Targeted long-read sequencing of the Ewing sarcoma 6p25.1 susceptibility locus identifies germline-somatic interactions with EWSR1-FLI1 binding

Authors

Olivia W. Lee, Calvin Rodrigues, Shu-Hong Lin, ..., Thomas G.P. Grünewald, Olivier Delattre, Mitchell J. Machiela

Correspondence

olivier.delattre@curie.fr (O.D.), mitchell.machiela@nih.gov (M.J.M.)

Lee et al. characterized germline variation in the 6p25.1 EwS susceptibility region. Longer GGAA microsatellite alleles at 6p25.1 were associated with EwS risk and the microsatellite showed chromatin features of an *EWSR1-FLI1* enhancer. *EWSR1-FLI1* binding altered expression of *RREB1*, promoting proliferation.

Lee et al., 2023, The American Journal of Human Genetics *110*, 427–441
March 2, 2023
https://doi.org/10.1016/j.ajhg.2023.01.017

Targeted long-read sequencing of the Ewing sarcoma 6p25.1 susceptibility locus identifies germline-somatic interactions with EWSR1-FLI1 binding

Olivia W. Lee,[1,2,32] Calvin Rodrigues,[3,4,32] Shu-Hong Lin,[1] Wen Luo,[1,5] Kristine Jones,[1,5] Derek W. Brown,[1] Weiyin Zhou,[1,5] Eric Karlins,[1,5] Sairah M. Khan,[1] Sylvain Baulande,[6] Virginie Raynal,[6] Didier Surdez,[3,4,7] Stephanie Reynaud,[4,8] Rebeca Alba Rubio,[9] Sakina Zaidi,[3,4] Sandrine Grossetête,[3,4] Stelly Ballet,[4,8] Eve Lapouble,[4,8] Valérie Laurence,[4] Gaelle Pierron,[4,8] Nathalie Gaspar,[10] Nadège Corradini,[11] Perrine Marec-Bérard,[11] Nathaniel Rothman,[1] Casey L. Dagnall,[1,5] Laurie Burdett,[1,5] Michelle Manning,[1,5] Kathleen Wyatt,[1,5] Meredith Yeager,[1,5] Raj Chari,[1,12] Wendy M. Leisenring,[13] Andreas E. Kulozik,[14] Jennifer Kriebel,[15,16,17] Thomas Meitinger,[18,19] Konstantin Strauch,[20,21] Thomas Kirchner,[22,23,24] Uta Dirksen,[25] Lisa Mirabello,[1] Margaret A. Tucker,[1] Franck Tirode,[3,4] Gregory T. Armstrong,[26] Smita Bhatia,[27] Leslie L. Robison,[26] Yutaka Yasui,[26] Laura Romero-Pérez,[9,22,23] Wolfgang Hartmann,[28] Markus Metzler,[29] W. Ryan Diver,[30] Adriana Lori,[30] Neal D. Freedman,[1] Robert N. Hoover,[1] Lindsay M. Morton,[1] Stephen J. Chanock,[1] Thomas G.P. Grünewald,[9,22,23,31] Olivier Delattre,[3,4,*] and Mitchell J. Machiela[1,33,*]

Summary

Ewing sarcoma (EwS) is a rare bone and soft tissue malignancy driven by chromosomal translocations encoding chimeric transcription factors, such as EWSR1-FLI1, that bind GGAA motifs forming novel enhancers that alter nearby expression. We propose that germline microsatellite variation at the 6p25.1 EwS susceptibility locus could impact downstream gene expression and EwS biology. We performed targeted long-read sequencing of EwS blood DNA to characterize variation and genomic features important for EWSR1-FLI1 binding. We identified 50 microsatellite alleles at 6p25.1 and observed that EwS-affected individuals had longer alleles (>135 bp) with more GGAA repeats. The 6p25.1 GGAA microsatellite showed chromatin features of an EWSR1-FLI1 enhancer and regulated expression of *RREB1*, a transcription factor associated with RAS/MAPK signaling. *RREB1* knockdown reduced proliferation and clonogenic potential and reduced expression of cell cycle and DNA replication genes. Our integrative analysis at 6p25.1 details increased binding of longer GGAA microsatellite alleles with acquired EWSR-FLI1 to promote Ewing sarcomagenesis by RREB1-mediated proliferation.

Introduction

Ewing sarcoma (EwS [MIM: 612219]) is a rare bone and soft tissue tumor occurring primarily in adolescents and young adults.[1,2] Despite its rarity (~3 affected individuals per 1,000,000 people in the United States), EwS is the second most diagnosed bone malignancy in children and adolescents with a higher frequency

[1]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD 20892, USA; [2]Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA; [3]Inserm U830, PSL Université, Research Center, Institut Curie, 75005 Paris, France; [4]SIREDO Oncology Centre, Institut Curie, 75005 Paris, France; [5]Cancer Genomics Research Laboratory, Frederick National Laboratory for Cancer Research, Leidos Biomedical Research Inc, Frederick, MD 21701, USA; [6]ICGex Next-Generation Sequencing Platform, PSL Université, Research Center, Institut Curie, 75005 Paris, France; [7]Balgrist University Hospital, Faculty of Medicine, University of Zurich (UZH), Zurich, Switzerland; [8]Unité de Génétique Somatique, Department of Genetics, Institut Curie Hospital, 75005 Paris, France; [9]Max-Eder Research Group for Pediatric Sarcoma Biology, Institute of Pathology, Faculty of Medicine, LMU, 80337 Munich, Germany; [10]Department of Oncology for Child and Adolescent, Institut Gustave Roussy, 94800 Villejuif, France; [11]Institute for Paediatric Haematology and Oncology, Leon Bérard Cancer Centre, University of Lyon, 69008 Lyon, France; [12]Genome Modification Core Laboratory, Frederick National Laboratory for Cancer Research, Frederick, MD 21701, USA; [13]Cancer Prevention and Clinical Statistics Programs, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA; [14]University Children's Hospital of Heidelberg, 69120 Heidelberg, Germany; [15]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, 85764 Neuherberg, Germany; [16]Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, 85764 Neuherberg, Germany; [17]German Center for Diabetes Research (DZD), 85764 München-Neuherberg, Germany; [18]Institute of Human Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, 85764 Neuherberg, Germany; [19]Institute of Human Genetics, Technische Universität München, 80333 Munich, Germany; [20]Institute of Genetic Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, 85764 Neuherberg, Germany; [21]Chair of Genetic Epidemiology, IBE, Faculty of Medicine, LMU, 80539 Munich, Germany; [22]Division of Translational Pediatric Sarcoma Research, German Cancer Research Center (DKFZ), German Cancer Consortium (DKTK), 69120 Heidelberg, Germany; [23]Hopp-Children's Cancer Center (KiTZ), Heidelberg, Germany; [24]Institute of Pathology, Faculty of Medicine, LMU, 80337 Munich, Germany; [25]University Children's Hospital of Essen, 45147 Essen, Germany; [26]Department of Epidemiology and Cancer Control, St. Jude Children's Research Hospital, Memphis, TN 38105, USA; [27]Institute for Cancer Outcomes and Survivorship, University of Alabama, Birmingham, AL 35294, USA; [28]Gerhard- Domagk Institute of Pathology, University Hospital of Münster, 48149 Münster, Germany; [29]University Children's Hospital of Erlangen, 91054 Erlangen, Germany; [30]Department of Population Science, American Cancer Society, Atlanta, GA, USA; [31]Institute of Pathology, Heidelberg University Hospital, Heidelberg, Germany
[32]These authors contributed equally
[33]Lead contact
*Correspondence: olivier.delattre@curie.fr (O.D.), mitchell.machiela@nih.gov (M.J.M.)
https://doi.org/10.1016/j.ajhg.2023.01.017.

of affected individuals with European ancestry compared to African and Asian ancestry.[3–5] This ancestral difference in incidence suggests a substantial germline genetic component to EwS risk; however, knowledge of the underlying genetic etiology of EwS is incomplete. As metastatic EwS survival rates have not appreciably improved over the past 30 years,[1] discoveries in understanding the genetic susceptibility to EwS could provide valuable insights for disease screening and improved management.

EwS arises from a gene fusion between a member of the FUS (MIM: 137070), EWSR1 (MIM: 133450), and TAF15 (MIM: 601754) (FET) family and an erythroblast transformation-specific (ETS) transcription factor. The most frequent fusion is between Ewing sarcoma breakpoint region 1 (*EWSR1*) at 22q12 and Friend leukemia integration 1 (*FLI1*) (MIM: 193067) at 11q24, observed in >85% of EwS-affected individuals.[6–8] The EWSR1-FLI1 fusion results in the expression of a chimeric oncoprotein that binds two types of GGAA motifs, one centered on a single GGAA that corresponds to the classical binding motif of the ETS family, the other composed of tandem repeats of the GGAA sequence forming microsatellites. Binding of EWSR1-FLI1 to the latter motif creates *de novo* enhancers, alters local transcription, and promotes the development of EwS.[9–12] Prior studies indicate that the EWSR1-FLI1 oncoprotein binds to a minimum of four consecutive GGAA motifs and that EWSR1-FLI1-dependent enhancer activity increases exponentially with more than 12 consecutive GGAA motifs until an optimal "sweet-spot" of GGAA-microsatellite length (18–26 GGAA repeats).[13–16]

To date, the known underlying genetic architecture of EwS susceptibility includes 6 independent genetic susceptibility loci (1p36.22, 6p25.1, 10q21.3, 15q15.1, 20p11.22, and 20p11.23) that have been identified in genome-wide association studies (GWASs).[17,18] Effect estimates for these loci have odds ratios in excess of 1.7, which is higher than most GWAS susceptibility loci for adult cancers. EwS GWAS susceptibility loci are tagged by common variants (MAF > 5%) despite its low incidence in the general population. Moreover, a subsequent investigation also identified low-frequency variants in these regions independently associated with EwS risk,[17] underscoring the complex nature of susceptibility to EwS. It is notable that EwS is rarely identified in pedigrees with highly penetrant mutations in cancer-predisposition genes.[18,19] An additional distinctive feature of EwS susceptibility is an enrichment of GGAA microsatellites near lead GWAS signals,[20] suggesting potential interactions between inherited germline variation and somatically acquired *EWSR1-FLI1* fusions. As such, EwS GWAS susceptibility loci might be tagging variation in or around GGAA microsatellites that promote EWSR1-FLI1 binding and alter downstream dysregulation of target gene expression.

Our prior study of the 10q21 EwS GWAS susceptibility locus examined the relationship between a GWAS-identified EwS susceptibility locus and EWSR1-FLI1 binding.[14] Regional targeted sequencing of 10q21 identified a single-nucleotide polymorphism, rs79965208, located within a GGAA microsatellite that increased the length of consecutive GGAA repeats, which resulted in both increased EWSR1-FLI1 binding and enhanced downstream expression of the nearby *EGR2* (MIM: 129010) gene, which has an impact on EwS cell growth and proliferation.

