Cooperation of cancer drivers with regulatory germline variants shapes clinical outcomes

J. Musa, F. Cidre-Aranaz, M.-M. Aynaud, M. F. Orth, M. L. Knott, O. Mirabeau, G. Mazor, M. M. Mirabeau, M. L. Gartlgruber, D. Surdez, S. Stein, G. Sannino, J. Li, L. Romero-Pérez, F. Westermann, W. Hartmann, U. Dirksen, M. Gymrek, N. D. Anderson, A. Shlien, B. Rotblat, T. G. P. Grünewald

Pediatric malignancies including Ewing sarcoma (EwS) feature a paucity of somatic alterations except for pathognomonic driver-mutations that cannot explain overt variations in clinical outcome. Here, we demonstrate in EwS how cooperation of dominant oncogenes and regulatory germline variants determine tumor growth, patient survival and drug response. Binding of the oncogenic EWSR1-FLI1 fusion transcription factor to a polymorphic enhancer-like DNA element controls expression of the transcription factor MYBL2 mediating these phenotypes. Whole-genome and RNA sequencing reveals that variability at this locus is inherited via the germline and is associated with variable inter-tumoral MYBL2 expression. High MYBL2 levels sensitize EwS cells for inhibition of its upstream activating kinase CDK2 in vitro and in vivo, suggesting MYBL2 as a putative biomarker for anti-CDK2-therapy. Collectively, we establish cooperation of somatic mutations and regulatory germline variants as a major determinant of tumor progression and highlight the importance of integrating the regulatory genome in precision medicine.

1 Max-Eder Research Group for Pediatric Sarcoma Biology, Institute of Pathology, Faculty of Medicine, LMU Munich, Munich, Germany. 2 INSERM U830, Equipe Labellisée LNCC Genetics and Biology of Pediatric Cancers, PSL Research University, SIREDO Oncology Centre, Institut Curie Research Centre, Paris, France. 3 Institute of Pathology, Faculty of Medicine, LMU Munich, Munich, Germany. 4 Department of Life Sciences, Ben-Gurion University of the Negev, Beer-Sheva, Israel. 5 Neuroblastoma Genomics, German Cancer Research Center (DKFZ), Heidelberg, Germany. 6 Division of Translational Pathology, Gerhard-Domagk Institute of Pathology, University Hospital of Münster, Münster, Germany. 7 Department of Pediatric Hematology and Oncology, University Hospital of Essen, Essen, Germany. 8 Department of Medicine, University of California, San Diego, La Jolla, CA, USA. 9 Department of Computer Science and Engineering, University of California, San Diego, La Jolla, CA, USA. 10 Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, ON, Canada. 11 Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada. 12 Department of Paediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, ON, Canada. 13 German Cancer Consortium (DKTK), Partner site Munich, Munich, Germany. 14 German Cancer Research Center (DKFZ), Heidelberg, Germany. Correspondence and requests for materials should be addressed to T.G.P.G. (email: thomas.gruenewald@med.uni-muenchen.de)
The advent of high-throughput "omics" technologies in oncology enabled assignment of patients to targeted therapies based on somatic mutations in the protein coding genome. However, many childhood cancers including Ewing sarcoma (EwS)—a highly aggressive bone-associated cancer—hardly exhibit any recurrent genetic alteration other than pathognomonic and uniformly expressed driver mutations. Yet, these tumors show substantial inter-individual heterogeneity concerning clinical behavior and treatment response, which cannot be solely explained by their few additional (epi-)genetic alterations.

Recent studies in humans and model organisms suggested that the effects of a dominant oncogene may depend on variations in the regulatory genome. Thus, we hypothesized that oncogenic cooperation of driver-mutations with specific regulatory germline variants may explain inter-individual diversity of clinical outcomes in cancer.

We explore this possibility in EwS, which constitutes a genuine model to study such cooperation for several reasons: First, it is characterized by a simple, nearly diploid genome with a single driver-mutation resulting from chromosomal rearrangements fusing the EWSR1 gene to various members of theETS family of transcription factors. Second, EWSR1-FLI1 steers ~40% of its target genes by binding DNA at GGAA-microsatellites, which are thereby converted into potent enhancers. Third, the enhancer activity of EWSR1-FLI1-bound GGAA-microsatellites strongly depends on the inter-individually variable number of consecutive GGAA-repeats. Together, these characteristics provide an ideal framework to analyze how cooperation of a dominant oncogene (here EWSR1-FLI1) with polymorphic germline regulatory elements (here GGAA-microsatellites) influences the expression of disease-promoting genes that could explain clinical diversity in cancer. In this study, we show, in the EwS model, how such cooperation steers the expression of the functionally and clinically relevant EWSR1-FLI1 target gene MYBL2, thereby determining tumor growth, patient survival, and drug response.

Results

EWSR1-FLI1 regulates MYBL2 via a polymorphic GGAA-microsatellite. To identify candidate genes with high clinical relevance, we crossed two datasets. The first comprised expression microarrays of A673 EwS cells harboring a doxycycline (DOX)-inducible shRNA against EWSR1-FLI1 (A673/TR/shEF1) profiled with/without DOX-treatment (Supplementary Data 1). The second comprised 166 transcriptomes of primary EwS with clinical annotation (Supplementary Data 2). We calculated for each gene represented in both datasets the fold change (FC) of its expression after DOX-induced EWSR1-FLI1 knockdown in A673/TR/shEF1 cells and the significance for association with overall survival (OS) stratifying patients by expression quintiles of the corresponding gene. Specifically, the latter analysis was carried out by a custom software (GenEx) that automatically calculates the P values for each gene in a given overall survival dataset with matched gene expression data by a Mantel–Haenszel test for patients grouped in the highest versus the lowest expression quintile of the given gene (adjusted for multiple comparisons by the Bonferroni method) (see Methods). This analysis identified MYBL2 (alias B-MYB), encoding a central transcription factor regulating cell proliferation, cell survival, and differentiation, as the top EWSR1-FLI1 upregulated gene, whose high expression was significantly associated with poor OS (nominal P = 9.6 × 10^-7, Bonferroni-adjusted P = 0.018) (Fig. 1a, b; Supplementary Data 3).

