From the Institute of Pathology

of the Ludwig Maximilian University Munich

Director Prof. Dr. med. Thomas Kirchner

Dissertation

Zum Erwerb des Doctors of Philosophy (Ph.D.)

an der Medizinischen Fakultät der

Ludwig-Maximilians-Universität München

Systematic multi-omics profiling of Ewing sarcoma cell lines

Vorgelegt von

Dr. med. Martin Franz Orth, M.Sc.

aus Ruppertsberg

München, 2021

First supervisor:	PD Dr. Dr. med. Thomas Grünewald
Second supervisor:	Prof. Dr. rer. nat. Andreas Jung
Dean:	Prof. Dr. med. dent. Reinhard Hickel
Day of oral defense:	22.03.2021

for Dr.-Ing. Lukas Sebastian Kaul

Affidavit

Orth, Martin Franz

aus Ruppertsberg

I hereby declare, that the submitted thesis entitled

“Systematic multi-omics profiling of

Ewing sarcoma cell lines”

Is my own work. I have only used the sources indicated and have not made unauthorised use of services of a third party. Where the work of others has been quoted or reproduced, the source is always given.

I further declare that the submitted thesis or parts thereof have not been presented as part of an examination degree to any other university.

Ruppertsberg, 30.03.2021 Martin Orth

Confirmation of congruency between printed and electronic version of the doctoral thesis

Orth, Martin Franz

aus Ruppertsberg

I hereby declare that the electronic version of the submitted thesis, entitled

“Systematic multi-omics profiling of

Ewing sarcoma cell lines”

Is congruent with the printed version both in content and format.

Ruppertsberg, 30.03.2021 Martin Orth

Table of contents

Summary

100 years after its first description, Ewing sarcoma (EwS), the second most common bone-associated cancer in children and young adults, is still poorly understood. Neither the cell of origin is known, nor the detailed mechanism of expression regulation by the pathognomic fusion oncogene. Similarly, factors causing overt clinical heterogeneity and advanced/targeted therapeutic strategies for patients with non-localized disease remain to be identified.

An apparent paradox of EwS is its clinical heterogeneity compared to its silent landscape of genomic mutations. The only highly recurrent mutation in EwS is the characteristic fusion oncogene composed of *EWSR1* and an *ETS*-transcription factor. Interactions of this single driver with the genome have been described and associated with gene expression regulation several times, but always in a small number of cell line models. This thesis aimed at creating a multidimensional dataset on a large number of EwS cell line models with and without fusion oncogene knockdown, the Ewing Sarcoma Cell Line Atlas (ESCLA), to both enable further investigations of expression regulation in EwS and model heterogeneity.

In 18 well-characterized EwS cell lines, with three distinct fusion types, an inducible shRNA construct targeting the fusion oncogene was stably integrated. The whole genomes of the cell lines were sequenced with relatively long reads (150 bp) and >30 coverage. For the respective fusion and the histone marks H3K27ac, H3K27me3 and H3K4me3 chromatin immunoprecipitation with subsequent next-generation sequencing (ChIP-Seq) was performed. The transcriptome of the cells with and without fusion knockdown was assessed by ClariomD DNA microarrays, as was the protein expression by mass spectrometry and the CpG island methylation by MethylationEPIC BeadChip arrays.

Whole genome sequencing enabled genotyping of several polymorphic potentially fusion binding microsatellites with GGAA motif. ChIP-Seq data were in line with previous publications and identified 50 additional consensus fusion binding sites. Transcriptome and proteome data strongly correlated with each other and displayed expression rearrangement upon fusion knockdown. Only for CpG methylation not any uniform effect of fusion oncogene knockdown was observed.

Cell lines with distinct fusion types, *EWSR1-FLI1* type 1, 2 and *EWSR1-ERG*, were for the first time systematically compared to each other. Neither expression regulation, nor methylation profile were dependent on the respective fusion. However, the fusion types differed in their rate of chromoplexy as developmental process. All *EWSR1-ERG* fusions and

55% of *EWSR1-FLI1* type 1 fusions developed from chromoplexy, whereas all *EWSR1-FLI1* type 2 fusions were the result of reciprocal translocation.

Binding of the fusion to GGAA motifs appeared as multifactorial and still poorly understood process. Among others, high numbers of consecutive GGAA motifs, additional nearby motifs and microsatellites as well as and copy number gains correlated with fusion binding probability. Genes differentially expressed upon fusion knockdown differed from not affected genes in their distance to the next fusion bound GGAA mSat, the number of nearby GGAA mSats, and in the presence of transcription factor bindings sites for NFAT5, NFYC, and E2F2 in their promoters. All these transcription factors were also regulated by the fusion oncogene.

A set of 22 genes were identified to be regulated to different extends in the 18 cell line models upon fusion knockdown. This heterogeneity in regulation was in line with heterogeneous expression in patients, which correlated with overall survival. These genes were mainly associated with cell-cycle progression and cell division, transcription factors and targets of those. Yet, evaluated and identified parameters of *EWSR1-ETS* mediated gene expression regulation were not sufficient to fully explain inter-cell line differences in gene regulation.

Several studies demonstrated previously an interaction between the fusion oncogene and GGAA microsatellites, but were limited to few loci. Previous whole exome sequencing projects missed out on these relevant regulatory regions. Reporter assays *in vitro* revealed enhancer activity of GGAA microsatellites, but in an artificial only mono-allelic approach. Studies and experiments on gene regulation in EwS with only two to three cell lines could hardly model heterogeneity. The here generated ESCLA overcame these obstacles, and supported, refined and expanded previously elaborated models of fusion oncogene mediated gene regulation genome wide.

In conclusion, a multidimensional and comprehensive dataset was generated on a collection of EwS cell line models clearly outnumbering previous studies. Moreover, the dataset has already enabled first novel insights on the mechanisms and dependencies of fusion mediated gene regulation and modelled heterogeneity. The generated cell lines and the ESCLA likely constitute a rich resource for the Ewing sarcoma research community. Additionally, the capability of the dataset to model heterogeneity might enforce research on personalized medicine and the development of new treatment strategies for so far incurable advanced disease patients.

1. Introduction

1.1. General and clinical aspects of Ewing sarcoma

Ewing sarcoma (EwS) is an aggressive bone and soft-tissue cancer (Grünewald et al., 2018). EwS mainly affects children and young adults with an incidence of 0.155 per 100,000 in Caucasians, thus constituting the second most common bone-associated sarcoma at younger ages after osteosarcoma (Kaatsch et al., 2016). Moreover, EwS occurs rarely at higher ages (Jawad et al., 2009). Males are more frequently affected than females (1.5 times) (Jawad et al., 2009). The most common EwS bone cancer sites are the lower extremities and the pelvis (Grünewald et al., 2018). Soft-tissue EwS occurs more frequent in elder patients, rather located in the trunk than in the extremities (Applebaum et al., 2011). First symptoms are intermittent local and regional pain, palpable swelling and only rarely fever and pathological fracture (Grünewald et al., 2018; PDQ Pediatric Treatment Editorial Board, 2002).