The current investigation seeks to characterize the previously identified EwS germline susceptibility locus at 6p25.1 tagged by rs7742053. This variant, rs7742053, is telomeric to the Ras-responsive element-binding protein 1 (*RREB1* [MIM: 602209]) gene, and expression analysis has demonstrated elevated *RREB1* expression in EwS, as well as an expression quantitative trait locus (eQTL) association with the GWAS risk variant.[20] Knockdown of EWSR1-FLI1 in A-673 EwS cell lines also showed strong downregulation of *RREB1* expression.[20] These prior findings suggest that EWSR1-FLI1 dysregulation of *RREB1* could be important for EwS etiology at the 6p25.1 locus. The 6p25.1 EwS susceptibility region, spanning 90 kb, has 1 microsatellite with more than 4 consecutive GGAA repeats, which is needed for EWSR1-FLI1 binding,[9] and chromatin immunoprecipitation sequencing (ChIP-seq) analyses confirmed that this GGAA microsatellite within the 6p25.1 EwS susceptibility region demonstrated evidence for FLI1 binding.[20] We hypothesize potential germline-somatic interactions at the 6p25.1 EwS locus in which germline variation at a nearby GGAA microsatellite could impact EWSR1-FLI1 binding, resulting in altered gene expression of *RREB1*. We employed targeted Pacific Biosciences (PacBio) single-molecule real-time (SMRT) sequencing[21] of EwS-affected and control blood DNA samples to construct haplotypes of the microsatellite region in the 6p25.1 EwS susceptibility locus as well as characterize associations of alleles with EwS risk. We further examined the potential of the 6p25.1 GGAA microsatellite as a EWSR1-FLI1 binding enhancer and regulator of *RREB1* expression and investigated the impact of *RREB1* expression on EwS proliferation.

Material and methods

Study populations (participants)

A total of 580 blood-derived DNA samples (348 EwS-affected individuals and 232 control subjects) were included in this study. Affected individuals were collected from 4 contributing studies including the Childhood Cancer Survivorship Study (n = 50), the Institut Curie (n = 50), LMU Munich (n = 202), and the Bone Disease Injury Prevention Study (n = 28). Control subjects originated from the American Cancer Society Cancer Prevention Study II (n = 232). Informed consent was obtained from all participants or their legal representatives previously through

contributing studies. The study was approved by Institutional Review Boards at each participating study center.

We applied stringent quality control steps to ensure a high-quality dataset (Figure S1). A random 15% of samples for each plate were run on Agilent TapeStation prior to amplification to ensure the overall high quality of each included plate of DNA. Likewise, random samples were evaluated by Agilent TapeStation after each round of PCR amplification to confirm high-quality PCR products. We quantified PCR products using PicoGreen and normalized the samples prior to pooling. We again utilized the TapeStation after pooling the barcoded amplification products to determine the pooled library size and molarity prior to sequencing.

We performed targeted Single Molecule Real Time (SMRT) sequencing on a total of 530 samples passing QC steps (328 EwS-affected individuals and 202 cancer-free control subjects) of the chr6: 6,835,220–6,839,996 region. Additional quality-control filtering of circular consensus sequencing (CCS) reads was performed to ensure an accuracy greater than 99% and a coverage depth minimum of 15 CCS reads after sequencing and mapping. Advanced microsatellite analyses were performed on a total of 385 samples (271 EwS-affected individuals and 114 control subjects) (Figure S1).

Laboratory methods

10 ng of DNA was prepared per sample to perform targeted PacBio Single Molecule Real-Time (SMRT) sequencing. The SMRT sequencing process protocols are described in the PacBio sequencing manual.[22] Briefly, the protocol contains a 2-step PCR. In the first step of PCR, template-specific primers amplify the region of chr6: 6,835,220–6,839,996 and incorporate a universal tag sequence. The second PCR utilizes the universal tag, which serves as a priming site for incorporating unique barcode sequences onto each sample for multiplexing. After the second round of PCR, the products were purified and quantified. Normalization was performed on the PCR products prior to pooling to ensure even representation of each sample in the pool. Hairpin adapters were ligated to the ends of the pooled amplicons during the SMRT bell library preparation. Each library pool (of up to 96 samples) had primer annealing and polymerase binding performed according to the protocol and sequenced on 1 Sequel SMRT Cell. After sequencing, CCS read generation was performed with the criteria of minimum 3 passes and accuracy of 99%.

Variant alignment, calling, and annotation

For variant alignment, CCS reads were used as input to lima to demultiplex the pooled samples. The demultiplexed data were then aligned to hg19 using pbmm2. GATK variant calling was performed with HaplotypeCaller (v.4.0.11.0) and Google DeepVariant (v.0.9.0) PacBio model and then merged using GLnexus (v.1.1.5). Variant annotation was processed with SnpEff (v.4.3r), ANNOVAR (v.2017-07-17), and GATK VariantAnnotator (v.3.8-1-0-gf15c1c3ef). PacBio RepeatAnalysis-Tools was used to generate the motif count for on-target reads spanning across the microsatellite repeat region and visualization of the repeat region.

Identification of microsatellite region and designation of microsatellite alleles for each sample

The microsatellite of interest in the 6p25.1 region was included in the sequenced amplicon. Microsatellite alleles were called by extracting the specific region of chr6: 6,837,193–6,837,301, corresponding to the coordinates of the microsatellite using UCSC STR information. The microsatellite region sequences were extracted from all CCS reads then regenerated for each sample in FASTQ format. All microsatellite sequences in each sample FASTQ file were read into R 4.1.2. The patterns of microsatellite repeat sequences were aggregated for each sample and the frequency of CCS reads for each pattern was calculated in each individual. The two most frequent read sequence patterns were assigned as microsatellite alleles for each sample, except in instances where greater than 60% of reads were one pattern, in which case the sample was assigned as homozygous at the microsatellite locus. All called microsatellite alleles were visually inspected and confirmed with the waterfall plot generated using CCS raw BAM files. Microsatellite alleles were then filtered to ensure each microsatellite allele contained greater than 15 CCS reads.

Phasing from sequenced variants

To harness the genomic phase information contained in long-read sequencing data, we developed a pipeline to phase microsatellite alleles as well as SNPs of interest directly from BAM files. Direct phasing from BAM files enables observation of long-range haplotypes without uncertainty resulting from phasing algorithms.[23] Using BAM files generated from SMRT sequencing, our custom pipeline enables phasing of microsatellite alleles as well as nearby SNP and indel alleles to determine whether alleles of specific variants of interest are on the same chromosome as a microsatellite allele. Samtools (v.1.8)[24] was used to read BAM files and extract CCS sequence reads and headers spanning the loci of interest. CCS reads that did not pass the previous filters were excluded in the phasing step. The output of the pipeline contains a unique CCS read header, sample information, and haplotype information for each variant.

Direct sequencing of rs17142617 and the 6p25.1 microsatellite

Genomic DNA from 38 EwS samples (cell lines and PDX) were used to perform targeted amplifications of the region encompassing both the lead SNP rs7742053 and the 6p25.1 GGAA microsatellite locus. Corresponding 5.7 kb amplicons were generated after 30 cycles of PCR amplifications using long range PrimerStar GXL DNA polymerase (Takara Bio) and checked by capillary electrophoresis on a Fragment Analyzer instrument (Agilent). 60 ng of each amplicon were used to prepare SMRTbell libraries according to "SMRTbell Barcoded Adapter Complete Prep Kit" protocol from Pacific Biosciences. After final quality assessment using BioAnalyzer system with DNA 12000 kit, the pool of libraries was hybridized with PacBio V3 Sequencing primer and bound to the Polymerase from the PacBio V3.0 Binding Kit. The final complex was loaded onto 1 SMRTcell of the Sequel I system. High-fidelity long reads (HiFi reads) were generated using Circular Consensus Sequencing tool (CCS), with 3 passes and accuracy of 99% as thresholds from PacBio's open-source SMRT Analysis software (SMRT Link). After demultiplexing using Lima, HiFi reads were aligned to hg19 using minimap2. Phased genotypes based on rs7742053 and GGAA microsatellite locus (chr6: 6,837,190–6,837,285) were extracted from each read. GGAA microsatellite structures were determined based on the predominant pattern found for each individual allele.

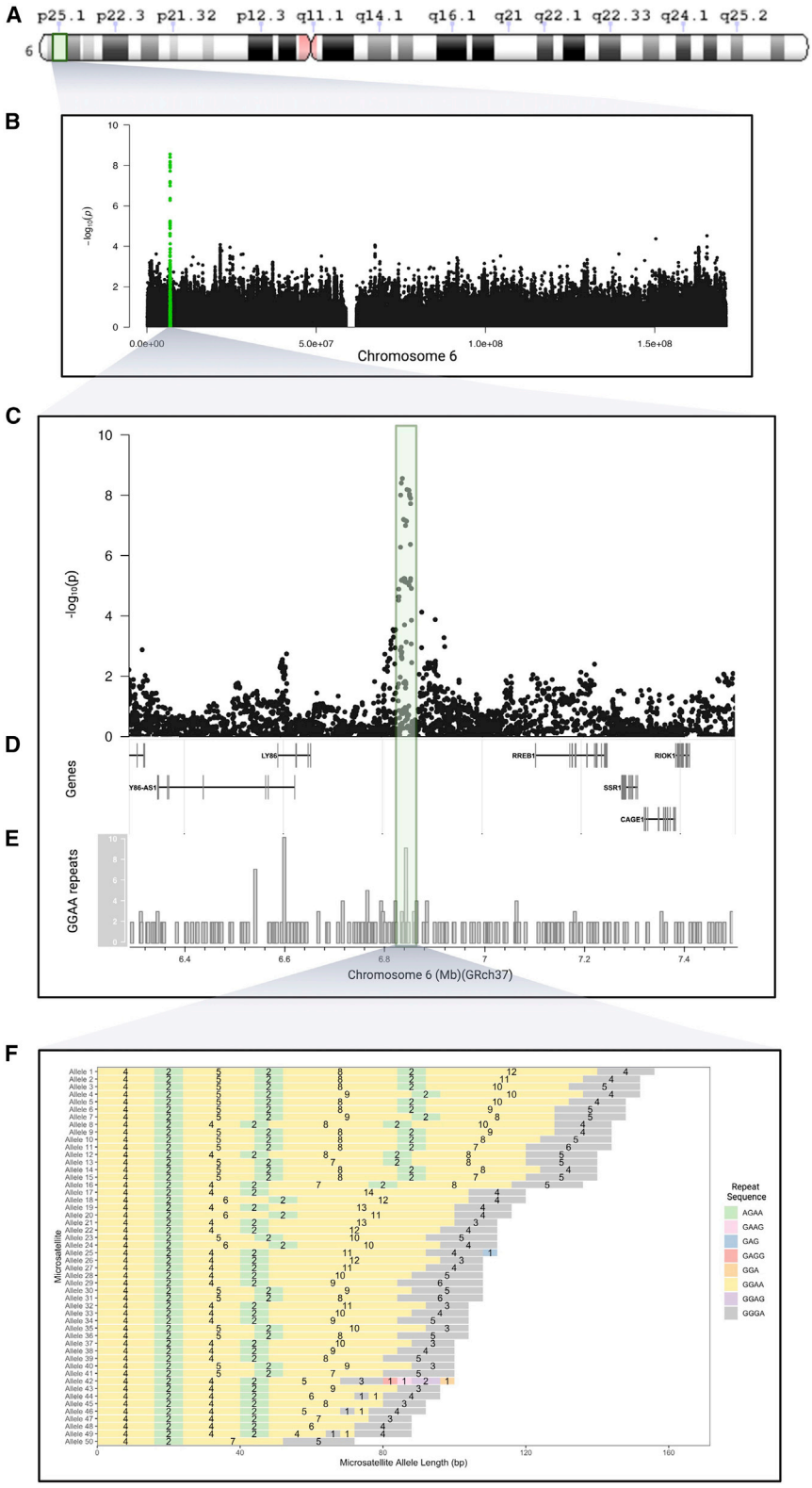

Figure 1. Regional fine mapping of 6p25.1 EwS susceptibility region and GGAA repeat microsatellite

(A) Cytoband location of 6p25.1 EwS susceptibility region.