The EWSR1-FLI1-dependency of MYBL2 expression was validated in time-course experiments in A673/TR/shEF1 on the mRNA and protein level in vitro and in vivo (Fig. 1c, Supplementary Fig. 1a–c), and in nine additional cultured EwS cell lines (Supplementary Fig. 1d).

Despite this tight regulation of MYBL2 by EWSR1-FLI1, we noted a marked inter-tumor heterogeneity of MYBL2 mRNA expression in 166 primary EwS (Supplementary Fig. 1e) and in an independent cohort of 208 EwS on protein level stained for p-MYBL2 (Supplementary Fig. 1f). Interestingly, MYBL2 expression did not correlate with minor variations of EWSR1-FLI1 expression (Supplementary Fig. 1g), suggesting that inter-individual diversity of MYBL2 transcription may be caused differently.

In accordance, re-analysis of published chromatin immunoprecipitation followed by next-generation sequencing (ChIP-seq) data from A673 and SK-N-MC EwS cells revealed strong signals for EWSR1-FLI1 that mapped to a polymorphic GGAA-microsatellite located ~150 kb telomeric of MYBL2 (Fig. 1d). In both cell lines, this GGAA-microsatellite exhibited EWSR1-FLI1-dependent epigenetic characteristics of an active enhancer indicated by H3K4me1 and H3K27ac marks (Fig. 1d).

The EWSR1-FLI1-dependent enhancer activity of this GGAA-microsatellite was confirmed in reporter assays, for which we cloned fragments of ~880 bp from cell line-derived haplotypes differing in their number of consecutive GGAA-repeats (6, 10, or 12 GGAA-repeats) in the pGL3-Fluc vector. Other regulatory variants in the flanking regions were excluded by whole-genome sequencing (WGS) of the parental cell lines and by Sanger sequencing of the cloned fragments (see Methods). In these assays, we observed a positive correlation of the measured enhancer activity and the number of consecutive GGAA-repeats (Fig. 1e).

To test whether EWSR1-FLI1 prefers haplotypes with more consecutive GGAA-repeats, we carried out ChIP-seq analysis using relatively long reads (single-end 150 bp) for EWSR1-FLI1 in the EwS cell line RDES that is heterozygous at the MYBL2-associated GGAA-microsatellite (12 versus 14 consecutive GGAA-repeats). In this analysis, we obtained 31 ChIP-seq reads spanning the entire GGAA-microsatellite. In line with our results from reporter assays (Fig. 1e), 71% of these spanning reads (22/31) mapped to the longer haplotype, whereas only 29% (9/31) mapped to the shorter one (P = 0.015).

Applying the haplotype inference and phasing for short tandem repeats (HapSTR) algorithm on 38 pairs of germline and EwS tumor WGS data covering the MYBL2-associated GGAA-microsatellite, we identified additional haplotypes with 6–17 consecutive GGAA-repeats (average 13.1 GGAA-repeats). Notably, all haplotypes (76/76) were entirely conserved between germline and tumor DNA (Supplementary Data 4).

We next performed expression quantitative trait locus (eQTL) analysis in 35 primary EwS tumors for which matched gene expression and WGS data were available. Prior reports suggested that more than 13 consecutive GGAA-repeats at EWSR1-FLI1 bound GGAA-microsatellites delineate a critical number beyond which very strong EWSR1-FLI1 binding and enhancer activity can be observed, in agreement with our ChIP-seq analysis showing preferential EWSR1-FLI1 binding to the longer haplotype as stated above. Classifying all haplotypes in either “short” (≤13 GGAA-repeats) or “long” (>13 GGAA-repeats), we detected a significantly higher MYBL2 expression in EwS tumors with long/long haplotypes compared to those with short/short haplotypes (Supplementary Data 4, Supplementary Fig. 1h).

We further validated the EWSR1-FLI1-mediated regulation of MYBL2 in time-course EWSR1-FLI1 ChIP-seq and RNA sequencing (RNA-seq) data generated in A673/TR/shEF1 cells. Removal of DOX after suppression of EWSR1-FLI1 for 7 days led to a gradual increase of MYBL2 transcription that correlated with increasing EWSR1-FLI1 recruitment to this GGAA-microsatellite...
MYBL2 is critical for proliferation and cell survival of EwS cells. To obtain first clues on the functional role of MYBL2 in primary EwS, we performed gene-set enrichment analysis (GSEA) of MYBL2 co-expressed genes in 166 EwS tumors. GSEA revealed that MYBL2 co-expressed genes were strongly enriched in human orthologs of known MYBL2 targets in zebrafish and in signatures related to proliferation, cell cycle progression, and sensitization to apoptosis mediated by a CDK-inhibiting protein. Suggesting that MYBL2 may constitute a key downstream mediator of EWSR1-FLI1-induced, evolutionarily conserved proliferation programs.

To test this hypothesis, we performed MYBL2 knockdown experiments in A673, SK-N-MC, and RDES EwS cell lines with moderate to high baseline MYBL2 expression (Supplementary Fig. 2a–c). Using four different siRNAs, we found that MYBL2 silencing reduced proliferation through blockage of G2/M progression, which was accompanied by increased apoptotic cell death (Fig. 2b–d).

To further explore the function of MYBL2 in EwS growth, we generated DOX-inducible anti-MYBL2 shRNA expression systems in A673 and SK-N-MC cells using two different shRNAs. In both cell lines, DOX-induced MYBL2 silencing significantly reduced clonogenic growth in vitro and tumor growth in vivo compared to a non-targeting control shRNA (Fig. 2e–g).

