambridge, UK | Cat#ab92513; RRID: AB_2630401 |
| Rabbit polyclonal anti-FLI1 | Abcam, Cambridge, UK | Cat#ab15289; RRID: AB_301825 |
| Rabbit polyclonal anti-H3K27ac | Abcam, Cambridge, UK | Cat#ab4729; RRID: AB_2118291 |
| Rabbit polyclonal anti-H3K27me3 | Diagenode, Seraing, Belgium | Cat#C15410069; RRID: AB_2814977 |
| Rabbit polyclonal anti-H3K4me3 | Diagenode, Seraing, Belgium | Cat#C15410003; RRID: AB_2616052 |
| Horse polyclonal anti-mouse/rabbit IgG (Vectastain Elite ABC HRP Kit) | Vector Laboratories, Burlingame, CA, USA | Cat#PK-6200; RRID: AB_2336826 |
| Mouse monoclonal anti-PAX7 | DSHB, Iowa City, IA, USA | Cat#PAX7-c; RRID: AB_2299243 |
| Rabbit monoclonal anti-pMYBL2 (phospho T487; EPR2204Y) | Abcam, Cambridge, UK | Cat#ab76009; RRID: AB_1309969 |
| Horse polyclonal anti-rabbit IgG (ImmPress HRP Polymer Detection Kit) | Vector Laboratories, Burlingame, CA, USA | Cat#MP-7401; RRID: AB_2336529 |
| Rabbit polyclonal anti-SOX6 | Atlas Antibodies, Stockholm, Sweden | Cat#HPA003908; RRID: AB_1080063 |
| Mouse monoclonal anti-GAPDH (6C4) | Santa Cruz Biotechnology, Dallas, TX, USA | Cat#sc-32233; RRID: AB_627679 |
| Mouse monoclonal anti-FLI1 (MRQ-1) | Cell Marque, Rocklin, USA | Cat#254M; RRID: AB_1516868 |
| Goat polyclonal anti-mouse IgG (H + L), HRP | Promega, Maddison, WI, USA | Cat#W402B; RRID: AB_430834 |
| Goat polyclonal anti-rabbit IgG (H + L) HRP | OriGene Technologies, Rockville, MD, USA | Cat#R1364HRP; RRID: AB_10262463 |
| Rabbit monoclonal anti-ERG (EP111) | Cell Marque, Rocklin, USA | Cat#434R |
| Bacterial and virus strains | | |
| Stellar Competent Cells | Clontech, TaKaRa, Saint-Germain-en-Laye, France | Cat#636763 |
| Biological samples | | |
| TMA of *in vivo* expanded EwS cell lines | This paper | N/A |
| Chemicals, peptides, and recombinant proteins | | |
| Doxycycline (dox) for *in vivo* experiments | belapharm, Vechta, Germany | Beladox |
| Doxycycline (dox) | Merck, Darmstadt, Germany | Cat#324385 |
| Lipofectamine RNAiMax | Thermo Fisher, Waltham, MA, USA | Cat#13778030 |
| Critical commercial assays | | |
| High-Capacity cDNA Reverse Transcription Kit | Thermo Fisher, Waltham, MA, USA | Cat#4368814 |
| NucleoSpin RNA kit | Macherey-Nagel, Düren, Germany | Cat#740955 |
| SYBR SELECT Master Mix | Thermo Fisher, Waltham, MA, USA | Cat#4472908 |
| iDeal ChIP-seq kit for Transcription Factors | Diagenode, Seraing, Belgium | Cat#C01010170 |
| NucleoSpin Tissue genomic DNA prep kit | Macherey-Nagel, Düren, Germany | Cat#740952 |
| PureYield Plasmid Midiprep System | Promega, Maddison, WI, USA | Cat#A2492 |
| Deposited data | | |
| WGS data from 18 EwS cell lines; raw data | This paper | SRA: PRJNA610192 |
| Transcriptome data (DNA microarrays) from 18 EwS cell lines in EWSR1-ETS-high and -low condition; raw and processed data | This paper | GEO: GSE176339 |
| DNA methylation data from 18 EwS cell lines in EWSR1-ETS-high and -low condition; raw data | This paper | GEO: GSE176339 |

(*Continued on next page*)