(B) EwS GWAS association data indicating susceptibility signal on chromosome 6. Association signals from the 6p25.1 region are highlighted in green.

(C) Regional linkage disequilibrium patterns around the 6p25.1 EwS susceptibility region tagging SNP, rs7742053 (chr6: 6,300,000–7,500,000; hg19/GRCh37).

(D) Genes located near rs7742053: *LY86-AS1*, *LY86*, *RREB1*, *SSRA1*, *CAGE1*, and *RIOK*.

(E) GGAA microsatellite repeat sequence located at 6p25.1. The microsatellite of interest is boxed in green.

(F) Detected microsatellite alleles including GGAA motif repeat sequences at chr6: 6,837,194–6,837,283 (hg19/GRCh37). In total, 50 different microsatellite alleles were found from 271 EwS-affected individuals and 114 control samples. Colors represent different types of repeat motifs and the number represents frequency of each repeat sequence appeared consecutively.

Statistical analysis

All statistical analyses and visualizations were performed in R 4.1.2 (https://www.R-project.org/)[22] and Graphpad prism (v.9.3.1). Logistic regression was performed with MASS package (v.7.3.53.1) and the pseudo R square of each model was calculated with rms package (v.6.2.0). The linear regression model was performed with the stats package (v.4.0.3). Plots were created using the ggplot2 package (v.3.3.3). The genomic track and the GGAA counts plot included in Figure 1 were created using Gviz (v.1.34.1). The EwS GWAS plot was created with qqman (v.0.1.4).[25] For microsatellite analyses, all statistical analyses were performed in allele level in which two alleles were counted for each sample. Each sample had an allele score of long/long (2), long/short (1), and short/short (0) based on the length threshold of 135 bp as previously described.

Haplotype tags for longer GGAA alleles

A total of 109 variants shared between HaplotypeCaller (v.4.0.11.0) and Google DeepVariant (v.0.9.0) were identified. We restricted our analyses to 29 common variants with minor allele frequency >0.05. Then, univariable logistic and linear regression were performed to identify SNPs associated with EwS risk and microsatellite length, respectively. Using a p value threshold of 0.05, we identified three common SNPs which were associated with both EWS risk and microsatellite length. These three SNPs were used for further analyses of haplotype regression with EwS risk and the number of GGAA repeats in microsatellite alleles. Regression and filtering of SNPs were performed in R 4.1.2 and PLINK 1.90 (https://www.cog-genomics.org/plink/),[26] respectively.

Epigenomic visualization of the 6p25.1 microsatellite locus

Visualization of ChIP-seq tracks was performed using UW Epigenome browser (http://epigenomegateway.wustl.edu/browser/), using A673 ChIP-seq tracks for EWSR1-FLI1, H3K27ac, H3K4me3, and CTCF. An hg19 GGAA/TTCC track was used to indicate the number of consecutive GGAA/TTCC repeats in the hg19 genome. FLI1 and H3K27ac ChIP-seq profiles were examined in A673/TR/shEF cells with EWSR1-FLI1 KD,[27] and in EWIma1 cells (expressing EWSR1-FLI1) and parental MSC Pat cells (which do not express EWSR1-FLI1).[28] CTCF HiChIP data were also examined to visualize CTCF-mediated chromosome looping.[29]

RREB1 expression analysis with EWSR1-FLI1 expression and knockdown

RREB1 expression was examined in RNA-sequencing datasets of EWIma1 cells expressing *EWSR1-FLI1* versus parental MSC cells (MSC Pat).[28] *RREB1* expression was also assessed in microarray RNA expression datasets, derived from EwS cell lines with dox-inducible *EWSR1-FLI1*.[16] Gene expression counts were plotted in R using ggplot2 package (v.3.3.5).[11]

Cell culture

Human A673 EwS cells and A673 cells stably transfected with dCas9-KRAB were cultured in DMEM with 10% of fetal bovine serum (Eurobio) at 37°C, with 5% CO_2; media also contained 100 UI/mL penicillin and 100 mg/mL streptomycin. Cells were tested for mycoplasma regularly and were consistently mycoplasma negative. The A673 dCas9-KRAB-expressing cell line was created as described previously[12]; the A673 cell line was lentivirally transduced using the plasmid Lenti-dCas-KRAB-blast (Addgene) with a MOI of 3, cell selection was performed using 20 μg/mL blasticidin.

siRNA transfection

For qPCR knockdown validation following siRNA transfection, 75,000 A673 cells were seeded in each well of a 6-well plate. The following day, individual siRNAs against *RREB1* (Dharmacon ON-TARGETplus Human *RREB1* siRNA J-019150-05-0005, J-019150-07-0005, and J-019150-08-0005, referred to as si*RREB1* #1, #3, and #4, respectively) (Dharmacon/Horizon Discovery), and control siRNA (siCT, Dharmacon/Horizon Discovery) were transfected using lipofectamine RNAiMAX (Thermo Fisher/Invitrogen) (50 nM siRNA, lipofectamine ratio 1), using manufacturers protocols. All siRNAs used targeted all RefSeq *RREB1* isoforms. 72 h after transfection, cells were collected, and RNA was extracted (Machery-Nagel total RNA extraction kit). RNA was converted to cDNA, and qPCR was carried out using ABI SYBR green (Thermo Fisher) using a BioRad CFX96 machine, using primers against *RREB1* mRNA (5′-GTCTTTCACTGCCCAGTATGT-3′, 5′-GTGGGTTATCTGAATGGGTCTC-3′, primers targeted all RefSeq RREB1 isoforms). Fold knockdown was calculated using the delta delta Ct method, using *RPLP0* as a loading control.

CRISPR interference (CRISPRi)

CRISPR guide RNAs targeting the *EWSR1-FLI1*-bound 6p25.1 microsatellite were designed (Table S5). These RNAs were annealed with tracrRNAs (each at a final concentration of 10 nM). These duplexes were transfected into the A673 dCas9-KRAB cell line. Cells were collected 96 h after transfection. RNA was extracted then converted to cDNA. qPCR was performed with primers targeting *RREB1* mRNA, using RPLP0 as a control.

Cell proliferation and clonogenic assays

120,000 A673 cells were seeded in each well of a 6-well plate. si*RREB1* siRNAs #1, #3, and #4 and control siRNAs were transfected in the following day as described above (see siRNA transfection). At day 2 post transfection, cells were trypsinized in 2 mL of media and processed for cell proliferation and clonogenic assays. For cell proliferation assays, 100 μL of cell suspension was seeded into a 6-well plate with 2 mL media. The plates were placed in an incubator, and cells were trypsinized and counted at day 6. For clonogenic assays, the cell suspension at day 2 was diluted and was seeded into wells of a 6-well plate (1,000 cells, 500 cells per well). Colonies were stained after 11 days post seeding using crystal violet, then were imaged and counted.

RNA sequencing

RNA extracted from siRNA siRREB1 #1 and siCT (72 h post transfection) was analyzed by RNA sequencing. RNA-sequencing libraries were prepared from 1 μg of RNA using Truseq Stranded mRNA library preparation kit (Illumina, ref. 20020594). 100 bp paired end sequencing was performed on a NovaSeq6000 instrument, and 3 biological replicate pairs were sequenced. Initial processing of reads was performed using Nextflow (v.19.04.0), RNA-Seq pipeline v.3.1.8.[30] Briefly, read quality was assessed using FastQC (v.0.11.8) and reads were aligned to the hg19 human genome build, using STAR (v.2.6.1a_08–27).[31] Read counts generated by STAR were analyzed further; non-expressed/very-low-expression genes were filtered (gene row sums <6, across 6 samples), and differential gene expression analysis was carried out using DESeq2 (v.1.34.0).[32,33] Genes having a false-discovery corrected p value (padj) < 0.05 and a fold change >1.5 were considered as differentially expressed. Heatmaps were generated using the R package pheatmap (https://cran.r-project.org/web/packages/pheatmap/index.html).

The R package clusterProfiler (v.4.2.2)[34,35] was used to examine gene ontology signatures. Overlaps of sets of genes were performed using R package GeneOverlap (v.1.30.0.), using a Fisher exact test to assess statistical significance of overlap. RNA-seq data generated by this study are accessible on the Gene Expression Omnibus GEO repository, under accession number GSE220780.

Gene set enrichment analysis (GSEA)

Gene set enrichment analysis[36,37] was performed using the GSEA tool (v.4.1.0) (https://www.gsea-msigdb.org/gsea/index.jsp). RNA sequencing raw gene counts derived from STAR were first normalized using DESeq2 (v.1.34.0).[32,33] Non-expressed or low-expressed genes were filtered out (gene row sums <9). A count of +1 was added to all counts, to avoid infinite scores caused by fold change calculations dividing by 0. The table of normalized gene counts generated was used for GSEA analysis, using signal to noise as the enrichment metric. 1,000 gene set permutations were run to generate an empirical distribution of enrichment scores, and p value was calculated based on score distribution. A p value of 0 indicates that the p value is below <0.001, the minimum possible p value with 1,000 permutations.