Fig. 1 MYBL2 is a clinically relevant direct EWSR1-FLI1 target gene regulated via a polymorphic GGAA-microsatellite. a Integrative analysis of gene expression microarrays of A673/TR/shEF1 cells profiled with/without DOX addition with 166 clinically annotated EwS transcriptomes; P values determined via Mantel-Haenszel test. The dashed line indicates the Bonferroni-adjusted P value threshold. b Kaplan-Meier survival analysis of 166 EwS patients stratified by quintile MYBL2 expression; P value determined via Mantel-Haenszel test. c Western blot using antibodies against (EWSR1)-FLI1 and MYBL2 in A673/TR/shEF1 cells. EWSR1-FLI1 was silenced for 7 days by DOX-treatment and re-expressed after DOX-removal for 10 days. Loading control: β-actin. d Epigenetic profile of the MYBL2 locus in indicated EwS cells transduced with either a control shRNA (shGFP) or a specific shRNA against EWSR1-FLI1 (shEF1) from published DNase-seq (DNase I hypersensitivity (HS)) data and ChIP-seq data for EWSR1-FLI1, H3K4me1, and H3K27ac. e Reporter assays of MYBL2-associated GGAA-microsatellite (mSat) haplotypes in A673/TR/shEF1 cells treated with/without DOX. Horizontal bars represent means, and whiskers represent the SEM; P values determined via one-tailed Mann-Whitney test. f Analysis of relative MYBL2 expression by qRT-PCR in RDES EwS cells with/without CRISPRi-mediated targeting of the MYBL2-associated GGAA-microsatellite. Horizontal bars represent means, and whiskers represent the SEM; P values determined via two-tailed Mann-Whitney test. Not significant, ns; *P < 0.05. Source data are provided as a Source Data file.
Fig. 2 MYBL2 is critical for proliferation and cell survival of EwS cells in vitro and in vivo. a Upper: Heat-map of genes whose expression is positively or negatively correlated with MYBL2 in 166 primary EwS. Lower: GSEA of the same dataset showing selected gene-sets enriched in MYBL2 high-expressing tumors. b Viable cell count 96 h after transfection of three EwS cell lines with either four different specific siRNAs against MYBL2 (summary of four different siRNAs shown) or a non-targeting siControl. Horizontal bars represent means, and whiskers represent the SEM, n ≥ 3 biologically independent experiments. c Analysis of cell cycle and cell death (sub GI/G0) 96 h after transfection of three EwS cell lines as described in b, using PI staining. Dots show the percentages of cells per experiment delineating cell cycle phases, bars show the fraction of cells (%) in each cell cycle phase, n ≥ 3 biologically independent experiments. d Analysis of apoptosis 96 h after transfection of three EwS cell lines as described in b, using Annexin V/PI staining. Horizontal bars represent means, and whiskers represent the SEM, n = 3 biologically independent experiments. e Relative colony number of A673 and SK-N-MC cells containing either DOX-inducible specific shRNA constructs directed against MYBL2 (shMY_4 refers to shMYBL2_4 and shMY_6 refers to shMYBL2_6) or a non-targeting shControl (shCtr). Cells were grown either with or without DOX. Horizontal bars represent means, and whiskers represent the SEM, n = 3 biologically independent experiments. f, g Kaplan-Meier survival analysis of NSG mice xenografted with A673 or SK-N-MC cells with/without DOX-inducible MYBL2 suppression. Once tumors were palpable, mice were randomized and treated with either vehicle (−) or DOX (+), n ≥ 4 animals per condition. P values determined via Mantel-Haenszel test. h Representative micrographs of xenografts stained with hematoxylin and eosin (HE) or for cleaved caspase 3 (CC3) by IHC. Scale bar is 100 μm. i Quantification of cells arrested in M-phase and automated quantification of the picture area positive for CC3 (relative to control) of micrographs described in h. Horizontal bars represent means, and whiskers represent the SEM, n = 5 samples per condition. ***P < 0.001, **P < 0.01, *P < 0.05; P values determined via two-tailed Mann-Whitney test. Source data are provided as a Source Data file.

Fig. 3 MYBL2 mediates its phenotype via direct upregulation of CCNF, BIRC5, and AURKB. a RNA-seq showing differentially expressed genes (DEGs) after siRNA-mediated MYBL2 knockdown compared to a non-targeting siControl. A summary of three cell lines is shown; n = 3 technical replicates per condition. b GSEA of RNA-seq data. Displayed are 275 gene-sets downregulated upon MYBL2 knockdown that had an FDR q < 0.05. c Analysis of MYBL2 ChIP-seq data from A673 cells showing MYBL2 peaks in the promoters of CCNF, BIRC5, and AURKB. Publicly available EWSR1-FLI1 ChIP-seq data from A673 cells was analyzed to exclude a direct regulation by EWSR1-FLI1. Whole-cell extract (WCE) served as a control. d Linear regression of CCNF, BIRC5, and AURKB expression onto MYBL2 expression in 166 EwS tumors. e Kaplan-Meier survival analyses of 166 EwS patients stratified by median expression levels of the indicated gene; P values determined via Mantel–Haenszel test. f Viable cell count 96 h after transfection of A673 and SK-N-MC cell lines with two different siRNAs directed against either CCNF, BIRC5, or AURKB (summary of two different siRNAs shown) or a non-targeting siControl. Horizontal bars represent means, and whiskers represent the SEM, n ≥ 3 biologically independent experiments. g Measurement of cell death using Trypan blue positivity 96 h after transfection of A673 and SK-N-MC cells transfected as described in f. Horizontal bars represent means, and whiskers represent the SEM, n ≥ 3 biologically independent experiments. ***P < 0.001, **P < 0.01, *P < 0.05; P values determined via two-tailed Mann-Whitney test. Source data are provided as a Source Data file.
We then focused on the 76 most significantly DEGs (mean log2 FC ≥ 1.5, Bonferroni-adjusted P < 0.05), of which representative genes were validated by quantitative real-time PCR (qRT-PCR) (Supplementary Fig. 3b, c, Supplementary Data 7). ChiP-seq analysis using a specific anti-MYBL2 antibody revealed that 50 of these 76 DEGs (66%) showed evidence for MYBL2 promoter-binding (Fig. 3c, Supplementary Data 8). Using microarray data of 166 patient tumors in which 92% of these direct MYBL2 targets were represented enabled correlation of their expression levels with that of MYBL2 and with OS of patients stratified by median expression of the corresponding gene (Supplementary Data 9, 10). Among these genes, CCNE, BIRC5, and AURKB stood out for being highly significantly co-expressed with MYBL2 (Bonferroni-adjusted P < 0.05, rPearson ≥ 0.7) (Fig. 3d), and associated with poor OS (Fig. 3e). To investigate their functional role, we individually knocked down either gene using two specific siRNAs in two different EwS cell lines (Supplementary Fig. 3d) and assessed proliferation and cell viability in vitro. Strikingly, knockdown of these genes broadly phenocopied the anti-proliferative and anti-survival effect of MYBL2 silencing (Fig. 3f, g), suggesting that they may constitute important mediators of the pro-proliferative EWSR1-FLI1/MYBL2 transcriptional program. However, as other functionally relevant genes (e.g. MKI67, KIF20A, PIF1) are also regulated by MYBL2 (Supplementary Fig. 3c), it is conceivable that other genes may contribute to the phenotype of MYBL2.