*Continued*

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Proteomics data from mass spectometry of 18 EwS cell lines in EWSR1-ETS-high and -low condition; processed data | This paper | Table S4A |
| ChIP-Seq data for 18 EwS cell lines (targets: FLI1, ERG, H3K27ac, H3K27me3, H3K4me3); raw and processed (BW) data | This paper | GEO: GSE176339 |
| ChIP-Seq data for A-673 and SK-N-MC from Riggi et al. 2014 | GEO | GEO: GSE61953; https://doi.org/10.1016/j.ccell.2014.10.004 |
| Transcriptome data (DNA microarrays) from A-673 cells with E2F2 and GATA2 knockdown; raw and processed data | This paper | GEO: GSE212063 |
| Survival cohort, 196 primary EwS with clinical annotation and gene expression microarray data | lab of T.G.P. Grünewald | GEO: GSE12102, GSE17618, GSE34620, GSE63157; Javier Alonso, Unidad hemato-oncología pediátrica, Hospital Infantil Universitario La Paz, 28029 Madrid, Spain |
| **Experimental models: Cell lines** | | |
| Ewing sarcoma cell line A-673 | ATCC, Manassas, VA, USA | RRID: CVCL_0080 |
| Ewing sarcoma cell line CHLA-10 | Children's Oncology Group, Monrovia, CA, USA | RRID: CVCL_6583 |
| Ewing sarcoma cell line CHLA-25 | Children's Oncology Group, Monrovia, CA, USA | RRID: CVCL_M152 |
| Ewing sarcoma cell line EW-1 | Delattre, O., Paris, France | RRID: CVCL_1208 |
| Ewing sarcoma cell line EW-22 | Delattre, O., Paris, France | RRID: CVCL_1214 |
| Ewing sarcoma cell line EW-24 | Delattre, O., Paris, France | RRID: CVCL_1215 |
| Ewing sarcoma cell line EW-3 | Delattre, O., Paris, France | RRID: CVCL_1216 |
| Ewing sarcoma cell line EW-7 | Delattre, O., Paris, France | RRID: CVCL_1217 |
| Ewing sarcoma cell line MHH-ES-1 | DSMZ, Braunschweig, Germany | RRID: CVCL_1411 |
| Ewing sarcoma cell line MIC | Delattre, O., Paris, France | RRID: CVCL_EI96 |
| Ewing sarcoma cell line POE | Delattre, O., Paris, France | RRID: CVCL_EJ01 |
| Ewing sarcoma cell line RD-ES | DSMZ, Braunschweig, Germany | RRID: CVCL_2169 |
| Ewing sarcoma cell line Rh1 | Delattre, O., Paris, France | RRID: CVCL_1658 |
| Ewing sarcoma cell line SK-ES-1 | DSMZ, Braunschweig, Germany | RRID: CVCL_0627 |
| Ewing sarcoma cell line SK-N-MC | DSMZ, Braunschweig, Germany | RRID: CVCL_0530 |
| Ewing sarcoma cell line TC-106 | Children's Oncology Group, Monrovia, CA, USA | RRID: CVCL_F531 |
| Ewing sarcoma cell line TC-32 | Children's Oncology Group, Monrovia, CA, USA | RRID: CVCL_7151 |
| Ewing sarcoma cell line TC-71 | DSMZ, Braunschweig, Germany | RRID: CVCL_2213 |
| Human HEK293T | DSMZ, Braunschweig, Germany | RRID: CVCL_0063 |
| **Experimental models: Organisms/strains** | | |
| Mouse: NOD/Scid/gamma mice | Charles River, Wilmington, USA | RRID: IMSR_ARC:NSG |
| **Oligonucleotides** | | |
| shRNA targeting *EWSR1-FLI1* type 1: 5'-GCAGCAGAACCCTTCTTATGA -3' | eurofins Genomics, Ebersberg, Germany | first described in https://doi.org/10.1158/1078-0432.CCR-06-1762 |
| shRNA targeting *EWSR1-FLI1* type 2: 5'-CTTTGGAGCCGCATCACAATA -3' | eurofins Genomics, Ebersberg, Germany | N/A |
| shRNA targeting *EWSR1-ERG*: 5'-GCTACGGGCAGCAGAATTTAC -3' | eurofins Genomics, Ebersberg, Germany | N/A |
| siPOOL against GATA2 | siTOOLs Biotech, Planegg, Germany | N/A |
| siPOOL against E2F2 | siTOOLs Biotech, Planegg, Germany | N/A |
| Primers | This paper | see Table S6 |

(*Continued on next page*)

*Continued*

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Recombinant DNA | | |
| lentiviral pLKO-Tet-On all-in-one vector | Addgene, Watertown, USA[52] | Cat#21915 |
| Software and Algorithms | | |
| Burrows-Wheeler Aligner | Li[84] | https://bio-bwa.sourceforge.net |
| Picard | Broad Institute, Cambridge, MA, USA | https://broadinstitute.github.io/picard/ |
| GATK | Broad Institute, Cambridge, MA, USA | https://gatk.broadinstitute.org/hc/en-us |
| FASTQC | Andrews[85] | https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ |
| Samtools | Li et al.[86] | http://www.htslib.org |
| HOMER | Heinz et al.[74] | http://homer.ucsd.edu |
| HipSTR | Willems et al.[32] | https://hipstr-tool.github.io/HipSTR/ |
| Tandem Repeats Finder | Benson[76] | https://github.com/Benson-Genomics-Lab/TRF |
| CNVnator | Benson[76] | https://github.com/abyzovlab/CNVnator |
| CNVkit | Talevich et al.[87] | https://cnvkit.readthedocs.io/en/stable/ |
| Bcftools | Li[88] | https://samtools.github.io/bcftools/bcftools.html |
| SnpEff | Cingolani et al.[89] | http://pcingola.github.io/SnpEff/ |
| LUMPY | Layer et al.[90] | https://github.com/arq5x/lumpy-sv |
| GRIDSS | Cameron et al.[91] | https://github.com/PapenfussLab/gridss |
| BreakDancer | Chen et al.[26] | https://github.com/genome/breakdancer |
| ChainFinder | Baca et al.[92] | https://software.broadinstitute.org/cancer/cga/chainfinder |
| bowtie2 | Langmead and Salzberg[93] | https://bowtie-bio.sourceforge.net/bowtie2/index.shtml |
| MACS2 | Zhang et al.[94] | https://hbctraining.github.io/Intro-to-ChIPseq/lessons/05_peak_calling_macs.html |
| IGV | Thorvaldsdóttir et al.[95] | https://software.broadinstitute.org/software/igv/ |
| CHIPIN | Polit et al.[96] | https://github.com/BoevaLab/CHIPIN |
| BEDTools | Quinlan and Hall[97] | https://bedtools.readthedocs.io/en/latest/ |
| Transcriptome Analysis Console (4.0) | Thermo Fisher, Waltham, MA, USA | https://www.thermofisher.com/de/en/home/life-science/microarray-analysis/microarray-analysis-instruments-software-services/microarray-analysis-software/affymetrix-transcriptome-analysis-console-software.html |
| fancyimpute (v. 0.5.4., IterativeSVD) | Alex Rubinsteyn and Sergey Feldman: https://github.com/iskandr/fancyimpute | https://pypi.org/project/fancyimpute/ |
| GenomeStudio | Illumina, San Diego, CA, USA | https://emea.illumina.com/techniques/microarrays/array-data-analysis-experimental-design/genomestudio.html |
| R minfi package | Aryeeet al.[98] | https://bioconductor.org/packages/release/bioc/html/minfi.html |
| Gene Set Enrichment Analysis | Subramanian et al.[99] | https://www.gsea-msigdb.org/gsea/index.jsp |
| MolecularNeuropathology.org | Capper et al.[100] | https://www.molecularneuropathology.org/mnp/ |

(*Continued on next page*)

*Continued*

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Weighted Correlation Network Analysis R package | Langfelder and Horvath[101] | https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-9-559 |
| GraphPad PRISM (version 8) | GraphPad Software Inc., CA, USA | https://www.graphpad.com/scientific-software/prism/ |
| Circos plots | Krzywinski et al.[102] | http://circos.ca |
| Other | | |
| Resources website for exploration of the ESCLA, R2 ESCLA | This paper | http://r2platform.com/escla/ |

## RESOURCE AVAILABILITY

### Lead contact

Requests for resources of this article and any additional information should be directed to the lead contact Thomas G. P. Grünewald (t.gruenewald@dkfz-heidelberg.de).

### Materials availability

This study did not generate new unique reagents.