Results

Characterizing germline complexity at 6p25.1

The 385 sequenced samples passing quality control filtering (Figure S1) consisted of 271 EwS-affected

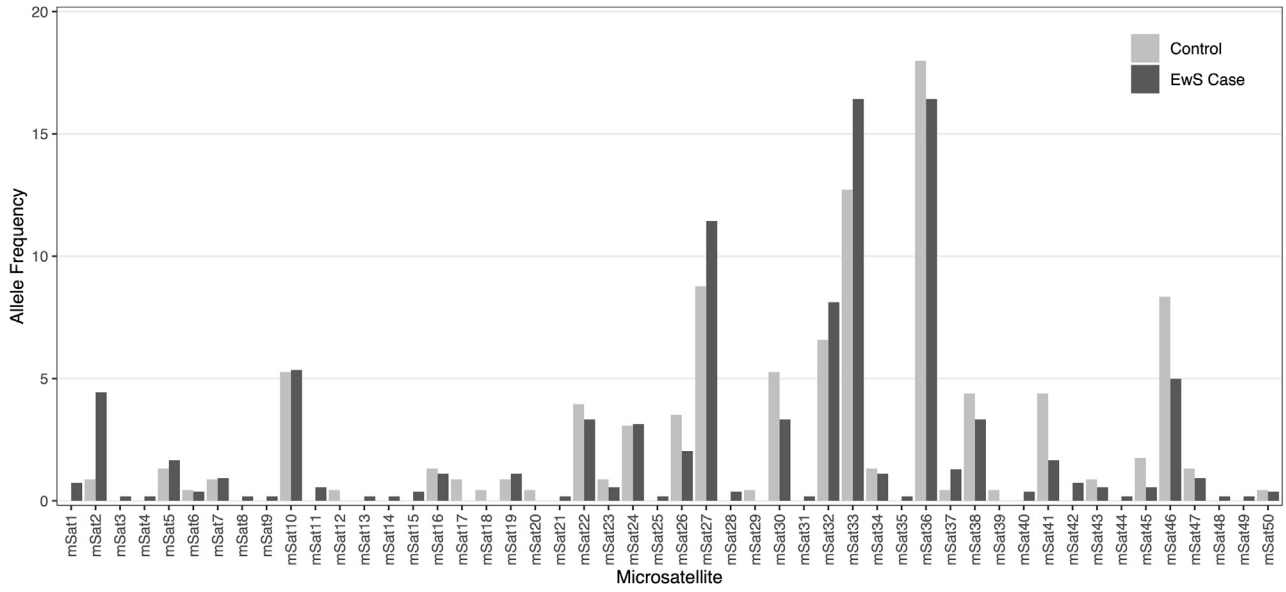

Figure 2. Allele frequency of the 50 microsatellite alleles at chr6: 6,837,194–6,837,283 (hg19/GRCh37) among EwS-affected individuals and control subjects

Distribution of EwS-affected individuals and control alleles among 50 identified microsatellite alleles (n = 542 for affected individuals and n = 228 for control subjects).

individuals and 114 control subjects with all samples of European ancestry. EwS-affected individuals were 50% female while the control subjects were 57% female. The linkage disequilibrium block for the 6p25.1 EwS susceptibility region spanned approximately 40 kb from 6,820,000–6,860,000 (GRCh37/hg19),[38] of which we chose an approximate 4.8 kb region from chr6: 6,835,220–6,839,996 (GRCh37/hg19) for targeted long-read sequencing that included the GGAA microsatellite of interest. The investigated GGAA microsatellite (chr6: 6,837,194–6,837,283) had more than 4 GGAA motifs, evidence of open chromatin and evidence of EWSR1-FLI1 oncoprotein binding from previous observervations.[20] We identified 70 common SNPs and 35 indels in the targeted sequencing region that corresponded to the region upstream of the lead GWAS SNP previously reported.[20] The GGAA microsatellite of interest was highly polymorphic with 50 distinct GGAA microsatellite alleles, which we annotated into 8 blocks (Figures 1 and S2). Most microsatellite allele patterns had two stretches of AGAA motifs located between long consecutive GGAA motifs, dividing the microsatellite into three shorter blocks of consecutive GGAA motifs (Figures 1 and S2). We detected a total of 44 microsatellite alleles in EwS-affected individuals and 31 alleles in ancestry-matched, cancer-free control subjects. There were 19 microsatellite alleles unique to EwS-affected individuals and 6 alleles unique to control subjects (Figure S3).

Longer microsatellite alleles are more common in EwS-affected individuals

The most frequent microsatellite allele for both EwS-affected individuals and control subjects (allele 36) had a length of 104 bp with a total of 17 GGAA motifs (Figure 2). Initially, microsatellite alleles were categorized into long and short groups based on a length threshold of greater than or equal to 135 bp, which was the 90th percentile of microsatellite length in control subjects. The long microsatellite allele group consisted of 16 microsatellite alleles and generally carried 2 additional AGAA repeats and 1–15 additional GGAA repeats compared to the short microsatellite group. The long microsatellite allele group was more frequently observed in EwS-affected individuals compared to control subjects (n = 90/452, 19.9% vs. n = 24/204, 11.8%; p = 0.04). When analyzing the length of the microsatellite allele as a continuous variable, we observed a positive relationship with EwS risk with a 16 base pair increase in microsatellite allele length (equal to one SD in length of all microsatellite alleles) associated with an increased odds of EwS of 24% (OR = 1.24, 95% CI = 1.05–1.48, p = 0.01) (Figure 3).

The number of GGAA repeats is associated with EwS risk

There were three core repeat motifs commonly observed in the 50 microsatellite alleles: GGAA, AGAA, and GGGA (Figures S4 and S5). To identify which motif was most strongly associated with EwS risk, we divided microsatellite alleles into counts of the common component repeat motifs and examined them for motif count associations with EwS risk. The average number of GGAA motifs in EwS-affected individuals was 19.20 compared to 18.45 among control subjects (p = 0.003) (Table S1). We additionally compared the length and number of GGAA motifs for the longest allele carried by each sample to minimize potential attenuation of effect by the other shorter

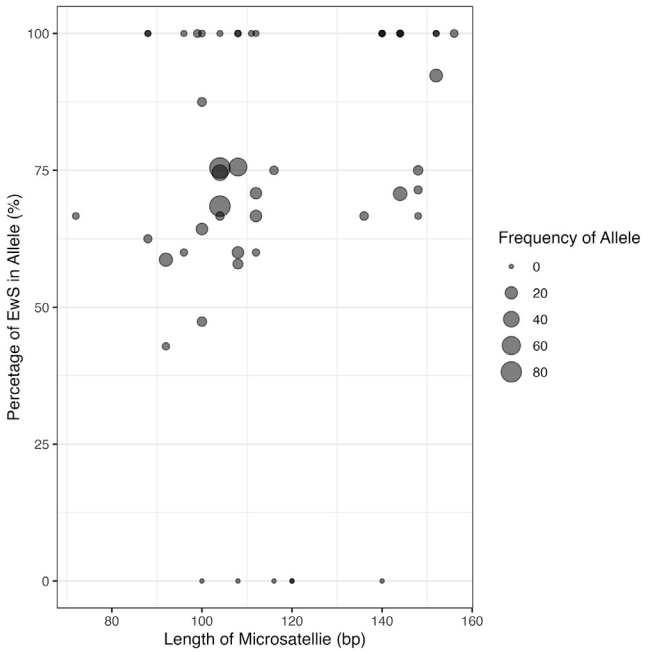

Figure 3. Proportion of EwS-affected individuals and length of microsatellite alleles

Each circle represents one of the 50 microsatellite alleles identified in chr6: 6,837,194–6,837,283 (hg19/GRCh37). The size of the circle represents the frequency of each allele in 385 EwS-affected individuals and control subjects. p value from logistic regression model = 0.015.

allele. We observed a slight increase in the difference of both the length of microsatellite (119.20 vs. 114.10, p = 0.02) and number of GGAA repeats (20.97 vs. 19.88, p = 0.02) between EwS-affected individuals and control samples. Logistic regression models confirmed an association between the number of GGAA motifs and increased EwS risk (OR = 1.07, 95% CI = 1.02–1.13, p = 0.006) (Table S2). Of the 50 microsatellite alleles observed, the longest consecutive stretch of GGAA repeats was generally located in the third or fourth GGAA repeat expansions (Figure S4). We tested whether the longest consecutive stretch of GGAA repeats significantly differed between the EwS-affected individuals and cancer-free control subjects and noted longer mean consecutive GGAA stretches among cancer-free control subjects (EwS mean = 9.11, control mean = 9.40; p = 0.029). These observations were surprising and suggest that the total count of GGAA motifs at 6p25.1 is more strongly associated with EwS risk than consecutive GGAA count. Longer microsatellite alleles at the 6p25.1 locus generally contained an additional AGAA repeat (AGAA expansion 3) and divided a long stretch of GGAA motifs into two shorter GGAA repeat expansions (GGAA expansions 3 and 4), resulting in a lower count of consecutive GGAA motifs but a higher count of total GGAA repeats (Figure S4). When combining GGAA repeat expansions 3 and 4, we found the total number of GGAA motifs in these expansions was higher in longer microsatellite alleles (mean = 10.73 vs. 9.99, p = 0.002). In addition, we divided total GGAA motif count into deciles and found no signs of a threshold above which EwS risk was substantially elevated (Figure S6), supporting a positive association between the number of GGAA motifs and EwS risk at the 6p25.1 EwS susceptibility region.

We also observed a marginally significant association with AGAA repeat motifs and increased EwS risk (OR = 1.28, 95% CI = 1.02–1.62, p = 0.04). The number of AGAA repeat motifs was highly correlated with the number of GGAA motifs (r = 0.87) (Figure S7), as alleles with more GGAA motifs contained two additional AGAA motifs in the middle of GGAA expansion 3 and 4 (Figure S4). We included both AGAA and GGAA count as predictors of EwS into the same logistic regression model to determine which was most associated with EwS risk. In this model, the resulting association estimate for AGAA count was substantially attenuated and the association p value was insignificant (OR = 0.89, 95% CI = 0.57–1.39, p = 0.59), suggesting GGAA motif count and not AGAA motif count was the main contributor to EwS risk. We further tested for an association between GGGA motifs and EwS risk but did not find evidence suggesting GGGA motifs were associated with EwS risk (OR = 0.88, 95% CI = 0.71–1.09, p = 0.23).

The risk allele of the GWAS tagging variant (rs7742053-A) is correlated with longer microsatellite alleles

In the previous EwS GWAS, the top associated tagging variant for the 6p25.1 region was rs7742053 (OR = 1.80, 95% CI = 1.48–2.18, p = 2.78×10^{-9}) and the minor (A) allele was associated with increased EwS risk.[20] Since by design, rs7742053 was not included within our targeted sequencing region, we used the highly correlated rs17142617 variant ($r^2_{EUR} = 0.95$, $D'_{EUR} = 0.99$),[38] which was the second-most GWAS-associated variant located in the targeted region (p = 9.93×10^{-9}), as a surrogate in high linkage disequilibrium with the minor (G) allele of rs17142617 corresponding to the EwS risk-associated allele. We observed the rs17142617 (G) risk allele was highly correlated with longer microsatellite length (r = 0.98, p = 3.55×10^{-10}) and was associated with a higher number of GGAA motifs (p = 3.97×10^{-79}). These findings indicate that the EwS risk locus tagged by rs7742053 is linked to longer microsatellite alleles with more GGAA motifs. To confirm the direct association between longer microsatellite length and the EwS risk locus, we performed additional targeted long-read sequencing of the 6p25.1 EwS susceptibility region encompassing both rs17142617 (the lead GWAS variant) and the GGAA microsatellite in 38 samples. We confirmed that haplotypes carrying the risk (A) allele at rs17142617 (n = 12) contained a greater number of GGAA motifs relative to haplotypes with the non-risk (C) allele (n = 64) (mean GGAA (A) allele = 25.75, mean GGAA (C) allele (C) = 18.09, p = 1.573×10^{-7}; Figure S8).