High MYBL2 levels sensitize EwS cells toward CDK2 inhibition. As there are—to the best of our knowledge—currently no direct MYBL2 inhibitors available, we reasoned that targeting its major upstream cyclin-dependent kinase, CDK2, which activates MYBL2 through phosphorylation, may offer a new therapeutic option for EwS patients with high MYBL2 expression. To test this possibility, we treated EwS cells with two small-molecule CDK2 inhibitors (CVT-313 and NU6140). While both inhibitors strongly reduced growth of A673 EwS cells at the lower micro-molar range, sensitivity toward them was dramatically diminished when MYBL2 was suppressed (Fig. 4a). Such differential effect was not observed in control cells expressing a non-targeting shRNA (Fig. 4a). Notably, NU6140 is a dual inhibitor of CDK2 and the major downstream MYBL2 target AURKB. Since this inhibitor enabled to specifically target EwS cells up- and downstream of MYBL2, we tested its effect on EwS growth in vivo. Treatment of NOD/scid/gamma (NSG) mice with NU6140 significantly (P < 0.05) reduced growth of EwS xenografts compared to vehicle (DMSO) (Fig. 4b), and was accompanied by reduced levels of phosphorylated MYBL2 and increased apoptotic cell death (Fig. 4c). However, this inhibitor had no additional effect on growth of xenografts with silenced MYBL2 expression (Fig. 4b), suggesting that MYBL2 is important for the anti-proliferative effect of CDK2 inhibitors. Consistently, different EwS cell lines with high MYBL2 levels showed higher sensitivity toward NU6140 than a EwS cell line with constitutively low MYBL2 expression (Supplementary Fig. 4a, b). A similar effect on growth of A673 EwS xenografts was observed using the CDK2 inhibitor CVT-313 (Supplementary Fig. 4c, e). Since we neither observed significant weight loss (Supplementary Fig. 4d) nor histomorphological changes in inner organs in mice treated for 14 days with up to 40 mg kg⁻¹ of either inhibitor, these results indicated that CDK2 inhibition can safely impair growth of EwS tumors and that MYBL2 may serve as a biomarker to predict its efficacy.

Interestingly, we observed in A673/TR/shEF1 cells that CDK2 appears to be moderately upregulated by EWSR1-FLI1 (Supplementary Fig. 4f), and found evidence for binding of EWSR1-FLI1 at the CDK2 locus in EwS cells (Supplementary Fig. 4g). However, whether EWSR1-FLI1 regulates CDK2 expression directly or indirectly remains to be elucidated in future studies.

Discussion
Collectively, our discoveries made in an aggressive childhood cancer exemplify how oncogenic cooperation between a cancer driver-mutation (here EWSR1-FLI1) and a regulatory germline variant (here a polymorphic enhancer-like GGAA microsatellite) can create a major source of inter-tumor heterogeneity determining clinical outcome and drug response through modulation of a druggable key downstream player (Fig. 4d).

To explore the possibility of such oncogenic cooperation in EwS beyond MYBL2, we analyzed the top five additional hits of our initial screen whose high expression was associated with worse patient overall survival (EXO1, CIORF112, ESPL1, HJURP, RAD54L; Supplementary Data 3) for the presence of EWSR1-FLI1 bound GGAA-microsatellites or ETS-like binding motifs in the vicinity of these genes, and in that case also possible eQTL effects. While no EWSR1-FLI1 binding was observed at the ESPL1 locus, we found evidence for EWSR1-FLI1 binding at GGAA-microsatellites or ETS-like single GGAA-motifs at the other loci. However, most of these EWSR1-FLI1-binding sites did not show genetic variability in WGS data from primary EwS samples, and if so, they appeared to have no eQTL properties (Supplementary Fig. 5), which may further support the special role of MYBL2 in EwS.

Our results suggest that cooperation between disease-promoting somatic mutations and regulatory germline variants could constitute a general mechanism to explain diversity of disease phenotypes, possibly beyond cancer. In line with this idea, recent reports for neurodegenerative and metabolic diseases showed that the same disease-causing somatic event/mutation can induce distinct phenotypes depending on (inherited) variations in regulatory elements. We anticipate that our findings made in the EwS model are translatable to other malignancies, and propose that integration of the regulatory genome in the process of developing new predictive markers and therapeutic strategies is necessary to refine and fully exploit "omics"-based precision medicine.

Methods
Provenance of cell lines and cell culture conditions. A673 and HEK293T cells were purchased from American Type Culture Collection (ATCC). MH-E-ESI, RDES, RH1, SK-E1, and SK-N-MC cells were provided by the German Collection of Microorganisms and Cell Cultures (DSMZ). TC-32, TC-71, and CHLA-10 cells were kindly provided by the Children’s Oncology Group (COG) and EW1, EW3, EW7, EW16, EW17, EW18, EW22, EW23, LAP35, MIC, ORS, POE, STA-ET1, STA-ET8 cells were provided by O. Delattre (Institute Curie, Paris). A673/TR/shEF1 cells were kindly provided by J. Alonso (Madrid, Spain). The SK-N-MC cell line is listed in the database of commonly misidentified cell lines, ICLAC (http://iclac.org/databases/cross-contaminations), as it was initially described to be a neuroblastoma cell line. Indeed, it is a EwS cell line expressing the pathogenic fusion oncogene EWSR1-FLI1. All cell lines were grown in humidified atmosphere at 37 °C and 5% CO2. Cells were cultured in RPMI 1640 medium supplemented with stable glutamine (Biochrom), 10% tetracycline-free FCS (Sigma-Aldrich), 100 U/ml penicillin (Biochrom), and 100 μg/ml streptomycin (Biochrom). Cells were routinely checked by nested PCR for mycoplasma infection, and their purity was confirmed by STR-profiling and, if applicable, by PCR-based detection of specific fusion onogenes.