### Data and code availability

The original WGS, ChIP-seq, MethylationEpic BeadChip and DNA microarray data generated during the current study have been deposited at the National Center for Biotechnology Information (NCBI) SRA and GEO, and are publicly available as of the date of publication. The data are accessible through the bioproject and series accession numbers SRA: PRJNA610192, GEO: GSE176339, and GEO: GSE212063. All other data supporting the findings of this study are available within the article, its supplementary data files or via the R2 platform (http://r2platform.com/escla/). Bioinformatic analyses were performed with published and freely available software indicated and cited in the respective STAR Methods section and in the key resources table. This paper does not report original code. Any additional information required to reanalyze the data reported in this work paper is available from the lead contact upon request.

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

### Provenience of cell lines

The EwS cell line A-673 (RRID CVCL_0080) was obtained from American Type Culture Collection (ATCC, Manassas, USA), and MHH-ES-1 (CVCL_1411), RD-ES (CVCL_2169), SK-ES-1 (CVCL_0627), SK-N-MC (CVCL_0530), TC-71 (CVCL_2213) were obtained from the German Collection of Microorganisms and Cell Cultures GmbH (DSMZ, Braunschweig, Germany). The EwS cell lines CHLA-10 (CVCL_6583), CHLA-25 (CVCL_M152), TC-32 (CVCL_7151) and TC-106 (CVCL_F531) were obtained from the Children's Oncology Group (COG),[30,103] and the EwS cell lines EW-1 (CVCL_1208), EW-3 (CVCL_1216), EW-7 (CVCL_1217), EW-22 (CVCL_1214), EW-24 (CVCL_1215), MIC (CVCL_EI96), Rh1 (CVCL_1658), and POE (CVCL_EJ01) were kindly provided by O. Delattre (Paris, France). HEK293T (CVCL_0063) was obtained from DSMZ. Cell identity was regularly controlled with in-house short tandem repeats (STR) profiling and, if applicable, by detection of specific fusion oncogenes by PCR, gel-electrophoresis, and Sanger sequencing. Genetic sex was ascertained from WGS data and is indicated in Figure 1F.

### Culture conditions for cell lines

To avoid any bias due to culture conditions, all cell lines were cultured in RPMI 1640 medium with stable glutamine (Biochrom, Berlin, Germany), supplemented with 10% FCS tested to be dox-free (Sigma-Aldrich, Taufkirchen, Germany), and penicillin and streptomycin (final concentrations 100 units/mL and 100μg/mL, respectively; Merck, Darmstadt, Germany) in tissue culture flasks and plates (TPP, Trasadingen, Switzerland). To improve cell attachment for CHLA-10, CHLA-25, EW-3, EW-24, MIC, and TC-106, culture dishes were coated with 2% gelatine (Sigma-Aldrich) for cell expansion, and collagen type I solution from bovine skin (Sigma-Aldrich) in experimental assays. Cells were incubated at 37°C and 5% $CO_2$ in a fully humidified environment. Cells were subcultured in ratios 1:2 to 1:8 after detachment with trypsin/EDTA (Biochrom), and spinning down the detached cells. Mycoplasma contamination was ruled out regularly by nested PCR with cell supernatant of each experiment.

### *In vivo* expansion of EwS cell lines

Animal experiments were conducted with allowance from the government of Upper Bavaria (ROB-2532.Vet_02-15-184) in the proprietary animal facility of the Institute of Pathology, LMU Munich, and in compliance with all relevant ethical regulations. Sample size was predetermined using power calculations with $\beta = 0.8$ and $\alpha = 0.05$ based on preliminary data and in compliance with the 3R

system (replacement, reduction, refinement). To generate EwS xenograft tumors for immunohistochemistry, $2.5\times10^6$ cells of six representative EwS cell lines with inducible knockdown of the fusion oncogene (A-673, EW-7, POE, SK-N-MC, Rh1, TC-71) were injected in 1:1 mix of PBS (Biochrom) and basement membrane matrix (Thermo Fisher) into the right flank of NOD/Scid/gamma mice (RRID:IMSR_ARC:NSG, Charles River, Wilmington, USA; all female except for the A-673-injected mice) at the age of 12 weeks. Mice were controlled every two days for health and tumor growth. Tumor size was measured with a caliper and tumor volume calculated as $V = a\times b^2/2$, with $a$ being the largest and $b$ the smallest diameter as previously described.[104] Once, the average tumor diameter reached 10 mm, mice were alternately allocated to treatment and control group, thus treated for 96 h with 2 mg/mL dox (belapharm, Vechta, Germany) in sucrose (Sigma-Aldrich) drinking water or only with sucrose drinking water, and then sacrificed by cervical dislocation. Other humane endpoints were determined as follows: Xenograft volume of 1,500 $mm^3$, ulcerated tumors, loss of 20% body weight, constant curved or crouched body posture, bloody diarrhoea or rectal prolapse, abnormal breathing, severe dehydration, visible abdominal distention, obese Body Condition Scores (BCS), apathy, and self-isolation. Xenografts were quickly extracted. Fragments of the tumors were snap frozen for RNA isolation. The remaining tumor was fixed in formalin and, at latest 72 h therafter, dehydrated and embedded in paraffin. For further procedures and analyses influences of sex were not considered, as all replicates and conditions of a specific cell line were expanded in mice of the same sex.

## METHOD DETAILS

### EWSR1-ETS-knockdown cell lines establishment

18 EwS cell lines with inducible knockdown of the fusion oncogene were generated. The expression of the given *EWSR1-ETS* fusion was assessed in every cell line by PCR of cellular cDNA and gel-electrophoresis. The following primers were used:

*EWSR1*-forward: 5'- GCCAAGCTCCAAGTCAATATAGC -3';
*FLI1*-reverse: 5'- GAGGCCAGAATTCATGTTATTGC -3';
*ERG*-reverse: 5'- TTGGGTTTGCTCTTCCGCTC -3'.