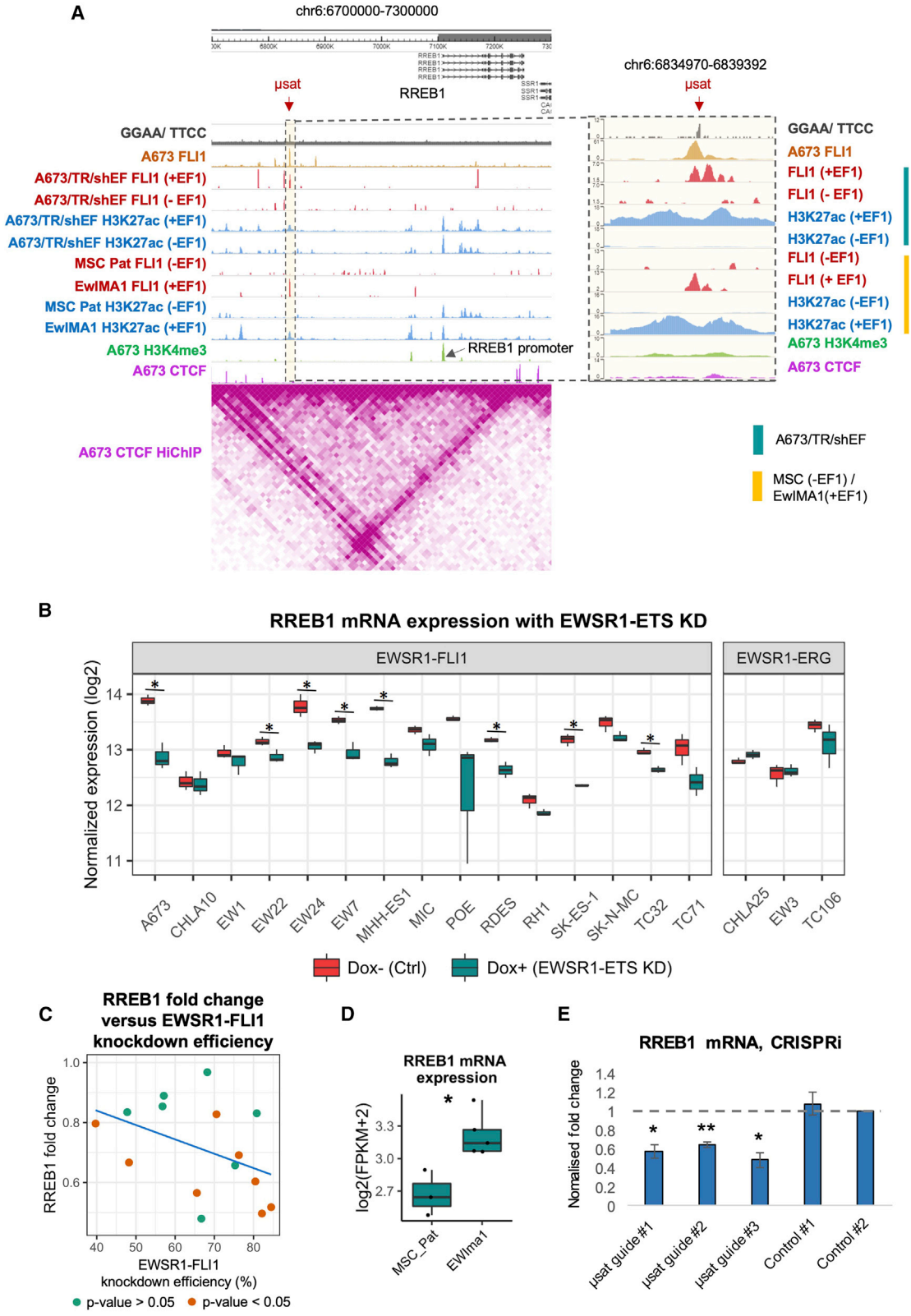

Figure 4. GWAS-identified locus at 6p25.1 shows a GGAA-microsatellite which shows features of an EWS-FLI1 enhancer, and which is linked to *RREB1* expression

(A) UW epigenome browser visualization of the 6p25.1 polymorphic microsatellite and downstream *RREB1*. Tracks shown for hg19 GGAA/TTCC repeats, ChIP-seq tracks for EWS-FLI1, H3K27ac, H3K4me3, CTCF, and HiChIP profile for CTCF. Right panel: zoom-in focusing on the region including the 6p25.1 microsatellite. Tracks were derived from A673, A673/TR/shEF (with and without dox inducible silencing of EWS-FLI1: +EF1 and −EF1), MSC Pat (−EF1) and EWIma1 (+EF1, derived from MSC Pat cells). FLI1 and H3K27ac

(legend continued on next page)

A three-SNP haplotype tags longer GGAA alleles at 6p25.1

In total, we identified three SNPs in the sequenced region (rs17142617, rs74781311, and rs2876045) for which the minor alleles showed a significant association with both EwS risk (OR = 1.655, 3.376, and 1.588, 95% CI = 1.04–2.73, 1.44–9.91, and 1.00–2.62, p = 0.004, 0.0003, and 0.008; respectively) and longer microsatellite length ($p = 3.55 \times 10^{-10}$, 2.72×10^{-21}, and 1.04×10^{-23}; respectively) in the 6p25.1 EwS susceptibility region (Table S2). We examined the association of haplotypes containing these three SNPs with the total number of GGAA motifs in microsatellite alleles (Table S3). Compared to the most frequently occurring haplotype (ATT), a haplotype carrying the strongest associated alleles (GGT) had on average an additional 10 GGAA repeats ($p = 4.82 \times 10^{-7}$) in the microsatellite (Table S3, Figure S9). We observed that haplotypes starting with the EwS risk allele (G) at rs17142617 had an additional 8 GGAA repeats on average, compared to the haplotypes starting with (A) at rs17142617 ($p = 1.31 \times 10^{-64}$) (Figure S9). Accompanying (G) at the first position, haplotypes carrying (G) in the middle position, the minor allele at rs74781311, had an additional 3 GGAA repeats on average compared to the haplotypes carrying the major (T) allele ($p = 5.27 \times 10^{-12}$). As these haplotypes are informative surrogates of GGAA repeat count in the 6p25.1 microsatellite region (regression $R^2 = 0.62$), haplotypes generated from these genotyped surrogates could be used in studies for which detailed sequencing of this region is not possible. Compared to models with individual SNPs, the R^2 improved significantly with the haplotype-based model ($R^2_{rs17142617} = 0.37$, $R^2_{rs74781311} = 0.19$, and $R^2_{rs2876045} = 0.35$). The haplotypes described are based on variants called from long-read sequencing. Haplotypes constructed from other approaches such as genotyping or imputation are anticipated to have similar performance but have not been tested.

The 6p25.1 GGAA microsatellite shows features of an EWSR1-FLI1 enhancer

Since EWSR1-FLI1 is known to bind GGAA microsatellites to create *de novo* enhancers in EwS,[2–4] we investigated whether this microsatellite at 6p25.1 showed features of an EWSR1-FLI1 enhancer. Using ChIP-seq data from A673 EwS cells, we examined the genomic binding of EWSR1-FLI1 and H3K27ac, an enhancer-associated histone mark indicating open chromatin. Consistent with a previously reported observation,[20] we noted that the GGAA microsatellite of interest showed binding of EWSR1-FLI1 and H3K27ac, suggesting that this microsatellite is in an area of open chromatin and interacts with EWSR1-FLI1 (Figure 4A). To assess whether EWSR1-FLI1 was linked with the presence of this enhancer, we examined H3K27ac binding in the presence and absence of EWSR1-FLI1. We first examined H3K27ac binding in A673/TR/shEF cells,[39] which contained a doxycycline-inducible shRNA targeting EWSR1-FLI1. In these cells, silencing of EWSR1-FLI1 was accompanied by a strong decrease in H3K27ac ChIP enrichment (Figure 4A, right panel). We also examined the effects of EWSR1-FLI1 expression in mesenchymal stem cells (MSCs), the putative cell of origin in EwS.[40] We compared with H3K27ac binding in EWIma1 cells which have been genetically engineered to recapitulate the *EWSR1-FLI1* chromosomal translocation[28] and which are derived from MSC parental cells (MSC Pat). We observed that the presence of *EWSR1-FLI1* fusion results in a gain of *EWSR1-FLI1* and H3K27ac enrichment at the 6p25.1 locus (Figure 4A, right panel). Collectively, these observations indicate that the 6p25.1 GGAA microsatellite shows features of a *de novo* EWSR1-FLI1 enhancer. In addition, examining CTCF HiChIP data[29] in the A673 EwS cell line, we found evidence that the GGAA microsatellite and the *RREB1* promoter were located within the same topologically associated domain (TAD) (Figure 4A, left panel). Since the TAD defines local regions of increased chromatin interaction and is relatively isolated from elements outside the TAD, the presence of the microsatellite and the *RREB1* promoter within this TAD may facilitate their interaction.

EWSR1-FLI1 microsatellite enhancer at 6p25.1 regulates the expression of RREB1

To investigate the effect of EWSR1-FLI1 on the expression of *RREB1*, we examined the effects of EWSR1-FLI1 knockdown in EwS cells, and of EWSR1-FLI1 expression in MSC cells. We first examined the effects of EWSR1-ETS (EWSR1-FLI1 and EWSR1-ERG) knockdown on *RREB1* mRNA expression, using a gene expression dataset from the Ewing Sarcoma Cell Line Atlas (ESCLA).[16] We analyzed Affymetrix gene expression data in 18 EwS cell lines with dox-inducible knockdown of EWSR1-FLI1 and EWSR1-ERG in cell lines

ChIP-seq track intensity in A673/TR/shEF and MSC Pat/EwIMA1 were scaled to equal min, max values, to allow quantitative comparison of −EF1 and +EF1 conditions.
(B) Effects of EWSR1-ETS (EWSR1-FLI1 and EWSR1-ERG) silencing on *RREB1* mRNA expression in EwS cell lines. *RREB1* mRNA expression (Affymetrix) was assessed with/without doxycycline-induced silencing of EWSR1-FLI1/EWSR1-ERG, in EwS cell lines bearing these respective gene fusions (ESCLA dataset, Orth et al.[16]).
(C) Correlation scatterplot examining *RREB1* fold change upon EWSR1-FLI1 knockdown (doxycycline induced) versus EWSR1-FLI1 knockdown efficiency, across 15 EwS cell lines bearing an EWSR1-FLI1 fusion (ESCLA dataset, Orth et al.[16]). Each dot represents a cell line, dot color indicates statistical significance ($p < 0.05$, t test) of *RREB1* reduction within individual cell lines, and linear regression is indicated in blue ($R = -0.43$, $p = 0.11$).
(D) *RREB1* expression (RNA sequencing) in MSC parental (MSC Pat) cells vs. EWIma1 cells, expressing EWSR1-FLI1.
(E) CRISPRi silencing of the 6p25.1 microsatellite, RNA qPCR showing *RREB1* expression in A673 dCas9-KRAB cells transfected with CRISPR guides targeting the microsatellite versus control guides. Mean ± SEM across biological replicates ($n \geq 3$) is indicated. *$p < 0.05$, **$p < 0.01$.