DNA/RNA extraction, reverse transcription, and qRT-PCR. DNA was extracted with the Nucleospin Tissue kit (Macherey-Nagel); plasmid DNA was extracted from bacteria with the PureYield kit (Promega). RNA extraction was performed with the Nucleospin RNA kit (Macherey-Nagel) and RNA was reverse transcribed using the High-Capacity cDNA Reverse Transcription kit (Applied Biosystems). qRT-PCRs were performed using SYBR Select Master Mix (Applied Biosystems) and reactions were run on a Bio-Rad CFX Connect instrument and analyzed using the Bio-Rad CFX Manager 3.1 software. All oligonucleotides were purchased from MWG Eurofins Genomics. For primer sequences see Supplementary Data 11.
Transient transfection. For siRNA transfection, cells were seeded in a six-well plate at a density of 1.5 x 10^5 per well in 1.6 ml of growth medium. The cells were transfected with either a negative control non-targeting siRNA (Sigma-Aldrich) or specific siRNAs targeting EWSR1-FLI1 or MYBL2 (both MWG Eurofins Genomics) for 48 h after the first transfection and harvested 96 h after the first transfection. siRNA sequences are given in Supplementary Data 1. For plasmid transfection, cells were seeded in a six-well plate at a density of 2 x 10^5 per well in 1.8 ml of growth medium. Plasmids were transfected with Lipofectamine LTX and Plus Reagent (Invitrogen). The pGL3 vector used for reporter assays has been described before.

Doxycline (DOX)-inducible shRNA constructs. Either a non-targeting negative control shRNA (MWG Eurofins Genomics) or specific shRNAs targeting EWSR1-FLI1 or MYBL2 (both MWG Eurofins Genomics) were cloned in the pLKO-Tet-on-all-in-one system. Oligonucleotide sequences are given in Supplementary Data 1. Lentiviruses were produced in HEK293T cells. A673 and SK-N-MC EwS cells were infected with respective lentiviruses and selected with 1.5 µg ml⁻¹ puromycin (Invitrogen). After single-cell cloning, knockdown efficacy of individual clones was assessed by qRT-PCR 48 h after addition of DOX (1 µg ml⁻¹; Sigma-Aldrich).

DNA constructs and reporter assays. MYBL2-associated GGAA-microsatellites (with ~440 bp 5’ and 3’ flanking regions) from three EwS cell lines were PCR-cloned upstream of the SV40 minimal promoter into the pGL3-Fluc vector (Promega). Primer sequences are given in Supplementary Data 1. The presence of additional variants devoid of the GGAA-microsatellite was ruled out by WGS of the parental cell lines and Sanger sequencing of the cloned fragments. A673/TR/ shEF1 cells (2 x 10^5 per well) were transfected with the Firefly pGL3-Fluc vector containing respective microsatellites and the Renilla pGL3-Rluc vector (Promega) (ratio 100:1) in a six-well plate with 1.8 ml of growth medium. Four hours after transfection, transfection medium was replaced with medium with/without DOX (1 µg ml⁻¹; Sigma-Aldrich). Cells were lysed and assayed with a dual luciferase assay system (Berthold) after 72 h. Firefly luciferase activity was normalized to that of Renilla.

CRISPR interference (CRISPRi) and analysis of cell growth. Due to the lack of functional DNase, CRISPRi does not cause a knockdown of the targeted DNA sequence, but blocks protein binding to it. For the reported experiments, a DNase-dead CAS9 (dCAS9) fused to the KRAB effector domain, which promotes an inhibiting chromatin state, is targeted to the genomic region of interest by specific gRNAs to silence the activity of a given enhancer. To achieve this, we used a pHAGE TRE dCas9-KRAB vector (Addgene #50917) and a pLKO.1-puro U6 shRNA BfuAI large stuffer vector (Addgene #52628), the latter containing either two gRNAs, targeting sequences adjacent to the MYBL2-associated GGAA-microsatellite, or a scrambled control (Supplementary Data 1). Lentivirus production was performed in HEK293T cells. RDES EwS cells were infected with the respective lentiviruses and selected with 1 µg ml⁻¹ puromycin and 1.5 µg ml⁻¹ G418 (both Invitrogen). The cells were induced with DOX (1 µg ml⁻¹; Sigma-Aldrich) for 5 days, after which MYBL2 and EWSR1-FLI1 levels were measured by qRT-PCR.

For measurement of cell growth, cells were grown in medium containing selection antibiotics and DOX (2 µg ml⁻¹) for 14 days as described. Thereafter, 8 x 10⁴ cells/well were plated in quadruplicate wells of 24-well plates in the presence of DOX. After four additional days, cells were washed and fixed with trichloroacetic acid for 1 h at 4°C. Then, plates were washed with phosphate-buffered saline (PBS), air dried, and cells were stained with crystal violet (Sigma-Aldrich) for 30 min. Surplus crystal violet was removed by rinsing the plates with PBS. Cell-bound crystal violet was dissolved in 10% acetic acid, and optical density was measured at 595 nm in a DS-11 spectrophotometer (DeNovix Inc.).
Western blot. Protein from A673/TR/shEFl cells was extracted at d0, d7, d11, d14, and d17 with RIPA and anti-protease cocktail (Roche). Western blots were performed according to standard protocols and probed with antibodies for MYB2 (Sigma-Aldrich, MISSION monoclonal anti-FIL1 antibody: 1:1000, A334385; Abcam) or rabbit polyclonal anti-MYBL2 antibody (1:500, sc-725; Santa Cruz) or mouse monoclonal anti-f-actin (1:10,000, A-5316; Sigma-Aldrich). Anti-rabbit IgG hors eradised peroxidase-coupled antibody (1:3,000; Amer sham Bioscience) was used as secondary antibody. Proteins were visualized using chemiluminescence (Pierce ECL, Western blot chemiluminescent substrate; Thermo Fisher Scientific).