Commercial Sanger sequencing (MWG Eurofins genomics, Ebersberg, Germany) of the PCR products was performed to confirm the described transcript sequences at the fusion point of *EWSR1* and *ETS* transcripts. For RNA interference (RNAi)-based knockdown of the fusion oncogene, the fusion-specific short hairpin RNA (shRNA) sequence published by Carrillo et al.[57] was chosen for the 11 selected EwS cell lines with *EWSR1-FLI1* subtype 1 fusion (*EWSR1* exon 7 to *FLI1* exon 6). For four EwS cell lines with *EWSR1-FLI1* subtype 2 fusion (*EWSR1* exon 7 to *FLI1* exon 5), no 21-bases target sequence with at least 8 bases ranging from one fusion partner to the other yielded a high intrinsic score when tested in the Genetic Perturbation Platform (GPP, Broad Institute, Cambridge, USA). This was most likely caused by the last nucleotide of exon 7 of *EWSR1* being an adenosine, as is the last nucleotide of exon 4 of *FLI1*, and the first five nucleotides of exon 5 of *FLI1* being GTTCA as are the first nucleotides of exon 8 of *EWSR1*. Since these six nucleotides do not provide any specificity for RNAi, an shRNA predicted to be effective for *FLI1* exon 9 was chosen for knockdown experiments in these cell lines. Expression of wildtype (not-fused) *FLI1* was excluded for these EWSR1-FLI1 subtype 2 cell lines by qRT-PCR. All three *EWSR1-ERG* positive cell lines, CHLA-25, EW-3, and TC-106 exhibited the same fusion transcript (*EWSR1* exon 7 to *ERG* exon 6). A fusion-specific shRNA target sequence for these *EWSR1-ERG* cell lines was predicted with GPP and specificity was tested with BLAST. Target sequences were as follows:

*EWSR1-FLI1* type 1: 5'- GCAGCAGAACCCTTCTTATGA -3';
*EWSR1-FLI1* type 2: 5'- CTTTGGAGCCGCATCACAATA -3';
*EWSR1-ERG*: 5'- GCTACGGGCAGCAGAATTTAC -3'; All oligonucleotides were purchased from Eurofins Genomics. For generation of cell liens with a dox-inducible knockdown of the respective *EWSR1-ETS* fusion oncogene, the lentiviral pLKO-Tet-On all-in-one vector[52] (Addgene, Watertown, USA) was used. Via a tetracycline responsive element, the expression of the cloned shRNA can be induced by application of dox to the cell culture medium. This vector carries a puromycin resistance cassette, enabling selection of successfully transfected cells. The vector was digested with EcoRI-HF and AgeI-HF (New England Biolabs, NEB, Frankfurt am Main, Germany), precipitated, and the opened plasmid without stuffer was extracted from agarose gel. The linearized vector was ligated with the annealed shRNA targeting the *EWSR1-ETS*-fusions or with non-targeting control sequence using T4 DNA ligase (Thermo Fisher, Waltham, USA). Stellar competent cells (Clontech, TaKaRa, Saint-Germain-en-Laye, France) were transformed with the ligated vector and plated. Bacteria clones were picked, and presence of the vector was controlled with colony PCR (Primers: Tet-pLKO_forward 5'-GGCAGGGATATTCACCATTAT-3', Tet-pLKO_reverse 5'-CTATTCTTTCCCCTGCACTG-3', Eurofin Genomics, Ebersberg, Germany). Corresponding clones were expanded, plasmids were extracted (PureYield Plasmid Midiprep System, Promega, Maddison, USA) and correct sequence of the shRNA was controlled with Sanger sequencing (MWG Eurofins Genomics; sequencing primer Tet-pLKO-Seq forward 5'-GGCAGGGATATTCACCATTATCGTTTCAGA-3', Tet-pLKO-Seq reverse 5'-GACGT GAAGAATGTGCGAGA-3', Eurofin Genomics, Ebersberg, Germany). Lentiviral particles were generated by transfecting HEK293T cells with 10 μg shRNA-carrying plasmid, 10 μg D8.9 and 3 μg pCMV-VSV-G (Addgene) packaging plasmids using Lipofectamine LTX with Plus Reagent (Thermo Fisher). The standard culture medium was replaced after 12 h by culture medium with 30% FCS. Virus-containing supernatant was collected after 48 h and filtered through 0.45 μm membrane. About $1\times10^6$ EwS cells were infected with 1 mL supernatant without polybrene. Upon confluence of the cells, cells were split, and successfully stably transduced cells were selected with 1 μg/mL puromycin (Invivogen, San Diego, USA). This concentration has been tested to be the lowest lethal dose for these EwS cell lines in wildtype conditions. For single cell cloning, virtually 0.8 cells per well of a 96-well-plate were seeded

in 150μL medium, containing 20% conditioned medium of the respective cell lines filtered through 0.45 μm membrane. The cloned shRNAs against the respective *EWSR1-ETS* fusion oncogenes was induced by addition of 1 μg/mL dox (Merck, Darmstadt, Germany) to the cell culture medium. Dox was refreshed every 48 h.

### Transient E2F2 and GATA2 knockdown

To induce transient *E2F2* or *GATA2* knockdown, 1 × $10^5$A-673 Ewing sarcoma cells were seeded on 6-well plates. The cells were immediately transfected with 30 nM siPOOLs targeting *E2F2* or *GATA2*, or with 30 nM of a non-targeting control siRNA pool (all siTOOLs Biotech, Planegg, Germany) using Opti-MEM (Thermo Fisher) and lipofectamine RNAiMax reagent according to the manufacturer's recommendations (Thermo Fisher). 48 h after the first transfection, cells were re-transfected. 96 h after the first transfection, total RNA was isolated and prepared for DNA microarray gene expression analysis as indicated below.

### RNA extraction, RT and qRT-PCR

RNA was extracted from cells with the NucleoSpin RNA kit from Macherey-Nagel (Düren, Germany) following the manufacturer's protocol. RNA (1 μg, quantified with NanoDrop) was reverse transcribed (RT) to cDNA with the High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher) according to the manufacturer's protocol. cDNA was diluted 1:10. For qRT-PCR 6.75 μL cDNA with 7.5 μL SYBR Select Mastermix (Thermo Fisher) and 0.75 μL equimolar mix of forward and reverse primer (final concentration 0.5 μM) were mixed per reaction (final volume 15 μL). Each reaction was run in duplicates. Gene expression levels for the tested gene were normalized to the expression of the housekeeping gene *RPLP0*, which has been previously used for EwS cells[18,19,45] and that does not show significant variation of its expression levels upon knockdown of the respective fusion gene as determined by Affymetrix Clariom D microarrays. qRT-PCR was run in a CFX Connect Real-Time PCR Detection System (Bio-Rad, Munich, Germany) with the following protocol: heat activation at 95°C for 2 min, annealing and elongation at 60°C for 30 s (50 cycles), final denaturation at 95°C for 30 s, followed by stepwise temperature increase from 65°C to 95°C, 0.5°C step every 5 s, for melting curve. Differential expression in the qRT-PCR samples was assessed with the Delta-Delta-Ct method.