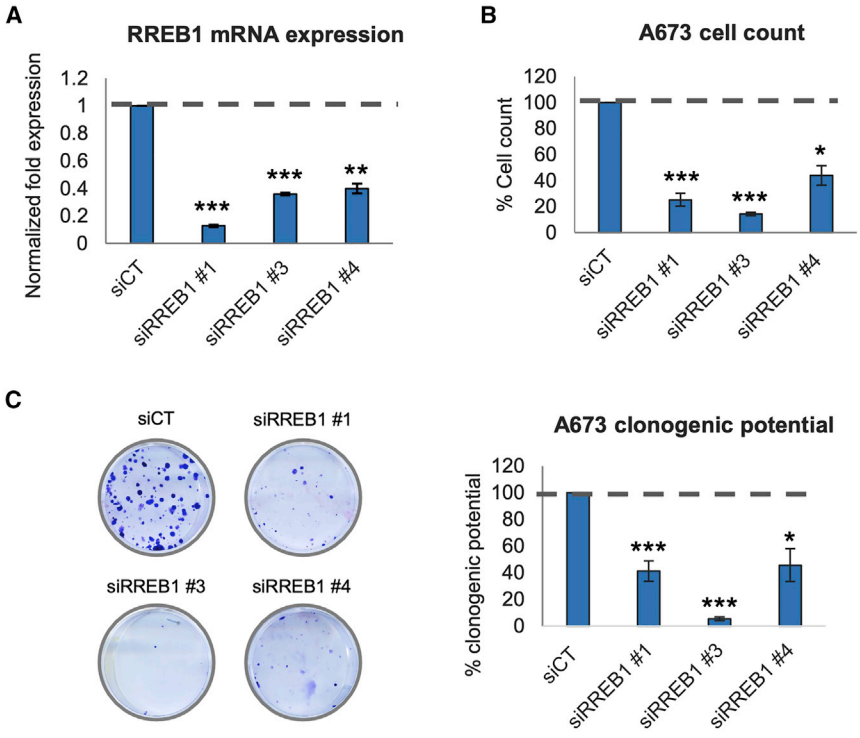

Figure 5. *RREB1* knockdown reduces proliferation and clonogenicity in A673 EwS cells

(A) RNA qPCR showing *RREB1* expression in A673 cells 72 h post transfection with siRNAs, using RPLP0 mRNA as a control.

(B) A673 cell count 96 h post siRNA transfection.

(C) Clonogenic potential assay, A673 colony count (expressed as a percentage relative to control siRNA) at day 10 post seeding ($n \geq 4$). For all bar plots, mean ± SEM across biological replicates ($n \geq 3$) is indicated. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

with these fusions (15 and 3 cell lines, respectively). We observed that 8 of 15 EWSR1-FLI1 cell lines showed a statistically significant downregulation of *RREB1* ($p = 0.009$ for A673), and 13 of 15 of these cell lines showed a trend of *RREB1* downregulation, upon EWSR1-FLI1 knockdown (Figure 4B). Further, in EWSR1-FLI1 cell lines, the fold reduction of *RREB1* upon EWSR1-FLI1 knockdown may be at least partially linked to EWSR1-FLI1 knockdown efficiency (Figure 4C, $R = -0.43$, $p = 0.11$). Cell lines showing a higher efficiency of EWSR1-FLI1 knockdown showed a tendency to also show a stronger downregulation of *RREB1*. On the other hand, for EWSR1-ERG-bearing EwS cell lines, no significant reduction of *RREB1* was noted upon EWSR1-ERG knockdown (Figure 4B). We also examined the effects of EWSR1-FLI1 expression in MSC cells by comparing EWIma1 and MSC Pat cells. Using RNA-sequencing data from both cell lines, we observed that EWIma1 cells which express the EWSR1-FLI1 fusion showed an elevated expression of RREB1 ($p < 0.021$) (Figure 4D). Thus, EWSR1-FLI1 knockdown leads to downregulation of *RREB1*, and EWSR1-FLI1expression in non-EwS cells leads to upregulation of *RREB1*, demonstrating that EWSR1-FL1 regulates *RREB1* expression.

To investigate whether EWSR1-FLI1 could regulate *RREB1* expression by this putative GGAA microsatellite enhancer in 6p25.1, we performed targeted silencing of this microsatellite in A673 EwS cells using CRISPR interference (CRISPRi). This approach allows transcriptional silencing of target loci through a KRAB domain bound to a catalytically inactive form of Cas9. We transfected three CRISPR guide RNAs targeting this microsatellite, as well as two control CRISPR guide RNAs into an A673 cell line which had been engineered to stably express dCas9-KRAB (A673 dCas9-KRAB). Using RNA qPCR, we observed that CRISPR guide RNAs targeting this microsatellite reduced the expression of RREB1 mRNA relative to control guide RNAs ($p = 0.024$, 0.008, and 0.022; for CRISPRi microsatellite/μsat guide RNAs #1, #2, #3, respectively) (Figure 4E) showing that silencing of the 6p25.1 EWSR1-FLI1 GGAA microsatellite enhancer can lead to reductions in *RREB1* expression. These findings suggest that *EWSR1-FLI1*-dependent regulation of *RREB1* expression at 6p25.1 is at least partially mediated through the GGAA microsatellite.

RREB1 knockdown attenuates proliferation and clonogenic potential in EwS cells

Observing that the GWAS risk allele at 6p25.1 is associated with elevated *RREB1* expression and that the 6p25.1 EWSR1-FLI1 microsatellite enhancer regulates expression of *RREB1*, we investigated the impact of *RREB1* expression on EwS proliferative phenotypes. We attempted to silence *RREB1* in A673 cells using transfection with 4 individual siRNAs targeting *RREB1* (si*RREB1* #1, #2, #3, and #4) and a control siRNA (siCT). Using RNA qPCR, we noted a strong reduction of *RREB1* mRNA with 3 siRNAs, siRREB1 #1, #3, and #4 (87.2%, 64.0%, and 60.1% knockdown, $p = 1.0 \times 10^{-4}$, 2.0×10^{-4}, and 3.6×10^{-3}; respectively) (Figure 5A). We examined the impact of reduced *RREB1* expression on cell viability and proliferation in control and *RREB1* knockdown conditions. We observed that *RREB1* knockdown reduced cell counts at 6 days after transfection with siRNAs ($p = 1.0 \times 10^{-4}$, 1.0×10^{-4}, and 1.7×10^{-2}; respectively) (Figure 5B). We also examined clonogenic potential in these cells and observed that all siRNAs reduced clonogenic potential ($p = 1.0 \times 10^{-4}$, 1.0×10^{-4}, and 2.1×10^{-2}; respectively) (Figure 5C).

Transcriptional effects of RREB1 knockdown in EwS cells

As our results suggest EWSR1-FLI1 binding to the 6p25.1 microsatellite regulates *RREB1* expression, we examined

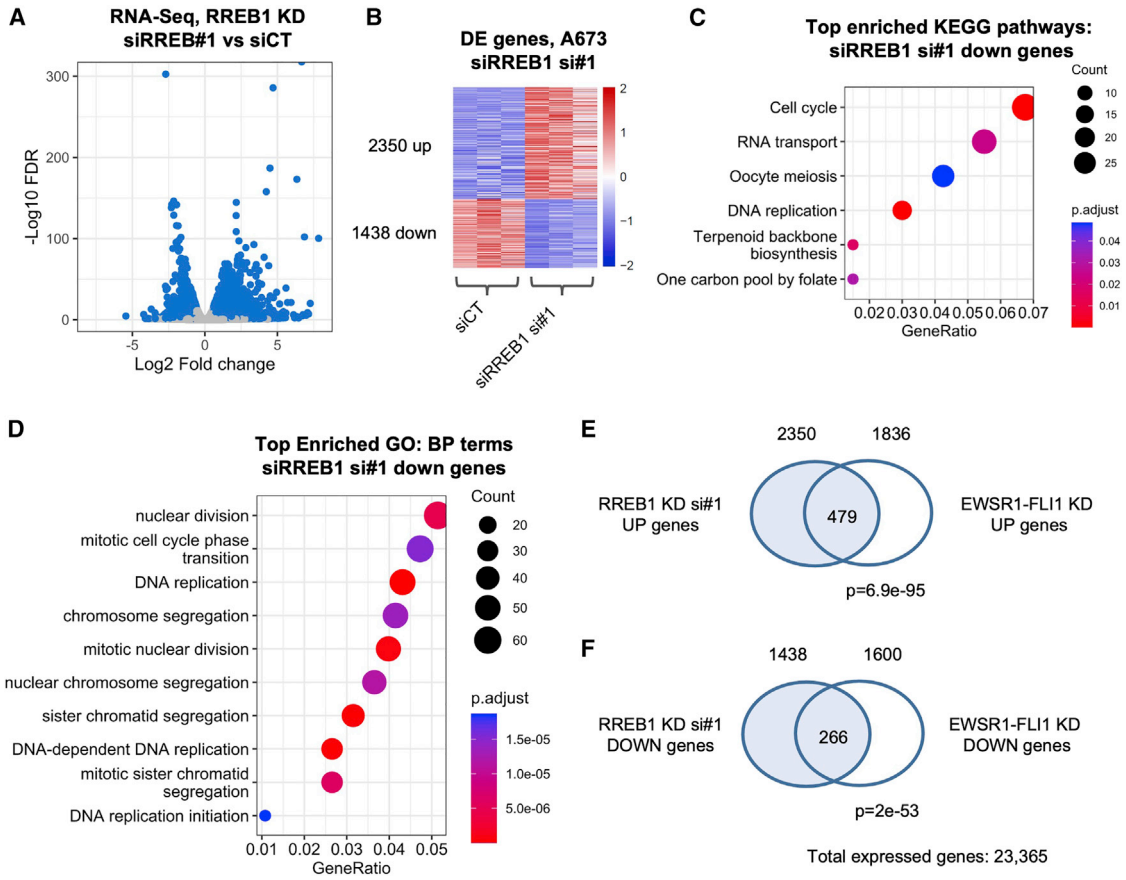

Figure 6. Transcriptional effects of *RREB1* silencing in A673 EwS cells

(A) Volcano plot showing differential gene expression upon *RREB1* knockdown using siRNA siRREB1 #1, relative to siRNA control siCT. Plot depicts log2 fold change vs. −log10 false discovery-corrected p value (FDR) for individual genes. Differentially expressed genes are represented in blue.

(B) Heatmap representation of differentially expressed genes upon transfection with siRREB1 si#1, relative to control siRNA (siCT). Expression values of genes are scaled by row.

(C and D) Plot showing enriched KEGG pathway (C) and gene ontology, biological process (GO BP) (D) terms in genes downregulated upon transfection with siRREB1 si#1, relative to control siRNA. Gene ratio indicates proportion of the GO/KEGG term containing query genes, dot size indicates number of query genes in the GO term, and color indicates the false-discovery rate corrected/adjusted p value.

(E) Overlap between genes upregulated with *RREB1* knockdown by siRREB1 si#1, and genes upregulated with EWSR1-FLI1 knockdown in A673 cells.