After assessment of symptoms, physical examination and radiography, EwS is diagnosed (molecular-)pathologically (Grünewald et al., 2018; PDQ Pediatric Treatment Editorial Board, 2002). However, diagnosing EwS is still challenging (Orth et al., 2020). Although EwS was first described by its eponym James Ewing in 1920 referring to its histology as "diffuse endothelioma of bone", a round cell sarcoma of unknown origin with small polyhedral cells, hyperchromatic nuclei and pale cytoplasm without inter-cellular stroma (Ewing, 2006), the histology (Figure 1) is not unique to EwS. Several other cancers share the small-round-cell histology, even differential diagnoses like neuroblastoma and rhabdomyosarcoma (Carter and Patel, 2019; Hung et al., 2017; Sbaraglia et al., 2020). The standard immunohistochemical marker for EwS is CD99 (Ambros et al., 1991). However, it has been shown several times that CD99 is highly sensitive for EwS, but also unspecific (Baldauf et al., 2018a; Orth et al., 2020; Zaccarini et al., 2018). Other proposed auxiliary markers like PAX7 (Baldauf et al., 2018b; Charville et al., 2017) and NKX2-2 (Hung et al., 2016) lack specificity. Friend leukemia integration 1 transcription factor, FLI1, also lacks specificity and additional sensitivity as it is rarely expressed in those EwS, which are not positive for *EWSR1-FLI1* (Crompton et al., 2014; Rossi et al., 2004). The recently introduced immunohistochemical markers BCL11B and GLG1 are highly specific, but capture only 63% of EwS (Baldauf et al., 2018a; Orth et al., 2020). Thus, further diagnostic procedures are based on the detection of the pathognomonic *EWSR1-ETS* fusion. The standard tool is fluorescence in situ hybridization for *EWSR1* breakapart (Machado et al., 2009). Nevertheless, *EWSR1* breakapart occurs also in morphological mimics of EwS like desmoplastic small round cell

tumor (DSRCT) and angiomatoid fibrous histiocytoma (Papp et al., 2017). Elsewise, the fusion can be detected via quantitative real-time PCR, whose reliability is highly dependent on the seleted primers, or RNASeq (Antonescu, 2014).

Standard treatment comprises neoadjuvant chemotherapy, often combinations with/of cyclophosphamide, doxorubicin and etoposide, followed by radiotherapy and/or resection (Gaspar et al., 2015; PDQ Pediatric Treatment Editorial Board, 2002). The further treatment should be adapted to histological response (Bosma et al., 2019). Prognosis is moderate in case of metastasis-free disease with a five-year survival rate of about 70%. Unfortunately, in about 25% of cases metastases have already formed at the time of diagnosis, leading to a by far worse survival rate (30%) (Bosma et al., 2018; Gaspar et al., 2015; Kridis et al., 2017). Noteworthy, the chemotherapeutics administered in young age and during development bear risks for permanent and late occurring adverse effects (Ginsberg et al., 2010; Longhi et al., 2012). Unfortunately, more advanced and specific treatments were not established in standard clinical protocols. However, several candidate drugs, like clofarabine (Çelik et al., 2018), and potential targets were described, for instance inhibition of PARP1 (Brenner et al., 2012), PRKCB (Surdez et al., 2012), HDAC enzymes (Pattenden et al., 2016), and CDK12 (Iniguez et al., 2018).

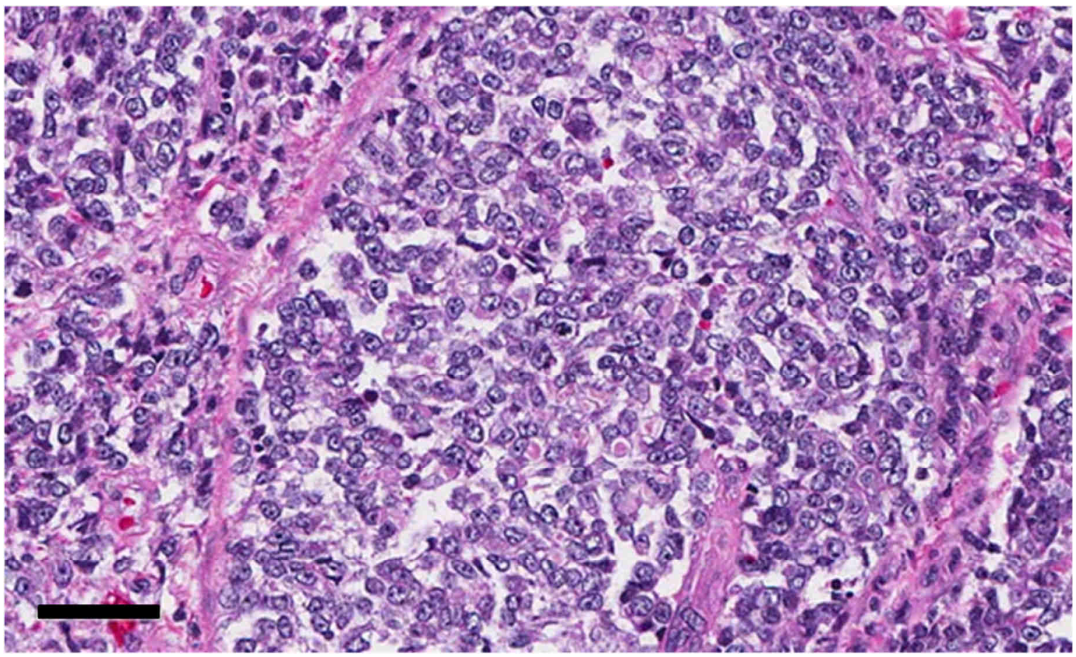

Figure 1: Histology of EwS. HE staining; bar indicates 50 µm. From the Atlas of Genetics and Cytogenetics in Oncology and Haematology (URL http://AtlasGeneticsOncology.org; PMID 23161685).

The cell of origin is still, even 100 years after the first description of EwS, a matter of debate. Analyses on the morphology, histological markers, transcriptome and epigenome, as well as the site of tumor growth suggest that EwS develops from mesenchymal stem cells (Gordon et al., 2016; Sheffield et al., 2017; Tirode et al., 2007). However, very recently this assumption was questioned again in a report favoring pluripotent stem and neural crest cells as origin of EwS (Miller et al., 2020).