Proliferation assays. Cells were seeded in a six-well plate at a density of 1.5 x 10^5 per well in 1.6 ml of growth medium. The cells were transfected with either negative control siRNA (Sigma-Aldrich) or siRNA targeting MYBL2 (Santa Cruz), harvested after 96 h (including supernatant), fixed in ethanol (70%) at 4 °C, and stained with PI solution (50 µg ml^-1 with 20 µg ml^-1 RNase A (Invitrogen)). Analysis of apoptosis has been performed by combined Annexin V-FITC/PI staining (BD Pharmingen, FITC Annexin V Apoptosis Detection Kit II; BD Biosciences). Cells were trans fected with siRNA targeting MYBL2 (Santa Cruz) and harvested 48 h after transfection (see above) and analyzed by FACS cytometry. Colony number was determined on scanned plates using Fiji (ImageJ)40,41. An example of the gating strategy is given in Supplementary Fig. 6.

Colony-forming assays. A673 and SK-N-MC cells containing either a DOX-inducible negative control siRNA or MYB2-targeting specific siRNAs were seeded in triplicate wells of a 12-well plate at a density of 500 cells (A673) or 1000 cells (SK-N-MC) per well in 2 ml of growth medium. Cells were grown with/without DOX (1 µg ml^-1; Sigma-Aldrich) for 10–14 days depending on the cell line and afterwards stained with crystal violet (Sigma-Aldrich). Colony number was determined on scanned plates using Fiji (ImageJ)40,41. Colony-forming assays were performed for 14 days depending on the cell line.

dNA library and RNA sequencing (RNAseq). A673, SK-N-MC, and RDES EwS cell lines were transfected with triplicates with either a negative control non-targeting siRNA or a specific siRNA targeting MYB2 (siMYB2_1). Total RNA was extracted using the NucleoSpin II kit (Macherey-Nagel). Complementary DNA libraries were sequenced with an Illumina HiSeq2500 instrument using 150 bp paired-end sequencing. Obtained reads were aligned on the human genome (hg19) using Tophat (version 2.1.0-rc5.32). Counting of reads annotated to the genome or the GRCh3737 gene build was done using htcount (v. HTSeq-0.5.3p9)43 with the following parameters: htsect-count -a 10 -q -s no -m union. Sample-to-sample normalization and differential expression analyses were performed using the R package GEO2R (v1.18.0)45. RNA-seq data were deposited at the Gene Expression Omnibus (GEO; accession code GSE119972).

Chromatin immunoprecipitation and sequencing (ChIPseq). DNA–protein cross-linking was done in the presence of 1% paraformaldehyde on 12 x 10^6 A673 or 4 x 10^6 RDES cells, respectively, for each condition for 10 min. Cell lysis, cross-linking was done in the presence of 1% of paraformaldehyde on 12 x 10^6 A673 or 4 x 10^6 RDES cells, respectively, for each condition for 10 min. Cell lysis, and afterwards stained with crystal violet (Sigma-Aldrich). Colony number was determined on scanned plates using Fiji (ImageJ)40,41. An example of the gating strategy is given in Supplementary Fig. 6.

Colony-forming assays. A673 and SK-N-MC cells containing either a DOX-inducible negative control siRNA or MYB2-targeting specific siRNAs were seeded in triplicate wells of a 12-well plate at a density of 500 cells (A673) or 1000 cells (SK-N-MC) per well in 2 ml of growth medium. Cells were grown with/without DOX (1 µg ml^-1; Sigma-Aldrich) for 10–14 days depending on the cell line and afterwards stained with crystal violet (Sigma-Aldrich). Colony number was determined on scanned plates using Fiji (ImageJ)40,41. Colony-forming assays were performed for 14 days depending on the cell line.

dNA library and RNA sequencing (RNAseq). A673, SK-N-MC, and RDES EwS cell lines were transfected with triplicates with either a negative control non-targeting siRNA or a specific siRNA targeting MYB2 (siMYB2_1). Total RNA was extracted using the NucleoSpin II kit (Macherey-Nagel). Complementary DNA libraries were sequenced with an Illumina HiSeq2500 instrument using 150 bp paired-end sequencing. Obtained reads were aligned on the human genome (hg19) using Tophat (version 2.1.0-rc5.32). Counting of reads annotated to the genome or the GRCh3737 gene build was done using htcount (v. HTSeq-0.5.3p9)43 with the following parameters: htsect-count -a 10 -q -s no -m union. Sample-to-sample normalization and differential expression analyses were performed using the R package GEO2R (v1.18.0)45. RNA-seq data were deposited at the Gene Expression Omnibus (GEO; accession code GSE119972).

Chromatin immunoprecipitation and sequencing (ChIPseq). DNA–protein cross-linking was done in the presence of 1% paraformaldehyde on 12 x 10^6 A673 or 4 x 10^6 RDES cells, respectively, for each condition for 10 min. Cell lysis, chromatin shearing, immunoprecipitation, and DNA purification were performed with reagents from iDeal ChIP-seq kit for Transcription Factors (Diagenode, ref: NATURE COMMUNICATIONS | (2019) 10:4128 | https://doi.org/10.1038/s41467-019-12071-2 | www.nature.com/naturecommunications

A673 or 4 x 10^6 RDES cells, respectively, for each condition for 10 min. Cell lysis, chromatin shearing, immunoprecipitation, and DNA purification were performed with reagents from iDeal ChIP-seq kit for Transcription Factors (Diagenode, ref: NATURE COMMUNICATIONS | (2019) 10:4128 | https://doi.org/10.1038/s41467-019-12071-2 | www.nature.com/naturecommunications