(F) Overlap between genes downregulated with *RREB1* knockdown by siRREB1 si#1, and genes downregulated with EWSR1-FLI1 knockdown in A673 cells. p values in (E) and (F) were computed using Fisher exact test.

transcriptional effects of *RREB1* knockdown in EwS cells. We used RNA sequencing to perform gene expression profiling in A673 cells transfected with si*RREB1* #1, which showed the strongest knockdown of *RREB1* compared to expression profiles of cells transfected with siRNA control siCT (n = 3) (Figure S10). Using an adjusted p value threshold <0.05 and fold change >1.5 to define differentially expressed (DE) genes, we noted that *RREB1* silencing showed a strong effect on gene expression patterns with 2,350 genes upregulated and 1,438 downregulated (Figures 6A and 6B). In enrichment analyses of differentially expressed genes in Gene Ontology Biological Process (GO BP) terms and KEGG pathways, downregulated genes were enriched for cell cycle and DNA replication terms and for cell adhesion molecules (Figures 6C and 6D). The effect of *RREB1* knockdown on DNA replication was consistent with the impact we observed of *RREB1* knockdown on cell proliferation and clonogenicity. We also confirmed these findings using Gene Set Enrichment Analysis (GSEA), a threshold-independent approach to examine enriched GO BP processes. Top GSEA-enriched terms similarly indicated a strong effect on DNA replication-related GO BP terms (Figure S11). Lastly, to assess whether *RREB1* and EWSR1-FLI1 could show similarities in transcriptional target profiles, we examined for overlaps between DE genes upon *RREB1* knockdown and DE genes upon EWSR1-FLI1 knockdown in A673 cells. Interestingly, we observed a strong overlap between genes upregulated by *RREB1* and EWSR1-FLI1 knockdown ($p = 6.9 \times 10^{-95}$) (Figure 6E). Similarly, we noted a significant overlap between genes downregulated upon knockdown of *RREB1* and those downregulated upon EWSR1-FLI1 knockdown ($p = 2.0 \times 10^{-53}$) (Figure 6F). Since *RREB1* is a downstream target of EWSR1-FLI1, these findings suggest that *RREB1*

could act as a transcriptional effector downstream of EWSR1-FLI1, contributing to the overall EWSR1-FLI1 transcriptional program.

Discussion

Our study characterized germline variants in a highly polymorphic microsatellite region at the 6p25.1 EwS susceptibility locus and identified a germline-somatic relationship between longer microsatellite alleles with higher numbers of GGAA motifs and increased expression of *RREB1* in the presence of EWSR1-FLI1. Based on targeted long-read sequencing, we found diverse patterns of microsatellite alleles consisting of three major motifs: GGAA, AGAA, and GGGA (Figures S3 and S4). Our analyses suggested that longer microsatellites are more frequent in EwS-affected individuals compared to control subjects, and specifically microsatellite alleles containing a higher number of GGAA motifs were more highly associated with EwS risk. The risk allele (A) of rs7742053, the lead marker variant of 6p25.1 EwS susceptibility locus, is in high linkage disequilibrium with a set of microsatellites containing increased counts of GGAA repeats and is associated with increased expression of *RREB1*. These observations suggest that risk at the 6p25.1 EwS susceptibility locus is at least partially explained by an oncogenic interaction between germline variation in the GGAA microsatellite with somatic EWSR1-FLI1 fusion binding. This is important for EwS susceptibility as germline variation that increases the binding of the somatically acquired EWSR1-FLI1 fusion could drive dysregulation of local genes, *RREB1* here, and EwS risk. Moreover, we observed evidence that the characterized GGAA microsatellite acts as an EWSR1-FLI1 enhancer by binding with EWSR1-FLI1. A difference in *RREB1* expression was observed based on the presence of EWSR1-FLI1, confirming that EWSR1-FLI1 binding increases expression of *RREB1*. Likewise, silencing of the GGAA microsatellite in 6p25.1 leads to reduced *RREB1* expression. Taken together, these observations suggest that regulation of *RREB1* expression is mediated through the interaction between germline GGAA microsatellite variation and the somatic EWSR1-FLI1 fusion protein.

RREB1 is a large zinc-finger transcription factor that binds specifically to the ras-responsive element (RRE) of select genes that, in turn, regulates promoter activity and gene transcription.[41] So far, *RREB1* has been shown to regulate cell growth, DNA damage repair, and epithelial-mesenchymal transition, and functions both as a transcriptional repressor and activator in the development of tumorigenesis by interacting with distinct signaling pathways.[41–43] We confirmed that *RREB1* knockdown reduces proliferation and clonogenicity in EwS cells, indicating that *RREB1* is a key gene for EwS cell growth and proliferation. In addition, RNA-sequencing analyses showed that siRREB1-downregulated genes were enriched with DNA replication-related biological pathways, suggesting a potential role of *RREB1* in DNA replication and cell cycle, particularly in EwS. Interestingly, two EwS susceptibility genes, *RREB1* and *EGR2*,[14] are downstream targets of the MAPK pathway[44,45] which plays a critical role in EwS development and chemoresistance as the pathway regulates cell proliferation and invasiveness of the bone malignancies, including EwS.[46,47]

Previously, it was found that DNA binding of EWSR1-FLI1 was observed only when four or more consecutive GGAA motifs were present[14] and that transcription activity increased as the number of GGAA motifs increased.[9] For several microsatellite alleles located in 6p25.1, AGAA motifs are located within consecutive GGAA motifs, creating 2 to 4 blocks of separate stretches of GGAA repeats. We tested whether consecutive stretches of GGAA repeats were more important for EwS risk than total number of GGAA repeats and found evidence that the total number of GGAA repeats in the 6p25.1 microsatellite region was more strongly associated with EwS risk than consecutive stretches. Recent evidence suggests that GGAA microsatellites bound by EWSR1-ETS fusions have additional GGAA repeats in the vicinity as well as a low number of interspersed bases contiguous to the adjacent GGAA repeats,[16] suggesting that EWSR1-FLI1 binding could be more complicated than simply depending on contiguous stretches of GGAA repeats. Microsatellite alleles located in 6p25.1 likewise have shorter GGAA blocks near the longest GGAA stretch and have AGAA motifs as flanking bases in the middle of two GGAA blocks. Further study is needed to investigate the functional dynamics of how variation in the 6p25.1 GGAA microsatellite impacts EWSR1-FLI1 binding and downstream transcriptional activity of *RREB1*. Other features of this microsatellite beyond length of GGAA motifs could be involved in EwS tumorigenesis, such as chromatin organization,[48–50] that need to be further investigated.

While we observed a strong association between EwS lead GWAS SNPs and the number of GGAA repeats, the difference in the number of GGAA repeats between EwS-affected individuals and cancer-free control subjects is less striking (19.20 versus 18.45, $p = 0.007$). This observation could indicate that *RREB1* susceptibility locus is functionally relevant for only a subset of EwS. For example, some EwS could be highly dependent upon *RREB1* activation, while others could be more dependent on other regions detected by the GWAS. Such etiologic heterogeneity has been previously observed in genetic susceptibility to other cancers.[51–53] In this respect, it is interesting to note that *RREB1* and *EGR2*, another susceptibility gene identified by the GWAS,[14] are both important downstream players of the MAPK pathways.

In conclusion, our findings describe the oncogenic interplay between a somatic driver mutation, EWSR1-FLI1, and a highly polymorphic GGAA microsatellite in the 6p25.1 EwS GWAS susceptibility locus. Longer GGAA repeats at 6p25.1 act as a *de novo* enhancer in the presence of EWSR1-FLI1 and confer an increase in EwS

risk through enhanced binding affinity with EWSR1-FLI1 fusion protein and upregulation of *RREB1*, which we observed to increase EwS cell proliferation and potentially DNA replication. A similar mechanism was observed for *EGR2* at the 10q21.3 EwS susceptibility locus[14] in which an oncogenic interaction between the EWSR1-FLI1 somatic driver mutation and regulatory germline susceptibility variants influences EwS. Our results demonstrate the utility of long-read sequencing for disentangling EwS germline risk in highly polymorphic tandem repeat sequences. Pursuit of this sequencing strategy to characterize additional EwS susceptibility loci as well as follow-up functional investigation of candidate genes could improve understanding of germline-somatic relationships in EwS susceptibility.

Data and code availability

The datasets reported during this study are available at dbGaP http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs003159.v1.p1(dbGaP: phs003159.v1.p1). RNA-sequencing data generated in this study are accessible on the Gene Expression Omnibus GEO repository(GEO: GSE220780). Code for gene expression analysis are available at https://github.com/calvin-s-rodrigues/EwS_6p25.1_RREB1. Code for 6p25.1 microsatellite and germline variants analyses and visualizations are available at https://github.com/machiela-lab/EwS-Pacbio.git. Additional software code in R and Unix for the analyses are available upon request.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2023.01.017.

Acknowledgments

The full list of the institutions and individuals who supported this work is available in the supplemental acknowledgments.

Author contributions

O.W.L. coordinated the study, performed all genomic and statistical analyses, drafted the manuscript, and designed and generated figures. C.R. performed all functional experiments, analyzed the sequencing data, drafted the manuscript, and designed and generated the figures. S.-H.L. and D.W.B. provided advice for long-read sequencing data analysis and statistical analyses. W.L., K.J., W.Z., E.K., C.L.D., L.B., M. Manning, K.W., M.Y., S. Baulande, and V.R. generated PacBio long-read sequencing data and quality control processes. S.M.K. assisted in generating figures. D.S., S.R., S.Z., S.G., S. Ballet, E.L., G.P., F.T., and R.C. provided biologic input and performed functional experiments. R.A.R., V.L., N.G., N.C., P.M.-B., N.R., W.M.L., A.E.K., J.K., T.M., K.S., T.K., U.D., L.M., M.A.T., G.T.A., S. Bhatia, L.L.R., Y.Y., L.R.-P., W.H., M. Metzler, W.R.D., A.L., N.D.F., R.N.H., and L.M.M. contributed to recruitment of affected individuals and sample collection. S.J.C. supervised the study, provided biological and genetic guidance, and drafted the manuscript. T.G.P.G. contributed to sample collection, and annotated and revised the manuscript. O.D. and M.J.M. initiated, designed, and supervised the study, provided biological and genetic guidance, analyzed the data, drafted the manuscript, and provided laboratory infrastructure and financial support. All authors have reviewed and approved the final version of the manuscript.

Declaration of interests

The authors declare no competing interests.

Received: September 29, 2022
Accepted: January 23, 2023
Published: February 13, 2023

Web resources

apps-scripts, https://github.com/PacificBiosciences/apps-scripts
bioinfo-pf-curie/RNA-seq, https://github.com/bioinfo-pf-curie/RNA-seq
fastqc, https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
GeneOverlap, https://bioconductor.org/packages/release/bioc/html/GeneOverlap.html
ggplot2, https://github.com/tidyverse/ggplot2
gviz, https://bioconductor.org/packages/release/bioc/html/Gviz.html
lima – Demultiplex PacBio Data, https://github.com/PacificBiosciences/barcoding
OMIM, https://www.omim.org/
pbmm2, https://github.com/PacificBiosciences/pbmm2
r, https://www.r-project.org/
RMassBank, https://www.bioconductor.org/packages/release/bioc/html/RMassBank.html
rms, https://cran.r-project.org/web/packages/rms/index.html
stats, https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html

References

1. Esiashvili, N., Goodman, M., and Marcus, R.B., Jr. (2008). Changes in incidence and survival of Ewing sarcoma patients over the past 3 decades: surveillance epidemiology and end results data. J. Pediatr. Hematol. Oncol. *30*, 425–430. https://doi.org/10.1097/MPH.0b013e31816e22f3.
2. 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.
3. Jawad, M.U., Cheung, M.C., Min, E.S., Schneiderbauer, M.M., Koniaris, L.G., and Scully, S.P. (2009). Ewing sarcoma demonstrates racial disparities in incidence-related and sex-related differences in outcome: an analysis of 1631 cases from the SEER database, 1973-2005. Cancer *115*, 3526–3536. https://doi.org/10.1002/cncr.24388.
4. 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.
5. Worch, J., Matthay, K.K., Neuhaus, J., Goldsby, R., and DuBois, S.G. (2010). Ethnic and racial differences in patients with Ewing sarcoma. Cancer *116*, 983–988. https://doi.org/10.1002/cncr.24865.