1.2. Genomic aspects of EwS

Genetically, EwS harbors a pathognomic aberrant fusion transcription factor. This fusion transcription factor results from in-frame translocation of the N-terminal part of the constitutively transcribed *EWS RNA-binding protein 1* (*EWSR1*, chr22q12.2) to the C-terminal part of a member of the *ETS* transcription factor family. In rare cases, *EWSR1* is replaced by *FUS*, another member of the FET protein family (Chen et al., 2016; Tsuda et al., 2020). The most common fusion partner for *EWSR1* is the *ETS* transcription factor *FLI1* (chr11q24.3; 85%) (Delattre et al., 1992; Sorensen et al., 1994). *EWSR1-FLI1* fusions (*EF1*) are differentiated by the last *EWSR1* and first *FLI1* exon in the fusion, most frequently *EWSR1* exon 7 and *FLI1* exon 6 (type 1) or *EWSR1* exon 7 and *FLI1* exon 5 (type 2) (Giovannini et al., 1994). Fusions to *ETS-related gene* (*ERG*; chr21q22.2), *EWSR1-ERG* (*EErg*) constitute another 10% of EwS fusion oncogenes (Sorensen et al., 1994; Zucman et al., 1993). Already at the time of the first *EErg* fusion description, it has been stated that this fusion cannot be the product of a simple balanced translocation, as the fusion partners are localized on opposite strands (Desmaze et al., 1997; Zucman et al., 1993). Recently, chromoplexy, a process involving multiple rearrangement loops over several chromosomes, was described as the developmental process for 42% of EwS fusion, especially for *EErg* (Anderson et al., 2018; Baca et al., 2013). Another rather seldom mutational process resulting in EwS fusion is chromothripsis (Anderson et al., 2018; Stephens et al., 2011).

Despite the fusion oncogenes, EwS presents a mainly unaltered genome. Copy number gains of chromosome 1q (18%), 8 (47%) and 12 (21%) are recurrently observed, as is loss of chromosome 16q (17%) (Mackintosh et al., 2012; Tirode et al., 2014). *CDKN2A* deletions are also common in EwS (12%). *TP53* mutations occur in about 7%, *STAG2* mutations in about 17%. Mutations in the coding sequence of *STAG2* are actually the most frequent somatic mutations in EwS despite the fusion oncogene. *STAG2*, and even more coincidence with *TP53* mutations are associated with poor outcome (Tirode et al., 2014).

Knockdown experiments of *EF1* resulted in impaired proliferation and invasion, and higher rates of apoptosis (Chansky et al., 2004; Patel et al., 2012). Several further knockdown experiments for the EwS fusions were performed, mainly to identify a multitude of regulated and potential target genes (Smith et al., 2006; Tirode et al., 2007). Expression of the fusion in other cell types than EwS leads to various phenotypes indicating dependency of the cellular context for the action of *EF1* (Kovar, 2005). For instance, while NIH3T3 cells were transformed (Arvand and Denny, 2001; May et al., 1993), mesenchymal stem cells from bone marrow were blocked in differentiation upon *EF1* expression (Torchia et al., 2003). Expression of *EF1* in human embryonic stem cells resulted in a similar transcriptomic profile to EwS (Gordon et al., 2016).

Therefore, these studies and the absence of other highly recurrent somatic mutations indicate that the fusion oncogene is the major driver of EwS. Differences in the biology of the distinct fusions were not extensively investigated. But a study on two type 1 and one type 2 cell line and one type 3 (*EWSR1* exon 10 to *FLI1* exon 6) found only 41 genes to be differentially expressed (Bandrés et al., 2005). Different fusion types were discussed to have impact on prognosis (de Alava et al., 1998). However, this hypothesis was later rejected (Le Deley et al., 2010).

1.3. Gene regulation in EwS

Contrary to the genome, the epigenome (Sheffield et al., 2017), transcriptome, and proteome are massively rewired in EwS compared to normal tissues, likely by the fusion oncogene. Previous studies could demonstrate that EF1 binds to microsatellites (mSats) with GGAA motif in ChIP experiments, resulting in enhancer activity (Gangwal et al., 2008; Guillon et al., 2009; Riggi et al., 2014; Wei et al., 2010). A mSat is defined as the repetitive sequential occurrence of motifs in the genome, the motif is two to six bases long (dependent on authors up to 10), and the minimal number of repeats is two to five (Richard et al., 2008). The repetitive motifs affect the DNA structure which leads to slippage during replication and finally to a high variability in mSat length (Mirkin, 2007). GGAA mSats are packed in nucleosomes, not transcribed and without any known function in normal tissues. Actually, mSats were for a long time believed to be genetic junk (Gangwal et al., 2008). However, there is growing evidence that mSats have impact on gene expression and constitute for 15% of all inherited gene regulation by nearby common genomic variants (Fotsing et al., 2019; Gymrek et al., 2016). The accessibility of GGAA mSats for EF1 binding and the presence of histone marks for open chromatin is actually dependent on EF1 (Patel et al.,

2012; Riggi et al., 2014; Tomazou et al., 2015). Knockdown of EF1 leads to a chromatin state more similar to mesenchymal stem cells. Vice versa, expression of *EF1* in mesenchymal stem cells resulted in a chromatin state more similar to EwS (Riggi et al., 2014). Moreover, in embryonic and multipotent stem cells, mSats are often likewise accessible as in EwS, indicating a permissive milieu for EwS fusion and, again, stem cells as presumable origin of EwS (Gomez et al., 2016). An EF1 dependent enhancer activity of GGAA mSats has been shown in several reporter assays (Dallmayer et al., 2019; Marchetto et al., 2020; Monument et al., 2014; Musa et al., 2019). Partly, GGAA mSats bound by the fusion oncogene depict even super-enhancer properties (Baldauf et al., 2018a), which are described to mediate cell identity (Whyte et al., 2013). Interestingly, the number of consecutive GGAA repeats correlates with the enhancer activity of the respective mSat (Gangwal et al., 2008). The minimal number of GGAA repeats for EF1 dependent enhancer activity is controversial. A minimum of four repeats for binding and five for enhancer activity (Gangwal et al., 2008) were described in reporter assays, in ChIP data four repeats were already sufficient for enhancer activity (Riggi et al., 2014). The highest enhancer activity was reported for 24 repeats, followed by a dip in enhancer activity before another peak at around 50 repeats (Monument et al., 2014).