Primer sequences (all from Eurofin Genomics, Ebersberg, Germany) were as follows:

*RPLP0* forward: 5'- GAAACTCTGCATTCTCGCTTC -3'; *RPLP0* reverse: 5'- GGTGTAATCCGTCTCCACAG -3';

*EWSR1-FLI1* forward: 5'- GCCAAGCTCCAAGTCAATATAGC -3'; *EWSR1-FLI1* reverse: 5'- GAGGCCAGAATTCATGTTATTGC -3'; *EWSR1-ERG* forward: 5'- TCCAAGTCAATATAGCCAACAGAG -3'; *EWSR1-ERG* reverse: 5'- CTGTGGAAGGAGATGGTTGAG -3'; *EWSR1* (wildtype) forward: 5'- CAGCCAAGCTCCAAGTCAATA -3'; *EWSR1* (wildtype) forward: 5'- TCCAGACTCCTGCCCATAAA -3'; *FLI1* (wildtype) forward: 5'- TGGATGGCAAGGAACTGTG -3'; *FLI1* (wildtype) reverse: 5'- CGGTGTGGGAGGTTGTATTA -3'; *ERG* (wildtype) forward: 5'- CGAACGAGCGCAGAGTTAT -3'; *ERG* (wildtype) forward: 5'- ACGTCTGGAAGGCCATATTC -3'; *E2F2* forward: 5'- AAAAGGAAGCTGGATCTGGAG -3'; *E2F2* reverse: 5'- GCGAAGTGTCATACCGAGTC -3'; *GATA2* forward: 5'- TTCAATCACCTCGAC TCGC -3'; *GATA2* reverse: 5'- GCTGTGCAACAAGTGTGG -3'.

Primers for the reverse fusion oncogene were published before.[25]

### Western blot

Western blot analyses were carried out as previously described.[105] Briefly, about 1×$10^6$ EwS cells were lysed in 100 μL RIPA buffer. Protein content was measured with Bradford assay (Bio-Rad). 20 μg protein were denaturated at 95°C for 5 min, and loaded on western blot gel. Samples were separated in a 10% acrylamide (Roth, Karlsruhe, Germany) gel and blotted onto nitrocellulose membrane (GE Healthcare, Freiburg, Germany) in a wet system. The membrane was blocked with milk (Roth) or BSA (for anti-FLI1) and incubated with the first antibody overnight at 4°C. Incubation with the corresponding horseradish peroxidase coupled second antibody followed the next day for 1 h at room temperature (RT). Bands were detected and quantified after addition of a chemiluminescent reagent (Merck) with LI-COR Odyssey (Homburg, Germany). Signal intensities were automatically quantified by the LI-COR system. The following antibodies were used: monoclonal anti-FLI1 raised in mouse (RRID:AB_1516868, 254M-16, 1:1,000; medac, Wedel, Germany), monoclonal anti-ERG raised in rabbit (EP111, 1:2,000; Cell Marque, Rocklin, CA, USA), anti-GAPDH raised in mouse (RRID:AB_627679, sc-32233, 1:2,000; Santa Cruz, Dallas, USA), polyclonal anti-mouse IgG-HRP raised in goat (RRID:AB_430834, W402B, 1:3,000; Promega) and polyclonal anti-rabbit IgG-HRP raised in goat (RRID:AB_10262463, R1364HRP, 1:5,000; OriGene, Herford, Germany)

### TMA construction and IHC

Analyses of EwS tissue was carried out as previously described.[106] On hematoxylin & eosin (HE) stained slides of the EwS xenograft tumors grown in NSG mice, representative areas with vital tumor tissue were marked. Three tissue cores (each 1 mm in diameter) per tumor were extracted from these areas. Tissue cores were integrated in tissue microarrays (TMA) scaffolds with human tonsils as control tissue. For subsequent immunohistochemical (IHC) stains, 4 μm sections were cut. Antigen retrieval was performed with microwave treatment using the antigen retrieval AR-10 solution (HK057-5K, DCS Innovative, Hamburg, Germany) for FLI1, the antigen retrieval ProTaqs I and V Antigen-Enhancer (Quartett, Berlin, Germany) for p-MYBL2 and PAX7, and the Target Retrieval Solution (S1699, Agilent Technologies, Waldbronn, Germany) for SOX6. For blockage of endogenous peroxidase, slides were incubated for 20 min in 7.5% aqueous $H_2O_2$ solution and blocking serum. Then, slides were incubated for 60 min with the primary anti-FLI1 (RRID:AB_1516868, 254M, 1:120; Cell Marque, Rocklin, USA), anti-p-MYBL2 (RRID:AB_1309969, ab76009, 1:100; Abcam,

Cambridge, UK), anti-PAX7 (RRID: AB_2299243, PAX7-c, 1:180; DSHB, Iowa City, IA, USA) or anti-SOX6 (RRID:AB_1080063, HPA003908, 1:1,600; Atlas Antibodies, Stockholm, Sweden) and afterwards with a secondary peroxidase-conjugated anti-rabbit/-mouse IgG antibody (RRID:AB_2336529, MP-7401, ImmPress Reagent Kit for FLI1, p-MYBL2 and SOX6; RRID:AB_2336826, PK6200, Vectastain Elite ABC HRP Kit for PAX7; Vector Laboratories, Burlingame, USA). Target detection was performed using DAB + chromogen (Agilent Technologies, Santa Clara, USA). Slides were counterstained with hematoxylin Gill's Formula (H-3401; Vector Laboratories). Immunoreactivity was semi-quantified with a slightly modified immunoreactivity score by Remmele and Stegner.[107] Staining intensity was scored as no, low, moderate, and strong with the values 0, 1, 2, 3, the area of stained cells was scored with 0–5 indicating quintiles of cell area positive for immunoreactivity. The product of the intensity and the area score resembled the final immunoreactivity score (IRS).