A673 or 4 x 10^6 RDES cells, respectively, for each condition for 10 min. Cell lysis, chromatin shearing, immunoprecipitation, and DNA purification were performed with reagents from iDeal ChIP-seq kit for Transcription Factors (Diagenode, ref: NATURE COMMUNICATIONS | (2019) 10:4128 | https://doi.org/10.1038/s41467-019-12071-2 | www.nature.com/naturecommunications

A673 or 4 x 10^6 RDES cells, respectively, for each condition for 10 min. Cell lysis, chromatin shearing, immunoprecipitation, and DNA purification were performed with reagents from iDeal ChIP-seq kit for Transcription Factors (Diagenode, ref: NATURE COMMUNICATIONS | (2019) 10:4128 | https://doi.org/10.1038/s41467-019-12071-2 | www.nature.com/naturecommunications

A673 or 4 x 10^6 RDES cells, respectively, for each condition for 10 min. Cell lysis, chromatin shearing, immunoprecipitation, and DNA purification were performed with reagents from iDeal ChIP-seq kit for Transcription Factors (Diagenode, ref:
non-targeting siControl in A673, SK-N-MC, and RDES EwS cell lines, all genes were ranked by the mean log2 FC and a pre-ranked GSEA was performed with 1000 permutations. GGA-microsatellite analysis using HipSTR. EwS tumors and/or matched blood samples were collected with informed consent from EwS patients treated in the Hospital for Sick Children (SickKids) in Toronto, Canada, in accordance with Research Ethical Board (REB) guidelines (approval no. 1000053452). In addition, publicly available EwS reference samples from the International Cancer Genome Consortium (ICGC) with matched tumor/germline WGS data were used for analysis. WGS was performed in all tumors and available matched germline samples using established protocols on Illumina instruments (paired-end 150/150 bp for the Toronto cohort, and paired-end 100/100 bp for the ICGC cohort). Paired FASTQ files were aligned to the human genome (hg19/GRCh37) using BWA-MEM (v0.7.8). Indel realignment and base quality scores were recalibrated using the Genome Analysis Toolkit (v.2.8.1). For the Toronto cohort, published gene expression data were available from RNA-seq which was deposited at the European Genome-phenome Archive (EGA) under accession number EGAS00001093062; and for the ICGC cohort from matched Affymetrix HG-U133A or HG-U133Plus2.0 gene expression arrays (GSE37371; GSE7707; GSE34620). Affymetrix gene expression data were normalized separately for each chip type by RMA using custom brainarray CDF (v20, ENTREZ)54. Batch effects were removed using Combat55. For eQTL analyses, only tumor samples with a minimum tumor purity of ≥60%, corresponding to TCGA standard tissue requirements (http://cancergenome.nih.gov/cancersselected/biospeccriteria), were used. Tumor purity estimates were made using the AscatNGS (Toronto cohort) or the ESTIMATE algorithm (ICGC cohort)46. To call the genotypes of the MYBBL2-associated GGA-microsatellite, we applied HipSTR (v0.6.2.247) on the WGS data using a minimum threshold of 0.8. All genotype calls of follow-up HipSTR default filters: −min-call-qual 0.9; −max-call-flank-inded 0.15; −max-call-stutter 0.15; −min-call-allele-bias 2; −min-call-strand-bias 2. Human samples and ethics approval. Archived human tissue samples were retrieved from the Institute of Pathology of the LMU Munich (Germany) and the Gerhard-Domagk Institute of Pathology of the University Hospital of Münster (Germany). All patients provided written informed consent. Retrospective and blinded analysis of anonymized samples was carried out upon ethical approval of LMU Munich’s ethics committee (approval no. 550-16 UE).

Tissue microarrays and immunohistochemistry. Formalin-fixed paraffin-embedded samples were collected at the Institute of Pathology of the LMU Munich47. We harvested at least two cores per sample with a core diameter of 1 mm from all blocks to construct tissue microarrays. All EwS samples showed cytogenetic evidence for a translocation of the EWSR1 gene either as determined by fluorescence in situ hybridization and/or qRT-PCR. The samples were reviewed by a reference pathologist. Four-micrometer sections were cut for immunohistochemistry and antigen retrieval was performed with microwave treatment using the anti-human ProTags I Antigen-Enhancer (Quartetti) for p-MYBBL2 or the Target Retrieval Solution (Agilent Technologies) for cleaved caspase 3. In total, 7.5% aqueous H2O2 solution (room temperature) and blocking serum from the corresponding kits were used for 20 min for blockage of endogenous peroxidase. Then slides were incubated for 60 min with the primary antibodies anti-p-MYBBL2 (1:1000; Abcam, ab121591) and anti-cleavage caspase 3 (1:1000, Cell Signaling, #9661). Afterwards slides were incubated with a secondary anti-rabbit IgG antibody (MP-7401, ImmunPrep Reagent Kit, Peroxidase-conjugated) followed by subsequent target detection using DAB® chromatogen (Agilent Technologies).Slides were counterstained with hematoxylin Gill’s Formula (H-3401; Vector). Evaluation of immunoreactivity and quantification of mitoses. Evaluation of p-MYBBL2 immunostaining was carried out semi-quantitatively by a blinded observer in analogy to the Immune Reactive Score (IRS), which is used routinely by pathologists for quantification of hormone receptor expression in mammmary carcinomas, ranging from 0 to 12 as described48. The intensity of p-MYBBL2 immunoreactivity (score 0 = none, score 1 = low, score 2 = intermediate, and score 3 = strong) and the percentage of cells stained with each intensity (score 0 = 0%, score 1 = 0–9%, score 2 = 10–50%, score 3 = 51–80%, and score 4 = 81–100%) was determined per high-power field (x40). The product of the predominant intensity score and its percentage score defined the final IRS. For cleaved caspase 3 immunostaining, automated quantification of the percentage of positive high-power field area was performed using Fiji (ImageJ49). Mitoses were identified in HE-stained slides by a blinded observer per high-power field. Final scores/quantifications were determined by examination of 4–16 high-power fields of at least one section for each sample.