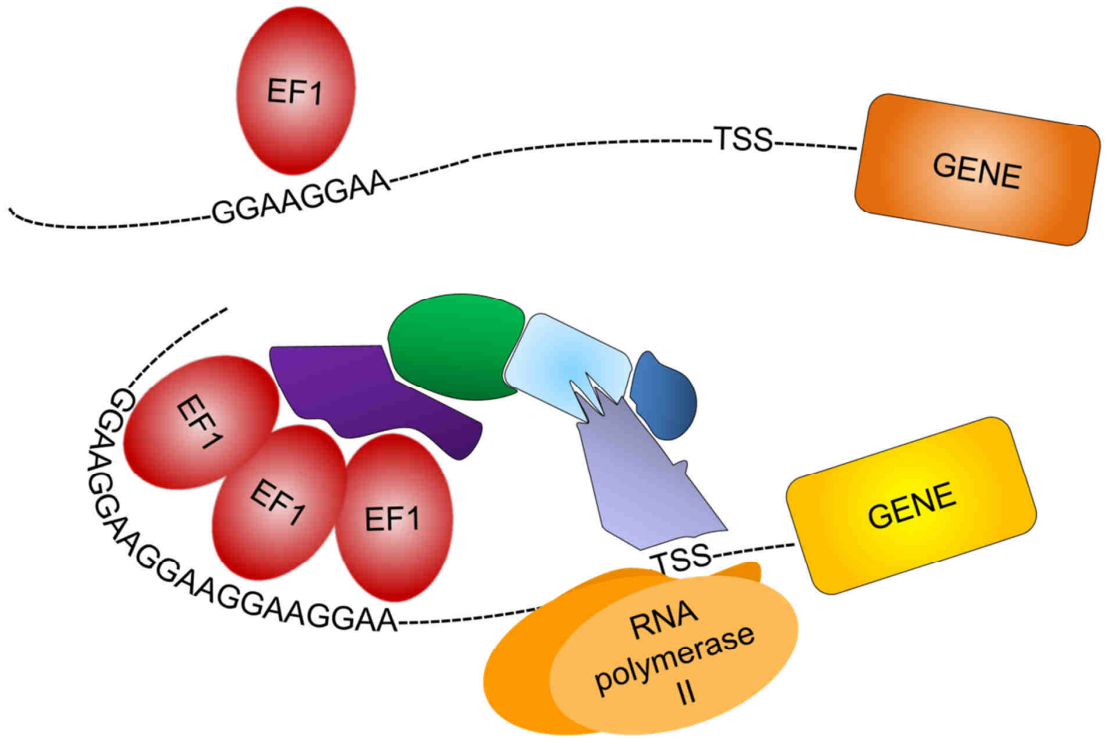

Figure 2: The current model of EWSR1-FLI1 mediated gene expression.

The mechanism, how EWSR1-ETS (EEts) confers *de novo* enhancer activity to otherwise nucleosomal GGAA mSats is not fully elucidated. Binding of EF1 to the genome is dependent on the presence of the DNA binding domain of the *ETS* transcription factor (May et al., 1993; Welford et al., 2001). However, the fused transcription factor lacks one of normally two *p300* binding domains, a transcriptional coactivator. Thus, monomeric EF1, which has been shown to bind to single GGAA motifs (Guillon et al., 2009; Riggi et al., 2014), recruits far less of this transcriptional coactivator (Riggi et al., 2014). Hence, one explanation for the EEts mediated enhancer activity at GGAA mSats might be that simply more EEts binds. Nevertheless, EWSR1 is highly important for the EEts mediated enhancer activity, and several interaction partners were described that might close the gap between the known fusion binding to GGAA mSats and the consequent gene regulation (Figure 2). EWSR1 is hypothesized to co-activate ETS transcription factors (Kedage et al., 2016). Additionally, EWSR1 is important for the open chromatin structure at GGAA mSats. A prion-like domain in EWSR1 binds the BAF chromatin remodelling complex thus recruiting BAF to GGAA mSats and open the chromatin structure (Boulay et al., 2017). Currently, an interaction of EWSR1 and RING1B, another interactor with p300, has been shown to be necessary for target location of the fusion (Sánchez-Molina et al., 2020). Furthermore, EWSR1 colocalizes with LSD1, whose expression is important for EEts mediated enhancer function (Theisen et al., 2020). Interestingly, fusion-EWSR1 even interacts with wildtype EWSR1, thus interfering with its regulatory activity, ultimately increasing transcription (Gorthi et al., 2018).

1.4. The relevance of gene regulation by the EwS-specific fusion oncogene

The first clinical relevance of *EF1* binding to a specific GGAA mSat has been shown in 2015 in the context of EwS susceptibility. Susceptibility highly varies between ethnicities. Africans depict a susceptibility by factor 9-10 less than Europeans (Jawad et al., 2009). In 2012, higher length of the *NR0B1* associated mSat in Africans was reported. However, this finding was not sufficient to explain lower susceptibility as *NR0B1* has been shown to promote EwS (Beck et al., 2012). Finally, a high diversity of the single nucleotide polymorphism (SNP) rs79965208 (T to A) has been shown, with a frequency of 64% in Caucasians, but only 25% in Africans. The A allele instead of a T creates a GGAA motif connecting two adjacent mSats, thus generating one long GGAA mSat nearby the susceptibility gene *EGR2* (Grünewald et al., 2015). Despite *EGR2* expression as susceptibility marker, six SNPs associated with susceptibility were identified in genome-wide association studies, most of

them closely located to GGAA repeats (Machiela et al., 2018). In 2019, the relevance of the repeat number of an EEts-bound GGAA mSat for EwS prognosis was reported as a correlation with the prognostically unfavorable *MYBL2* expression (Musa et al., 2019). Noteworthy, MYBL2 acts likely via AURKB. The same gene has been reported to mediate the enhancer activity of RING1B/EF1 complexes (Musa et al., 2019; Sánchez-Molina et al., 2020)

To further understand the biology of EwS, find prognostic markers, and inroads for new therapeutic approaches, several targets of the fusion oncogene were identified and evaluated for their functional relevance. *NR0B1* was the first identified direct target of an *EF1*-bound GGAA mSat that is necessary for cell line growth (Gangwal et al., 2008; Kinsey et al., 2006). Similar results were published for *CCK* (Carrillo et al., 2007), *FOXM1* (Christensen et al., 2013) and *NKX2-2*, which is closely located to one of the six susceptibility SNPs (Smith et al., 2006). Additionally, knockdown of *PRKCB* impaired many genes affected by *EF1*, hinting towards *PRKCB* as a major downstream effector of *EF1* (Surdez et al., 2012). The expression of *STEAP1* was associated with oxidative stress in EwS which promotes aggressiveness (Grunewald et al., 2012). Another *EEts* driven gene upregulating oxidative stress, *SOX6*, was recently identified (Marchetto et al., 2020). One of the few investigated *EF1* repressed targets is *SPRY1*. Reexpression of *SPRY1* reduces EwS cell proliferation as it inhibits the fibroblast growth factor receptor (Cidre-Aranaz et al., 2017). Another gene suppressed in EwS is LOX, which acts as tumor suppressor in EwS (Agra et al., 2013).