6. Delattre, O., Zucman, J., Plougastel, B., Desmaze, C., Melot, T., Peter, M., Kovar, H., Joubert, I., de Jong, P., Rouleau, G., et al. (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.
7. Sankar, S., and Lessnick, S.L. (2011). Promiscuous partnerships in Ewing's sarcoma. Cancer Genet. *204*, 351–365. https://doi.org/10.1016/j.cancergen.2011.07.008.
8. Riggi, N., and Stamenkovic, I. (2007). The biology of ewing sarcoma. Cancer Lett. *254*, 1–10. https://doi.org/10.1016/j.canlet.2006.12.009.
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. Toomey, E.C., Schiffman, J.D., and Lessnick, S.L. (2010). Recent advances in the molecular pathogenesis of Ewing's sarcoma. Oncogene *29*, 4504–4516. https://doi.org/10.1038/onc.2010.205.
11. May, W.A., Gishizky, M.L., Lessnick, S.L., Lunsford, L.B., Lewis, B.C., Delattre, O., Zucman, J., Thomas, G., and Denny, C.T. (1993). Ewing sarcoma 11;22 translocation produces a chimeric transcription factor that requires the DNA-binding domain encoded by FLI1 for transformation. Proc. Natl. Acad. Sci. USA *90*, 5752–5756. https://doi.org/10.1073/pnas.90.12.5752.
12. May, W.A., Lessnick, S.L., Braun, B.S., Klemsz, M., Lewis, B.C., Lunsford, L.B., Hromas, R., and Denny, C.T. (1993). The Ewing's sarcoma EWS/FLI-1 fusion gene encodes a more potent transcriptional activator and is a more powerful transforming gene than FLI-1. Mol. Cell Biol. *13*, 7393–7398. https://doi.org/10.1128/mcb.13.12.7393-7398.1993.
13. 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.
14. 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.
15. Johnson, K.M., Mahler, N.R., Saund, R.S., Theisen, E.R., Taslim, C., Callender, N.W., Crow, J.C., Miller, K.R., and Lessnick, S.L. (2017). Role for the EWS domain of EWS/FLI in binding GGAA-microsatellites required for Ewing sarcoma anchorage independent growth. Proc. Natl. Acad. Sci. USA *114*, 9870–9875. https://doi.org/10.1073/pnas.1701872114.
16. Orth, M.F., Surdez, D., Faehling, T., Ehlers, A.C., Marchetto, A., Grossetête, S., Volckmann, R., Zwijnenburg, D.A., Gerke, J.S., Zaidi, S., et al. (2022). Systematic multi-omics cell line profiling uncovers principles of Ewing sarcoma fusion oncogene-mediated gene regulation. Cell Rep. *41*, 111761. https://doi.org/10.1016/j.celrep.2022.111761.
17. Lin, S.H., Sampson, J.N., Grünewald, T.G.P., Surdez, D., Reynaud, S., Mirabeau, O., Karlins, E., Rubio, R.A., Zaidi, S., Grossetête-Lalami, S., et al. (2020). Low-frequency variation near common germline susceptibility loci are associated with risk of Ewing sarcoma. PLoS One *15*, e0237792. https://doi.org/10.1371/journal.pone.0237792.
18. Zhang, J., Walsh, M.F., Wu, G., Edmonson, M.N., Gruber, T.A., Easton, J., Hedges, D., Ma, X., Zhou, X., Yergeau, D.A., et al. (2015). Germline mutations in predisposition genes in pediatric cancer. N. Engl. J. Med. *373*, 2336–2346. https://doi.org/10.1056/NEJMoa1508054.
19. Gröbner, S.N., Worst, B.C., Weischenfeldt, J., Buchhalter, I., Kleinheinz, K., Rudneva, V.A., Johann, P.D., Balasubramanian, G.P., Segura-Wang, M., Brabetz, S., et al. (2018). The landscape of genomic alterations across childhood cancers. Nature *555*, 321–327. https://doi.org/10.1038/nature25480.
20. Machiela, M.J., Grünewald, T.G.P., Surdez, D., Reynaud, S., Mirabeau, O., Karlins, E., Rubio, R.A., Zaidi, S., Grossetete-Lalami, S., Ballet, S., et al. (2018). Genome-wide association study identifies multiple new loci associated with Ewing sarcoma susceptibility. Nat. Commun. *9*, 3184. https://doi.org/10.1038/s41467-018-05537-2.
21. McCarthy, A. (2010). Third generation DNA sequencing: pacific biosciences' single molecule real time technology. Chem. Biol. *17*, 675–676. https://doi.org/10.1016/j.chembiol.2010.07.004.
22. Procedure & Checklist - Preparing SMRTbell Libraries Using PacBio Barcoded Universal Primers for Multiplexing Amplicons. (2020). https://www.pacb.com/wp-content/uploads/Procedure-Checklist-Preparing-SMRTbell-Libraries-using-PacBio-Barcoded-Universal-Primers-for-Multiplexing-Amplicons.pdf.
23. Narasimhan, V., Danecek, P., Scally, A., Xue, Y., Tyler-Smith, C., and Durbin, R. (2016). BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. Bioinformatics *32*, 1749–1751. https://doi.org/10.1093/bioinformatics/btw044.
24. 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.
25. Turner, S.D. (2014). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. Preprint at bioRxiv. https://doi.org/10.1101/005165.
26. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575. https://doi.org/10.1086/519795.
27. Aynaud, M.M., Mirabeau, O., Gruel, N., Grossetête, S., Boeva, V., Durand, S., Surdez, D., Saulnier, O., Zaïdi, S., Gribkova, S., et al. (2020). Transcriptional programs define intratumoral heterogeneity of ewing sarcoma at single-cell resolution. Cell Rep. *30*, 1767–1779.e6. https://doi.org/10.1016/j.celrep.2020.01.049.
28. 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.
29. 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.

30. Servant, N., La Rosa, P., Hupe, P., and Allain, F. (2022). bioinfo-pf-curie/RNA-seq: v3.1.8 (v3.1.8). Zenodo. https://doi.org/10.5281/zenodo.7446922.
31. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.
32. Love, M.I., Anders, S., and Huber, W. (2017). Analyzing RNA-seq data with DESeq2. http://bioconductor.org/packages/devel/bioc/vignettes/DESeq2/inst/doc/DESeq2.html.
33. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.
34. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation *2*, 100141. https://doi.org/10.1016/j.xinn.2021.100141.
35. Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS *16*, 284–287. https://doi.org/10.1089/omi.2011.0118.
36. 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.
37. Mootha, V.K., Lindgren, C.M., Eriksson, K.F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., et al. (2003). PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat. Genet. *34*, 267–273. https://doi.org/10.1038/ng1180.
38. Machiela, M.J., and Chanock, S.J. (2015). LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics *31*, 3555–3557. https://doi.org/10.1093/bioinformatics/btv402.
39. 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.
40. Tirode, F., Laud-Duval, K., Prieur, A., Delorme, B., Charbord, P., and Delattre, O. (2007). Mesenchymal stem cell features of Ewing tumors. Cancer Cell *11*, 421–429. https://doi.org/10.1016/j.ccr.2007.02.027.
41. Deng, Y.N., Xia, Z., Zhang, P., Ejaz, S., and Liang, S. (2020). Transcription Factor RREB1: from Target Genes towards Biological Functions. Int. J. Biol. Sci. *16*, 1463–1473. https://doi.org/10.7150/ijbs.40834.
42. Nitz, M.D., Harding, M.A., Smith, S.C., Thomas, S., and Theodorescu, D. (2011). RREB1 transcription factor splice variants in urologic cancer. Am. J. Pathol. *179*, 477–486. https://doi.org/10.1016/j.ajpath.2011.03.038.
43. Su, J., Morgani, S.M., David, C.J., Wang, Q., Er, E.E., Huang, Y.H., Basnet, H., Zou, Y., Shu, W., Soni, R.K., et al. (2020). TGF-beta orchestrates fibrogenic and developmental EMTs via the RAS effector RREB1. Nature *577*, 566–571. https://doi.org/10.1038/s41586-019-1897-5.
44. Kent, O.A., Saha, M., Coyaud, E., Burston, H.E., Law, N., Dadson, K., Chen, S., Laurent, E.M., St-Germain, J., Sun, R.X., et al. (2020). Haploinsufficiency of RREB1 causes a Noonan-like RASopathy via epigenetic reprogramming of RAS-MAPK pathway genes. Nat. Commun. *11*, 4673. https://doi.org/10.1038/s41467-020-18483-9.
45. Gregory, K.J., Morin, S.M., and Schneider, S.S. (2017). Regulation of early growth response 2 expression by secreted frizzled related protein 1. BMC Cancer *17*, 473. https://doi.org/10.1186/s12885-017-3426-y.
46. Chandhanayingyong, C., Kim, Y., Staples, J.R., Hahn, C., and Lee, F.Y. (2012). MAPK/ERK Signaling in Osteosarcomas, Ewing Sarcomas and Chondrosarcomas: Therapeutic Implications and Future Directions. Sarcoma *2012*, 404810. https://doi.org/10.1155/2012/404810.
47. Jin, W. (2020). The Role of Tyrosine Kinases as a Critical Prognostic Parameter and Its Targeted Therapies in Ewing Sarcoma. Front. Cell Dev. Biol. *8*, 613. https://doi.org/10.3389/fcell.2020.00613.
48. 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.
49. 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.
50. Sánchez-Molina, S., Figuerola-Bou, E., Blanco, E., Sánchez-Jiménez, M., Táboas, P., Gómez, S., Ballaré, C., García-Domínguez, D.J., Prada, E., Hontecillas-Prieto, L., et al. (2020). RING1B recruits EWSR1-FLI1 and cooperates in the remodeling of chromatin necessary for Ewing sarcoma tumorigenesis. Sci. Adv. *6*, eaba3058. https://doi.org/10.1126/sciadv.aba3058.
51. Mobuchon, L., Derrien, A.C., Houy, A., Verrier, T., Pierron, G., Cassoux, N., Milder, M., Deleuze, J.F., Boland, A., Scelo, G., et al. (2022). Different Pigmentation Risk Loci for High-Risk Monosomy 3 and Low-Risk Disomy 3 Uveal Melanomas. J. Natl. Cancer Inst. *114*, 302–309. https://doi.org/10.1093/jnci/djab167.
52. Bigot, P., Colli, L.M., Machiela, M.J., Jessop, L., Myers, T.A., Carrouget, J., Wagner, S., Roberson, D., Eymerit, C., Henrion, D., and Chanock, S.J. (2016). Functional characterization of the 12p12.1 renal cancer-susceptibility locus implicates BHLHE41. Nat. Commun. *7*, 12098. https://doi.org/10.1038/ncomms12098.
53. Landi, M.T., Bishop, D.T., MacGregor, S., Machiela, M.J., Stratigos, A.J., Ghiorzo, P., Brossard, M., Calista, D., Choi, J., Fargnoli, M.C., et al. (2020). Genome-wide association meta-analyses combining multiple risk phenotypes provide insights into the genetic architecture of cutaneous melanoma susceptibility. Nat. Genet. *52*, 494–504. https://doi.org/10.1038/s41588-020-0611-8.