### Whole-genome sequencing (WGS)

DNA was extracted from wildtype EwS cell lines using the NucleoSpin Tissue kit (Macherey Nagel) following manufacturer's protocol and eluted in $H_2O$. For sequencing, 50 μL of 50 ng/μL DNA were used. After initial DNA quality assessment on a bioanalyzer (DNA Integrity Number at least 7.0), DNA was sequenced (150 bp, paired-end, PCR-free protocol) on Illumina HiSeq Xten or NovaSeq6000 (Illumina, San Diego, USA) at the Genomics and Proteomics Core Facility of the German Cancer Research Center (DKFZ, Heidelberg, Germany) or Life&Brain (Bonn, Germany). Raw sequencing data were aligned to the human reference genome (hg19) with Burrows-Wheeler Aligner (bwa mem).[84] PhiX contamination was excluded, Illumina adapters and duplicates were marked with picard (Broad Institute). Base quality scores were recalibrated with GATK (Broad Institute). Quality of the final alignments was controlled with FASTQC.[85] Coverage was assessed with samtools depth[86] and displayed as average of 90 kb bins. *De novo* motif finding was performed with HOMER.[74]

### Genotyping of GGAA mSats

To genotype and phase GGAA mSats in EwS cell lines, the haplotype inference and phasing for short-tandem repeats (HipSTR) tool was employed.[32] A library of all potential GGAA mSats was generated with Tandem Repeats Finder[76] running on the human reference genome (hg19), calling for those repeats with four nucleotide motifs, either high guanine-adenine or cytosine-thymine content and at least 4 sequential GGAA or TTCC motifs. The final library contained 8,311 loci. The library, the processed and aligned Illumina WGS data, and the reference genome were used as input for HipSTR. HipSTR mines all STR alleles for each locus, aligns reads accounting for artifacts due to the diversity and structure of STRs, and phases the STRs. The resulting variant calling file (VCF) was filtered for minimal call quality (0.9), maximum number of stutter artifacts and indels in STR flanking regions (0.15), minimal call rate (0.3), minimal call depth per locus (10) and minimal supporting reads per allele (3). Readouts were displayed for up to 18 consecutive GGAA-repeats as up to this repeat number more than 100 genotypes were called.

### Copy number variation and ploidy analysis

Copy numbers of genomic regions were inferred from the WGS data with CNVnator[108] by extraction of read mapping information, building a read depth histogram, calculating statistics and copy number variation (CNV) calling with a bin size of 1,000 nt. CNVkit was employed for CNV analyses for single genes with 300 bp bind and exclusion of not accessible chromosomal regions.[87] Ploidy was estimated with an algorithm from the ploidyNGS tool[29] comparing the most and second most frequent allele of heterozygous loci for read numbers, expecting a nearly 1:1 distribution in case of euploidy. In case of higher numbers of chromosome sets (n), different ratios would have been expected (2:1 in 3n, 3:1 in 4n cells, etc.). Due to high RAM consumption, the tool was not directly applied on the WGS data, but the method was adopted on read counts for InDels generated by VCMM, which also indicated the number of reads per genotype. The rates of reads per allele at the first 100,000 heterozygous loci were plotted in a stacked histogram for visualization of the ploidy status.

### Variant calling

Single nucleotide variants (SNVs) were called with bcftools[88] and GATK Mutect2 for somatic mutations, and annotated with ANNOVAR and SnpEff.[109,89] Structural variants were assessed with LUMPY, giving split and discordant reads as input, and GRIDSS.[90,91,110] Transchromosomal fusions were analyzed with BreakDancer.[26] Genomic rearrangement loops were extrapolated using the CNVkit and GRIDSS output in ChainFinder.[92]

### ChIP-seq

ChIP-seq analyses of EwS cell lines were carried out following established guidelines.[111] To identify EWSR1-ETS-bound genomic regions, ChIP was performed. For ChIP, at least $1\times10^7$ EwS cells were cultured. Only cells from not confluent cultures without consumed culture medium were used. Cells were fixed with 1% methanol-free formalin for 10 min at RT. Formalin was quenched with glycine (final 200 mM) for 5 min. Cells were lysed in two steps with the Diagenode iDeal ChIP-seq kit for transcription factors (Diagenode, Seraing, Belgium). DNA was sheared with Bioruptor Plus sonication device (Diagenode). Cell debris was spun down, and sheared DNA was added to antibodies coupled to washed magnetic beads (anti-FLI1 2 μg, RRID:AB_301825, ab15289, Abcam; anti-ERG 2 μg, RRID:AB_2630401, ab92513, Abcam; anti-H3K4me3 1.4 μg, RRID:AB_2616052, C15410003, Diagenode; anti-H3K27me3 2.9 μg, RRID:AB_2814977, C15410069, Diagenode; anti-H3K27ac 1 μg, RRID:AB_2118291, ab4729, Abcam). As

wildtype *FLI1* or *ERG* are not expressed in *EWSR1-FLI1*- or *EWSR1-ERG*-positive EwS cell lines, respectively, anti-FLI1 and anti-ERG antibodies can be assumed to specifically target the fusion oncoprotein. After incubation overnight at 4°C, beads were washed 4× with Diagenode washing buffers, bound DNA was eluted from the beads and purified. DNA was quantified with Qubit. As control for successful immunoprecipitation and suitability of the selected antibodies ChIP-PCR was performed with primers for genomic regions known to be bound by the respective marker. Then, Illumina compatible libraries were prepared using the Illumina TruSeq ChIP library preparation kit according to the manufacturer's protocol. Briefly, 1–10 ng of input or IP DNA were subjected to subsequent steps of end-repair, dA-tailing and ligation of TruSeq indexed Illumina adapters. After a final PCR amplification step, barcoded libraries were equimolarly pooled and quantified using a qPCR method (KAPA library quantification kit, Roche) before sequencing on the Illumina HiSeq2500 instrument. Each pool was loaded on 1 rapid flow cell and sequenced using a single-end mode (100 or 150 bases). Raw sequencing data were aligned to human reference genome hg19 with bowtie2.[93] Alignments were purged from duplicates with samtools. Peaks were called with MACS2,[94] selecting narrow peak for transcription factors and broad peak for histone marks. Super-enhancers were called with ROSE.[66]

For data display in IGV,[95] MACS2 output was converted into bigwig files. These files were normalized between cell lines based on genes with steady expression levels across cell lines using CHIPIN (https://github.com/BoevaLab/CHIPIN).[96] Merged peaks of ETS-ChIP across cell lines were calculated with HOMER and 200 nt width around peak center. For merged H3K27ac peaks, peak files were used as given by MACS2. Overlaps and genomic distances were analyzed with BEDTools.[97]

The core subset of EWSR1-ETS binding sites was defined as those sites shared in more than 80% of cell lines (i.e., at least in any 15 of the 18 cell lines). The cut-off of 80% was chosen empirically based on published data on EWSR1-ETS bound GGAA mSats with eQTL analyses[7,9,18,19,39,58] in overlay with our ChIP results from the ESCLA. Since these sites are likely *bona fide* EWSR1-ETS binding sites, we reasoned that they should be identified in the majority of EwS cell lines even though technical limitations may hinder to detect EWSR1-ETS binding of these sites in all ESCLA cell lines. To account on the one hand for technical limitations and avoid false negatives by choosing a too stringent cut-off (e.g. 18/18 cell lines), and to avoid on the other hand too many false positive core binding sites (e.g. by choosing 9/18 as a cut-off), we selected the cut-off of at least any 15 of the 18 cell ESLCA cell lines. Using this cut-off we re-identified the vast majority (90%) of previously identified direct EWSR1-ETS-target GGAA mSats lending support that this approach may constitute a reasonable core subset of EWSR1-ETS binding sites.