In sum, previous reports on EwS identified the pathognomonic *EEts* fusion as the major driver of EwS, which affects the expression of several genes by binding to GGAA mSats, thereby converting otherwise closed chromatin structure to *de novo* enhancers. Besides the quiet genomic landscape, EwS is clinically heterogeneous, partly depicting favorable and partly fatal prognosis. Little is known about the heterogeneity causing factor. Current studies focus on different fusion expression levels resulting in different EwS behavior from a more proliferative to more invasive state (Aynaud et al., 2020; Franzetti et al., 2017). Thus, further insights into *EEts* mediated gene regulation are urgently needed to better understand clinical heterogeneity and find inroads to advanced and personalized therapeutic strategies. For several reasons, previous studies were restricted to further clarifying EwS heterogeneity. Only a limited number of cell lines was tested (most often two), which cannot be sufficient to represent heterogeneity in patients. Moreover, most previous studies were based exclusively on *EF1* positive EwS, not addressing potential commonalities or differences between fusion

types. Furthermore, analyses on *EEts* mediated enhancer activity were mainly performed in a mono-allelic approach.

To overcome these obstacles, this thesis aimed at the generation of a multi-dimensional and comprehensive dataset of a relatively high number of EwS cell line models in fusion-high and -low state to assess commonalities between cell lines with distinct fusions, further investigate gene regulation genome wide, and to identify heterogeneously regulated clinically relevant genes as potential inroads for personalized medicine.

Three distinct fusion types were represented by 18 cell lines. All cell lines were transduced with an inducible shRNA targeting the respective fusion. The cells were characterized on the genome, transcriptome, proteome and methylome level. Furthermore, fusion binding to the genome and histone marks were assessed. These components constituted the Ewing Sarcoma Cell Line Atlas (ESCLA).

The ESCLA enabled first insights into several factors influencing fusion binding to GGAA mSats and expression regulation, confirming previously reported targets and dependencies and extending those with new identified targets, binding sites, and structure-effect associations. 22 heterogeneously regulated and survival correlated genes were identified, of which most were associated with cell-cycle progression.

2. Research objectives and scientific aims

2.1. Research objectives

EwS lacks highly recurrent mutations except for the pathognomonic chimeric *EWSR1-ETS* transcription factor. Hence, the reason for clinically overt inter-patient heterogeneity might not be mutations, but rather germline variations in polymorphic EWSR1-ETS binding sites. This thesis focused on deciphering the genetic architecture of preferred EWSR1-ETS binding sites and parameters affecting the binding. The results should further elucidate gene regulatory mechanisms in EwS and hint towards heterogeneously regulated clinically relevant genes as potential targets for personalized medicine. The analyses were based on the construction of a relatively large cohort of EwS cell lines and a comprehensive multidimensional omics-dataset on these cells, which will likely serve as rich resource for EwS researchers.

2.2. Scientific aims

1st aim: Generation of EwS cell lines with inducible knockdown of the fusion oncogene.

2nd aim: Generation of a multidimensional and comprehensive dataset on the EwS cell lines including data on genome, fusion binding to the genome, transcriptome, proteome and methylome.

3rd aim: Comparison of EwS cell lines with distinct fusion types based on the dataset.

4th aim: Demonstration and further investigation of EwS fusion binding and gene expression regulation in the dataset.

5th aim: Identification of differentially expressed and clinically relevant genes as potential biomarkers for personalized medicine.

3. Materials and Methods

3.1. Materials

3.1.1. Cell lines

Article	*Specification*	*Supplier*
A673	EwS cell line	ATCC, Manassas, VA, USA
CHLA10	EwS cell line	Children's Oncology Group, Monrovia, CA, USA
CHLA25	EwS cell line	Children's Oncology Group, Monrovia, CA, USA
CHLA99	EwS cell line	Children's Oncology Group, Monrovia, CA, USA
COGE352	EwS cell line	Children's Oncology Group, Monrovia, CA, USA
ES7	EwS cell line	Delattre, O., Paris, France
EW1	EwS cell line	Delattre, O., Paris, France
EW16	EwS cell line	Delattre, O., Paris, France
EW17	EwS cell line	Delattre, O., Paris, France
EW18	EwS cell line	Delattre, O., Paris, France
EW22	EwS cell line	Delattre, O., Paris, France
EW24	EwS cell line	Delattre, O., Paris, France
EW3	EwS cell line	Delattre, O., Paris, France
EW7	EwS cell line	Delattre, O., Paris, France
LAP35	EwS cell line	Delattre, O., Paris, France
MHHES1	EwS cell line	DSMZ, Braunschweig, Germany
MIC	EwS cell line	Delattre, O., Paris, France
ORS	EwS cell line	Delattre, O., Paris, France
POE	EwS cell line	Delattre, O., Paris, France
RDES	EwS cell line	DSMZ, Braunschweig, Germany
RH1	EwS cell line	Delattre, O., Paris, France
SBKMSKS1	EwS cell line	Burdach, S., Munich, Germany (Grunewald et al., 2012)
SKES1	EwS cell line	DSMZ, Braunschweig, Germany
SKNMC	EwS cell line	DSMZ, Braunschweig, Germany
SKNPLI	EwS cell line	Delattre, O., Paris, France
SKPNDW	EwS cell line	Delattre, O., Paris, France
TC106	EwS cell line	Children's Oncology Group, Monrovia, CA, USA
TC205	EwS cell line	Children's Oncology Group, Monrovia, CA, USA
TC32	EwS cell line	Children's Oncology Group, Monrovia, CA, USA
TC71	EwS cell line	Children's Oncology Group, Monrovia, CA, USA

HEK293T	Human embryonic kidney cell line	DSMZ, Braunschweig, Germany
SAOS2	Osteosarcoma cell line	DSMZ, Braunschweig, Germany
U2OS	Osteosarcoma cell line	DSMZ, Braunschweig, Germany

3.1.2. Mice

Article	***Supplier***
NOD/scid/gamma mice	Charles River, Wilmington, MA, USA

3.1.3. Bacteria

Article	***Supplier***
Stellar Competent Cells	Clontech, TaKaRa, Saint-Germain-en-Laye, France

3.1.4. Media and supplements for cell culture

Article	***Manufacturer***
Accutase	Sigma-Aldrich, Taufkirchen, Germany
Collagen solution from bovine skin	Sigma-Aldrich, Taufkirchen, Germany
Doxycycline	Merck, Darmstadt, Germany
Doxycycline (for mice)	Belapharm, Vechta, Germany
Fetal bovine serum	Sigma-Aldrich, Taufkirchen, Germany
OptiMEM medium	Thermo Fisher, Waltham, MA, USA
Penicilin / Streptomycin (10.000 U/ml, 10.000µg/ml)	Biochrom, Berlin, Germany
Phosphate Buffered Saline	Biochrom, Berlin, Germany
Plasmocure	InvivoGen, San Diego, CA, USA
Puromycin	InvivoGen, San Diego, CA, USA
RPMI 1640 medium with stable glutamin, 2 g/L NaHCO3	Biochrom, Berlin, Germany
Trypsin/EDTA (10X)	Biochrom, Berlin, Germany