Statistical analysis and software. Statistical data analysis was performed using GraphPad Prism 5 (GraphPad Software Inc., CA, USA) on the raw data. If not otherwise specified in the figure legends comparison of two groups in functional in vitro experiments was carried out using a two-tailed Mann–Whitney test. ***P<0.001, **P<0.01, *P<0.05. Comparison of three groups with data in ordinal scale was performed using Kruskal–Wallis test. If not otherwise specified in the figure legends, data are presented as dot plots with horizontal bars representing means, and whiskers representing the standard error of the mean (SEM). Sample size for all in vitro experiments was chosen empirically. In Kaplan–Meier survival analyses, curves were calculated for all individual survival times of patients or mice, respectively. Curves were compared by Mantel–Haenszel test to detect significant differences between the groups. For in vivo experiments, sample size was pre-determined using power calculation with β = 0.05 and a = 0.05 based on previous preliminary data and in compliance with the 3R system (replacement, reduction, refinement).

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability. RNA-seq and ChIP-seq data have been deposited at the Gene Expression Omnibus ( GEO) under the accession code GSE119972. Microarray data of 166 primary EwS tumors are available from the GEO and displayed in the UCSC genome browser. The following samples were used: GSM1517544 SK-N-MC_shGFP_48h_FLI1; GSM1517553 SK-N-MC_shGFP_48h_FLI1; GSM1517569 SK-N-MC_shGFP_48h_FLI1; GSM1517572 SK-N-MC_shGFP_48h_FLI1; GSM1517575 SK-N-MC_shGFP_48h_FLI1; GSM1517578 SK-N-MC_shGFP_48h_H3K27ac; GSM1517577 SK-N-MC_shGFP_48h_H3K4me1; and GSM1517582 SK-N-MC_shGFP_48h_H3K4me1 (WCE). For gene expression analysis of tumors for which matched germline/tumor WGS was available, published gene expression data from the Toronto cohort was available from RNA-seq data which was deposited at the European Genome-phenome Archive (EGA) under accession number EGA500001003062(ref. 12); and for the ICGC cohort from matched Affymetrix HG-U133A or HG-U133Plus2.0 gene expression arrays (GSE37371; GSE7707(ref. 49); GSE63157(ref. 59)); the source data underlying Figs. 1a–e, 2a–f, 3a–b, 4a–c, and Supplementary Figs. 1a–b, 1d–j, 2a–d, 3a–d, 4a–g, and 5a–e are provided as a Source Data file. All the other data supporting the findings of this study are available in the article and its supplementary information files and from the corresponding author upon reasonable request. A reporting summary for this article is available as a Supplementary Information file.

Code availability. Source code is available from the corresponding author upon reasonable request.

Received: 15 February 2019 Accepted: 16 August 2019
Published online: 11 September 2019

References. 1. Gartway, L. A., Verweij, J. & Ballman, K. V. Precision oncology: an overview. J. Clin. Oncol. J. Am. Soc. Clin. Oncol. 31, 1803–1805 (2013). 2. Gröbner, S. N. et al. The landscape of genomic alterations across childhood cancers. Nature 555, 321–327 (2018). 3. Grünewald, T. G. P. et al. Ewing sarcoma. Nat. Rev. Dis. Prim. 4, 5 (2018). 4. Brohl, A. S. et al. The genomic landscape of the Ewing Sarcoma family of tumors reveals recurrent STAG2 mutations. PLoS Genet. 10, e1004475 (2014). 5. Crompton, B. D. et al. The genomic landscape of pediatric Ewing sarcoma. Cancer Discov. 4, 1326–1341 (2014). 6. Tirode, F. et al. Genomic landscape of Ewing sarcoma defines an aggressive subtype with co-association of STAG2 and TP53 mutations. Cancer Discov. 4, 1342–1353 (2014). 7. Vu, V. et al. Natural variation in gene expression modulates the severity of mutant phenotypes. Cell 162, 391–402 (2015). 8. Stoehr, C. G. et al. Homozygous G/G variant of SNP309 in the human MDM2 gene is associated with earlier tumor onset in Caucasian female renal cell carcinoma patients. Oncogenes 5, e205 (2016). 9. herein. A. M. et al. foci of morbidly diseased tissues in chronic lymphocytic leukemia: results of an individual patient data-based meta-analysis. Haematologica 99, 1285–1291 (2014). 10. Zeron-Medina, J. et al. A polymorphic p53 response element in KIT ligand influences cancer risk and has undergone natural selection. Cell 155, 410–422 (2013). 11. Post, S. M. et al. A high-frequency regulatory polymorphism in the p53 pathway accelerates tumor development. Cancer Cell 18, 220–230 (2010). 12. Anderson, N. D. et al. Rearrangement bursts generate canonical gene fusions in bone and soft tissue tumors. Science 361, https://doi.org/10.1126/science. aab4819 (2018).
Author contributions
J.M. coordinated the study, performed most functional in vitro and in vivo experiments, analyzed the data, wrote the paper, and designed the figures. F.-C.A. carried out functional in vitro and in vivo experiments. M.-M.A. and D.S. carried out RNA- and ChIP-seq experiments. M.F.O. helped in functional experiments, bioinformatic analyses, and lentiviral transduction of cell lines. G.M., M.V., and B.R. carried out CRISPRi experiments. S.G. and O.M. helped in analysis of the RNA- and ChIP-seq data, and provided genetic and statistical guidance. S.O. supported animal experiments. J.S.G. helped in microarray data processing and bioinformatic analyses. A.M., M.D. M.C.B, M.G., F.W., S.S. T.L.B.H., M.M.L.K, G.S., J.L., and L.R.P. assisted in experimental procedures. W.H. and U.D. provided tissue microarrays. M.G., N.D.A., and A.S. provided patient sequencing data and carried out HipSTR-analyses. T.K. and O.D. provided biological and genetic guidance, laboratory infrastructure, and financial support. J.M. and T.G.P.G. interpreted and analyzed all data. T.G.P.G. initiated, designed and supervised the study, provided biological and genetic guidance, analyzed the data, wrote the paper together with J.M., and provided laboratory infrastructure and financial support. All authors read and approved the final manuscript.

Additional information
Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-019-12071-2.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Peer review information Nature Communications thanks Peter Scacheri, Janet Shipley and other anonymous reviewer(s) for their contribution to the peer review of this work.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2019