1,785 EWSR1-FLI1 binding sites identified in both SK-N-MC and A-673 by Riggi et al. (defined as core set of 1,785 peaks by the original authors)[31] has been employed as comparison dataset for EWSR1-ETS binding sites identified in this manuscript.

To maximize the statistical power to identify principles of EWSR1-ETS-mediated gene regulation via GGAA mSats and genomic features promoting EWSR1-ETS-binding to GGAA mSats, we compared those 251 GGAA mSats being bound by EWSR1-ETS in every cell line (likely optimal EWSR1-ETS binding sites) with those 4,934 GGAA mSats that were never found to be bound by EWSR1-ETS in any of the 18 cell lines.

### DNA microarray gene expression analyses

For identification of EWSR1-ETS-regulated genes, DNA microarray gene expression analyses were performed. To meet our aim of providing a widely applicable dataset that can be processed by established and user-friendly tools and as the high number of samples ($n$ = 108) required a pragmatic and cost-efficient approach, we have chosen the Affymetrix human Clariom D array, which contains >6 million individual probes, captures >285,000 transcripts, has an excellent dynamic range and can assess global gene expression and splicing simultaneously at much lower costs per sample as compared to deep RNA-sequencing approaches. Accordingly, from all 18 EwS cell lines with inducible knockdown of the respective fusion oncogene, cells were seeded in three technical replicates in standard medium or standard medium supplemented with dox (1 μg/mL). Medium was renewed (including dox) after 48 h. After 96 h, cells were lysed, and RNA was extracted with the NucleoSpin RNA kit (Macherey Nagel). Knockdown of the respective fusion oncogene was evaluated by qRT-PCR. Quality of the RNA (1 μg) was controlled (RNA Integrity Number $\geq$9.0 in all samples) before reverse transcription, fragmentation, and hybridization with the Affymetrix Clariom D microarray (Thermo Fisher) at IMGM (Planegg, Germany). The resulting raw data (CEL files) were normalized and summarized with the SST-RMA analysis algorithm[112] in the Transcriptome Analysis Console (version 4.0, Thermo Fisher) and manufacturer's array description file (version 2). Gene expression values were log2-transformed. Readouts not corresponding to actual gene transcripts were filtered out.

To minimize potential biases conferred by variability in the achieved EWSR1-ETS knockdown, differentially expressed genes (DEGs) were defined like super-enhancers by the ROSE algorithm instead of by application of arbitrary FCs and p value cut-offs. Specifically, genes were plotted ranked by their mean expression FC in wildtype versus knockdown condition. Linear functions connecting the lowest FC with the value 0, and connecting the highest fold change with the value 0 were calculated. The y-intercept for the linear functions were adapted until the function for the positive and negative FCs had only one intersection point with the plotted positive and negative FCs, respectively. All genes from the intersection point to the extremes were defined as DEGs (Figure S1C). Using these cell line specific cut-offs for defining DEGs revealed a much better consistency in EWSR1-ETS-regulated genes across cell lines as compared to a fixed cut-off (Figure S1D). For example, while only 163 DEGs are shared between those cell lines with worst and best EWSR1-ETS knockdown efficiency (TC-32 and MHH-ES-1, respectively; total number of DEGs: 355 and 2,135) when applying a strict FC-cut-off > 1 and p-value < 0.01, the overlap was around 2.6-fold greater by applying our cell line specific cut-offs as defined by the ROSE algorithm (here 417 overlapping genes of 1,109 and 2,026 total DEGs for TC-32 and MHH-ES-1, respectively).

Those genes were depicted as heterogeneously regulated, which were in at least 6 cell lines strongly regulated (top 33% of regulated genes) and in at least another 6 cell lines weakly (lower 33% of regulated genes) or not regulated. Overlaps of regulated genes were calculated for each cell line versus all others. Only the top 924 regulated genes (min. number of genes defined as regulated in a single cell lines) were considered. To increase overlap rates, the top 33% of the 924 regulated genes per cell line were compared to the top 924 regulated genes of each other cell line. Nonlinear dimensionality reduction was used to examine potential EwS cell line clusters dependent on the respective fusion oncogene. To this end, gene expression values on the transcriptome level in EWSR1-ETS-high condition were analyzed with the Rtsne package (version 0.15; perplexity = 5; max_iter = 500).

To assess potential impact of EWSR1-ETS on alternative splicing, for each probe selection region (PSR, exon level data, filtered for only expressed exons) of each gene the expression fold change between EWSR1-ETS-low and -high condition was calculated ($FC_{PSR}$) in each cell line individually. The fold change for gene level was calculated accordingly ($FC_{gene}$). The difference between the fold change of each PSR and the fold change of the corresponding gene was calculated ($FC_{Delta} = FC_{PSR}-FC_{gene}$). In the absence of alternative splicing mediated by EWSR1-ETS, the dynamics on the exon level should correspond to the dynamics on the gene level. Thus, $FC_{Delta} = 0$ was the null hypothesis. For each PSR, a single sample t test was performed across all cell lines to test for the deviation of the $FC_{Delta}$ from 0. The occurrence of alternative splicing was evaluated by comparison of expected deviations from the null hypothesis (i.e. 5% of investigated PSRs) versus observed.