3.1.5. siRNAs, shRNAs, vectors, and transfection reagents

Article	***Description/Target sequence***	***Manufacturer***
siRNA_EF1_TypeII_1	5'-CAGAGTTCACTGCTGGCCTAT-3'	eurofins Genomics, Ebersberg, Germany
siRNA_EF1_TypeII_2	5'-AGCAGAGTTCACTGCTGGCCT-3'	eurofins Genomics, Ebersberg, Germany

shRNA_Control	5'-CAACAAGATGAAGAGCACCAA-3'	eurofins Genomics, Ebersberg, Germany
shRNA_EF1_Carillo	5'-GCAGCAGAACCCTTCTTATGA-3'	eurofins Genomics, Ebersberg, Germany
shRNA_EF1_Tirode	5'-AAGGCAGCAGAACCCTTCTTA-3'	eurofins Genomics, Ebersberg, Germany
shRNA_EWSR1-ERG	5'-GCTACGGGCAGCAGAATTTAC-3'	eurofins Genomics, Ebersberg, Germany
shRNA_FLI1_Ex5	5'-AGTTCACTGCTGGCCTATAAT-3'	eurofins Genomics, Ebersberg, Germany
shRNA_FLI1_Ex9	5'-CTTTGGAGCCGCATCACAATA-3'	eurofins Genomics, Ebersberg, Germany
pCMV_EF1_puro	Vector	Surdez, D., Paris, France
Tet-pLKO-puro	Vector	Addgene, Watertown, MA, USA
D8.9	Packaging plasmid	Creative Biogene, Shirley, NY, USA
VSVG	Packaging plasmid	Addgene, Watertown, MA, USA
HiPerFect	Transfection reagent	Qiagen, Venlo, Netherlands
Lipofectamin LTX and Plus Reagent	Transfection reagent	Thermo Fisher, Waltham, MA, USA

3.1.6. Primers

All primers were retrieved from Eurofin Genomics, Ebersberg, Germany.

Name/target	***Sequence***	***Purpose***
CCND1-CTCF_Fw	5'-GGTGGGAGGTCTTTTTGTTTC-3'	ChIP-PCR
CCND1-CTCF_Rv	5'-CACGCAATCCCAGATCAAAAC-3'	ChIP-PCR
CCND1-EF1_Fw	5'-CACAGTGTGGGTATTTCCATCAAGCA-3'	ChIP-PCR
CCND1-EF1_Rv	5'-GGTGTGTAGGAAAAACAGCTCTCTGGA-3'	ChIP-PCR
CDKN1A_Fw	5'-ACTGACTCATCACTACTCCCTC-3'	ChIP-PCR
CDKN1A_Rv	5'-GTGTGCTATTCCCGCCAG-3'	ChIP-PCR
CUL1_Fw	5'-TCCTCCCTTCTAGAAAGAGCTGAC-3'	ChIP-PCR
CUL1_Rv	5'-AGGCCCGAAACCACAGAGCATAAA-3'	ChIP-PCR
GATA2 _Fw	5'-CTCAGGACCCATGGAAGTATTG-3'	ChIP-PCR
GATA2_Rv	5'-CTGCAATCCTCTTAGCCTCTAG-3'	ChIP-PCR
IGF2_Fw	5'-AGTGCTCGGAATGTTTGGGAACTG-3'	ChIP-PCR
IGF2_Rv	5'-AGTTACCAGGAGGTGCTCAAGTGT-3'	ChIP-PCR

Sec14L2_Fw	5'-GCCCCCGCTGATGCACTTCC-3'	ChIP-PCR
Sec14L2_Rv	5'-AAGTGCGCCAGCAGAGCCAG-3'	ChIP-PCR
Myco-F1	5'-ACACCATGGGAGCTGGTAAT-3'	Mycoplasma PCR
Myco-F1t	5'-ACACCATGGGAGTTGGTAAT-3'	Mycoplasma PCR
Myco-F2	5'-GTTCTTTGAAAACTGAAT-3'	Mycoplasma PCR
Myco-F2a	5'-ATTCTTTGAAAACTGAAT-3'	Mycoplasma PCR
Myco-F2cc	5'-GCTCTTTCAAAACTGAAT-3'	Mycoplasma PCR
Myco-R1	5'-CTTCATCGACTTTCAGACCCAAGGCAT-3'	Mycoplasma PCR
Myco-R1ac	5'-CTTCATCGACTTCCAGACCCAAGGCAT-3'	Mycoplasma PCR
Myco-R1cat	5'-CCTCATCGACTTTCAGACCCAAGGCAT-3'	Mycoplasma PCR
Myco-R1tt	5'-CTTCTTCGACTTTCAGACCCAAGGCAT-3'	Mycoplasma PCR
Myco-R2	5'-GCATCCACCAAAAACTCT-3'	Mycoplasma PCR
Myco-R2at	5'-GCATCCACCAAATACTCT-3'	Mycoplasma PCR
Myco-R2ca	5'-GCATCCACCACAAACTCT-3'	Mycoplasma PCR
EF1_Fw	5'-GCCAAGCTCCAAGTCAATATAGC-3'	qRT-PCR, PCR (fusion oncogene transcript)
EF1_Rv	5'-GAGGCCAGAATTCATGTTATTGC-3'	qRT-PCR, PCR (fusion oncogene transcript)
ERG_Fw	5'-CGAACGAGCGCAGAGTTAT-3'	qRT-PCR (wt gene)
ERG_Rv	5'-ACGTCTGGAAGGCCATATTC-3'	qRT-PCR (wt gene)
EWSR1_Fw	5'-CAGCCAAGCTCCAAGTCAATA-3'	qRT-PCR (wt gene)
EWSR1_Rv	5'-TCCAGACTCCTGCCCATAAA-3'	qRT-PCR (wt gene)
EWSR1-ERG_Fw	5'-TCCAAGTCAATATAGCCAACAGAG-3'	qRT-PCR
EWSR1-ERG_Rv	5'-CTGTGGAAGGAGATGGTTGAG-3'	qRT-PCR
FLI1_TV1-3_Fw	5'-TGGATGGCAAGGAACTGTG-3'	qRT-PCR (wt gene)
FLI1_TV1-3_Rv	5'-CGGTGTGGGAGGTTGTATTA-3'	qRT-PCR (wt gene)
FLI1_TV4_Fw	5'-CATCACCATCCCACCGTC-3'	qRT-PCR (wt gene)
FLI1_TV4_Rv	5'-TCTGAAACATCGTGGGTAGC-3'	qRT-PCR (wt gene)
RPLP0_Fw	5'-GAAACTCTGCATTCTCGCTTC-3'	qRT-PCR
RPLP0_Rv	5'-GGTGTAATCCGTCTCCACAG-3'	qRT-PCR
ERG_Rv2	5'-TTGGGTTTGCTCTTCCGCTC-3'	PCR (fusion oncogene transcript)
FLI1-EWS#1_Fw	5'-AATACAACCTCCCACACCGA-3'	PCR, reverse fusion
FLI1-EWS#1_Rv	5'-ACTCCTGCCCATAAACACCC-3'	PCR, reverse fusion
FLI1-EWS#2_Fw	5'-GTGCTGTTGTCACACCTCAG-3'	PCR, reverse fusion
FLI1-EWS#2_Rv	5'-GTTCTCTCCTGGTCCGGAAA-3'	PCR, reverse fusion
FLI1-FLI1_Fw	5'-AATACAACCTCCACACCGA-3'	PCR, reverse fusion
FLI1-FLI1_Rv	5'-CTTACTGATCGTTTGTGCCCC-3'	PCR, reverse fusion

Tet-pLKO_Fw	5'-GGCAGGGATATTCACCATTAT-3'	colony PCR
Tet-pLKO_Rv	5'-CTATTCTTTCCCCTGCACTG-3'	colony PCR
Tet-pLKO-Seq_Fw	5'-GGCAGGGATATTCACCATTATCGTTTCAGA-3'	Sequencing
Tet-pLKO-Seq_Rv	5'-GACGTGAAGAATGTGCGAGA-3'	Sequencing

3.1.7. Antibodies

Target	***Species, clonality***	***Application***	***Product number***	***Manufacturer***
ERG (EPR3864)	Rabbit, monoclonal	ChIP	ab92513	Abcam, Cambridge, UK
FLI1	Rabbit, polyclonal	ChIP	ab15289	Abcam, Cambridge, UK
H3K27ac	Rabbit, polyclonal	ChIP	ab4729	Abcam, Cambridge, UK
H3K27me3	Rabbit, polyclonal	ChIP	C15410069	Diagenode, Seraing, Belgium
H3K4me3	Rabbit, polyclonal	ChIP	C15410003	Diagenode, Seraing, Belgium
Mouse/rabbit IgG (Vectastain Elite ABC HRP Kit)	Horse, polyclonal	IHC	PK-6200	Vector Laboratories, Burlingame, CA, USA
PAX7	Mouse, monclonal	IHC	PAX7-c	DSHB, Iowa City, IA, USA
pMYBL2 (phospho T487; EPR2204Y)	Rabbit, monoclonal	IHC	ab76009	Abcam, Cambridge, UK
Rabbit IgG (ImmPRESS® HRP Polymer Detection Kit)	Horse, polyclonal	IHC	MP-7401	Vector Laboratories, Burlingame, CA, USA
SOX6	Rabbit, polyclonal	IHC	HPA003908	Atlas Antibodies, Stockholm, Sweden
ERG (EP111)	Rabbit, monoclonal	Western blot, primary	434R	Cell Marque, Rocklin, USA
GAPDH (6C4)	Mouse, monclonal	Western blot, primary	sc-32233	Santa Cruz Biotechnology, Dallas, TX, USA
FLI1 (MRQ-1)	Mouse, monclonal	Western blot, primary; IHC	254M	Cell Marque, Rocklin, USA
Mouse IgG (H+L), HRP	Goat, polyclonal	Western blot, secondary	W402B	Promega, Maddison, WI, USA
Rabbit IgG (H+L) HRP	Goat, polyclonal	Western blot, secondary	R1364HRP	OriGene Technologies, Rockville, MD, USA

3.1.8. Kits

Article	*Manufacturer*
Antigen retrieval AR-10 solution	DCS Innovative, Hamburg, Germany
High-Capacity cDNA Reverse Transcription Kit	Thermo Fisher, Waltham, MA, USA
iDeal ChIP-seq kit for Transcription Factors	Diagenode, Seraing, Belgium
NucleoSpin Gel and PCR Clean-up	Macherey-Nagel, Düren, Germany
NucleoSpin RNA	Macherey-Nagel, Düren, Germany
NucleoSpin Tissue genomic DNA prep kit	Macherey-Nagel, Düren, Germany
ProTaqs I and V Antigen-Enhancer	Quartett, Berlin, Germany
PureYield Plasmid Midiprep System	Promega, Maddison, WI, USA
Qubit dsDNA HS Assay Kit	Thermo Fisher, Waltham, MA, USA
SYBR SELECT Master Mix	Thermo Fisher, Waltham, MA, USA
Target Retrieval Solution S1699	Agilent Technologies, Waldbronn, Germany

3.1.9. Enzymes

Article	*Manufacturer*
AgeI HF	New England Biolabs, Frankfurt (Main), Germany
Complete Protease Inhibitor Mini	Roche, Mannheim, Germany
EcoRI HF	New England Biolabs, Frankfurt (Main), Germany
GoTaq Hot Start Polymerase	Promega, Maddison, WI, USA
Proteinase Inhibitor Cocktail	Sigma-Aldrich, Taufkirchen, Germany
RNAse cocktail	Thermo Fisher, Waltham, MA, USA
T4 DNA ligase	Thermo Fisher, Waltham, MA, USA
Taq DNA polymerase	New England Biolabs, Frankfurt (Main), Germany

3.1.10. Chemicals and reagents

Article	*Manufacturer/supplier*
Acetic acid (100%)	Neolab, Heidelberg, Germany
Acrylamid, Rotiphorese Gel 30 Bisacrylamid	Roth, Karlsruhe, Germany
AEC+ chromogen	Agilent Technologies, Santa Clara, CA, USA
Agarose	Sigma-Aldrich, Taufkirchen, Germany
Ammonium persulfate (APS)	Sigma-Aldrich, Taufkirchen, Germany
Ampicillin	Sigma-Aldrich, Taufkirchen, Germany
Aqua bidest	Kerndl, Weißenfeld, Germany
Bradford	Bio-Rad, Munich, Germany