## Protein quantification

For identification of EWSR1-ETS-regulated proteins, mass spectrometry on protein lysates of EwS cell lines was performed. From all 18 EwS cell lines with inducible knockdown of the fusion oncogene, cells were seeded in three technical replicates in standard medium or standard medium supplemented with dox (1 μg/mL). Medium was changed to fresh medium after 48 h (including dox). After 96 h, the cell surface was washed with FCS-free medium, before incubation in FCS-free medium for 20 min and a second wash step to clean the cells from serum proteins. Cells were lysed in Nonidet-P40 buffer (1% Nonidet P40 (Thermo Fisher) and complete protease inhibitor (Sigma-Aldrich)). Lysates were collected with cells scrapers into protein low-binding tubes and sonicated with 60% amplitude, 6 × 30 s. Protein content of the lysates was quantified with Bradford assay (Bio-Rad) and 10 μg proteins were proteolysed with trypsin and peptides analyzed by quantitative LC-MSMS on a Q Exactive HF (Thermo Fisher) by a data-independent acquisition approach as described earlier.[113] Expression of 3,248 proteins were quantified consistently across all 18 cell lines. The expression levels of additional 1,336 proteins were imputed from patchy data with a machine learning based algorithm implemented in fancyimpute (v. 0.5.4, IterativeSVD option). The Pearson's correlation coefficients in the smaller dataset versus the imputed one were roughly the same with 0.56 versus 0.58, respectively, both highly significant. Proteins regulated upon EWSR1-ETS knockdown were defined as genes on the transcriptome level. Overlaps of regulated proteins were calculated for each cell line versus all others. Only the top 216 regulated proteins (minimal number in a single cell lines) were considered. To increase overlap rates, the top 33% of the 216 regulated proteins per cell line were compared to the top 216 regulated proteins of each other cell line. Clustering of EwS cell lines, dependent on the respective fusion oncogene, was investigated using t-SNE analysis (see DNA microarray expression gene analyses).

## Methylation analysis

For analysis of EWSR1-ETS-dependent CpG island methylation, genomic DNA of all 18 EwS cell lines with and without knockdown of the respective fusion oncogene was genotyped on Illumina Infinium MethylationEPIC BeadChip arrays. Therefore, all EwS cell lines were seeded in three technical replicates in standard medium or standard medium supplemented with dox (1 μg/mL). Medium was changed to fresh medium after 48 h (including dox). After 96 h, samples were lysed and DNA was extracted with the NucleoSpin Tissue kit (Macherey Nagel). Genomic DNA was genotyped with the Illumina Infinium Methylation EPIC BeadChip arrays at the Molecular Epidemiology unit of the German Research Center for Environmental Health (Helmholtz Center, Munich, Germany). Readout and analysis of the EPIC arrays was performed with GenomeStudio (Illumina) and the R minfi package and bumphunter algorithm.[98] For t-SNE analysis, the same approach and tools as for transcriptome and proteome data were applied. Methylation profile classification was run on the MolecularNeuropathology.org website.[61,100]

## GSEA, WGCNA, gene ontology

To identify gene sets that are enriched among EWSR1-ETS-regulated genes, all genes from DNA microarray expression analysis were ranked by their expression FC in EWSR1-ETS-high versus -low condition. Pre-ranked gene set enrichment analysis (GSEA) was performed with 1,000 permutations[99] on previously described gene sets obtained from the Molecular Signature Database hosted at the Broad Institute (MSigDB, c2.all.v6.2). For network analysis, the weighted correlation network analysis R package (WGCNA R)[101] was employed. A matrix of functionally annotated gene sets versus genes (indicating presence of the gene in the respective set) was built and the Jaccard's distance for all possible pairs was computed generating a symmetric GSEA adjacent matrix. The dynamicTreeCut algorithm was employed to identify GSEA term clusters. Top results were selected for visualization (min. absolute NES = 2.5). The cluster label corresponds to the highest scoring node of each cluster. For network visualization Cytoscape (v 3.8.2) was used.[114] Selected gene lists were probed for enriched gene ontology terms using PANTHER.[115]

### Survival analysis

For identification of genes associated with overall survival of EwS patients, an established dataset[116] composed of gene expression microarray data of 196 primary EwS tumors with clinical annotation available at GEO (accession codes: Affymetrix HG-U133plus2: GSE12102, GSE17618, GSE34620; Affymetrix HuEX-1.0st: GSE63157) or obtained from J. Alonso (Amersham/ GE Healthcare CodeLink microarrays, unpublished data) was used. Signal raw data were normalized and summarized with Robust Multi-array Average (RMA)[112] and custom brainarray chip description files (CDF, v20), yielding one optimized probe-set per gene.[117] Batch effects between microarray types were removed with ComBat.[118] Only samples with tumor purity of at least 60%, calculated with the ESTIMATE algorithm,[119] were further analyzed (TCGA standard). Survival association of the genes represented on all microarrays was calculated with the Mantel-Haenszel test using an in-house tool (GenEx) and GraphPad PRISM (version 8; GraphPad Software Inc., CA, USA). *P* values were adjusted for multiple testing using the Bonferroni method.

## QUANTIFICATION AND STATISTICAL ANALYSIS

Association of gene expression levels with patients' overall survival was assessed using the Kaplan-Meier method and the Mantel-Haenszel test. For comparison of two groups with normal data distribution (as assessed by Kolmogorow-Smirnow test), the two-sided independent Student's *t*-test was used. For statistical comparison of two groups in which normal distribution could not be assumed, the two-sided Wilcoxon rank-sum test was applied. For assessment of statistical significance assessment between two groups with two discrete categories, the Fisher's exact test or chi-square test was used. Most relevant statistical details are mentioned in the main text, additional information, including replicate numbers, are given in the figures and legends. *P* values under 0.05 (if applicable, Bonferroni corrected for multiple testing) were considered as significant.

Readouts of 'omics' analyses for specific genomic regions were visualized in the Integrative Genomics Viewer (IGV; version 2.6.2). Data encompassing the entire human genome were displayed in Circos plots.[102] Heatmaps were generated with GEN-E (Broad Institute) and Venn diagrams in BioVenn. t-stochastic neighbor embedding (t-SNE) was performed in R. Other plots were generated in GraphPad PRISM (version 8; GraphPad Software Inc) and R.

## ADDITIONAL RESOURCES

To facilitate data mining and exploration of the data by biomedical researchers without prior bioinformatics training, most of the data presented in this manuscript have been implemented in interactive analysis tools in a dedicated data scope in the R2 genomics analysis and visualization platform which can be accessed directly via http://r2platform.com/escla/. These implementations entail the mRNA, methylation, copy number, structural variant, protein, mutation and ChIP data.