Bromphenol blue	Sigma-Aldrich, Taufkirchen, Germany
Buffer for T4 ligase (10X)	Thermo Fisher, Waltham, MA, USA
Buffer GoTag Green	Promega, Maddison, WI, USA
Buffer Smart Cut	New England Biolabs, Frankfurt (Main), Germany
Chloroform/isoamylalcohol 24/1	Sigma-Aldrich, Taufkirchen, Germany
Crystal violet	Sigma-Aldrich, Taufkirchen, Germany
DAB+ chromogen	Agilent Technologies, Santa Clara, CA, USA
Dimethyl sulfoxide (DMSO)	Sigma-Aldrich, Taufkirchen, Germany
Dithiothreitol (DTT)	Sigma-Aldrich, Taufkirchen, Germany
DNA Ladder 1kbp	Roth, Karlsruhe, Germany
DNA Ladder GeneRuler 100 bp Plus	Thermo Fisher, Waltham, MA, USA
dNTPs	Sigma-Aldrich, Taufkirchen, Germany
Ethanol (denaturated >99.8%)	Roth, Karlsruhe, Germany
Ethidiumbromide solution	Sigma-Aldrich, Taufkirchen, Germany
Ethylenediaminetetraacetic acid (EDTA)	Roth, Karlsruhe, Germany
Formalin (4%)	Sigma-Aldrich, Taufkirchen, Germany
Gelatin from procine skin, powder	Sigma-Aldrich, Taufkirchen, Germany
Geltrex LDEV-Free Reduced Growth Factor Basement Membrane Matrix	Thermo Fisher, Waltham, MA, USA
Glycerol	Roth, Karlsruhe, Germany
Glycine	Roth, Karlsruhe, Germany
Hematoxylin Gill's Formula	Vector Laboratories, Burlingame, CA, USA
HRP Substrate Immobilion	Merck, Darmstadt, Germany
Isopropanol	Sigma-Aldrich, Taufkirchen, Germany
LB Broth (Miller)	Sigma-Aldrich, Taufkirchen, Germany
LB Broth Agar (Lennox)	Sigma-Aldrich, Taufkirchen, Germany
Magnesium chloride (MgCl2)	New England Biolabs, Frankfurt (Main), Germany
Methanol	Sigma-Aldrich, Taufkirchen, Germany
Milk powder, non-fat dried	Roth, Karlsruhe, Germany
Nonidet P40	Thermo Fisher, Waltham, MA, USA
Phenol/chloroform/isoamylalcohol 25/24/1	Thermo Fisher, Waltham, MA, USA
PageRuler Prestained Protein Ladder	Thermo Fisher, Waltham, MA, USA
RNAseZap	Sigma-Aldrich, Taufkirchen, Germany
Sodium acetate buffer solution 3M	Sigma-Aldrich, Taufkirchen, Germany
Sodium chloride (NaCl)	Roth, Karlsruhe, Germany
Sodium deoxycholate	Sigma-Aldrich, Taufkirchen, Germany
Sodium dodecyl sufate (SDS)	Sigma-Aldrich, Taufkirchen, Germany

Sodium orthovanadate (Na3VO4)	Sigma-Aldrich, Taufkirchen, Germany
Sucrose	Sigma-Aldrich, Taufkirchen, Germany
SYBR Safe DNA gel stain	Thermo Fisher, Waltham, USA
Tetramethylethylenediamine (TEMED)	Roth, Karlsruhe, Germany
TRIS	Roth, Karlsruhe, Germany
TRIS HCl	Sigma-Aldrich, Taufkirchen, Germany
Triton X-100	Sigma-Aldrich, Taufkirchen, Germany
Trypan blue solution	Sigma-Aldrich, Taufkirchen, Germany
Tween-20	Sigma-Aldrich, Taufkirchen, Germany
Water, DNAse und RNAse free	Roth, Karlsruhe, Germany

3.1.11. Consumeables

Type	***Specification***	***Manufacturer/supplier***
Autoclavation tape		Hartenstein, Würzburg, Germany
Cell culture flasks	T25, T75, T150	TPP, Trasadingen, Switzerland
Cell culture multiwell plates	96, 24, 12, 6 wells	TPP, Trasadingen, Switzerland
Cell scraper		Hartenstein, Würzburg, Germany
Centrifugation tubes	15 ml, 50 ml	Greiner Bio-One, Kremsmünster, Austria
Counting chambers	C-Chip	NanoEntek, Waltham, MA, USA
Counting chambers	for Countess cell counter	Thermo Fisher, Waltham, MA, USA
Cryo tubes	2 ml, CryoGen CLEARLine	Hartenstein, Würzburg, Germany
Filters for syringes	Rotilabo 0.45µm, 0.22µm	Roth, Karlsruhe, Germany
Gloves	Nitril NextGen	MeiTrade, Krimpen aan den Ijseel, Netherlands
	Purple Nitrile	Kimberly-Clark Professional, Koblenz-Rheinhafen, Germany
Injection needles	size 18, Sterican	Braun, Melsungen, Germany
Multi-dispenser tips	0.5 ml, 5 ml, sterile	Hartenstein, Würzburg, Germany
Nitrocellulose	Amersham Protran 0.45	GE Healthcare, Freiburg, Germany
Parafilm	M laboratory film	Hartenstein, Würzburg, Germany
Pasteur pipettes		Kimble Chase, Meiningen, Germany
PCR plates	FrameStar 480/96 clear, white	4titude, Wotton, UK
PCR Tubes	8 Well PCR Tube Strip, attached caps	4titude, Wotton, UK
Petri dishes	10 cm	Greiner Bio-One, Kremsmünster, Austria
Pipette tips	10 µl, 20 µl, 100 µl, 200 µl, 1250	Biozym, Hessisch Oldendorf,

	µl, SafeSeal SurPhop filter tips	Germany
Protective Coat S2	SplashCoat pro+	DACH, Raststatt, Germany
Reaction tubes	1.5 ml, 2 ml	Eppendorf, Hamburg, Germany
	1.5 ml, DNA lobind	Eppendorf, Hamburg, Germany
	1.5 ml, protein lobind	Eppendorf, Hamburg, Germany
Scalpel	Fig.21, ratiomed	Hartenstein, Würzburg, Germany
Serological pipettes	5 ml, 10 ml, 25 ml	Corning, Wiesbaden, Germany
Sonication tubes	1.5 ml TPX microtubes for Bioruptor Plus	Diagenode, Seraing, Belgium
Syringes	5 ml, Inject	Braun, Melsungen, Germany
Whatman paper		Hartenstein, Würzburg, Germany

3.1.12. Instruments

Type	***Specification***	***Manufacturer/supplier***
Aspiration pump	Vacusafe	Integra, Biebertal, Germany
	table device FTA-1	Hartenstein, Würzburg, Germany
Autoclave	Varioklav	HP Labortechnik, Oberschleißheim, Germany
Bacteria incubator	Kelvitron T	Heraeus, Hanau, Germany
Bacteria shaker	Certomat IS	Braun, Melsungen, Germany
Camera	µ850SW	Olympus, Tokio, Japan
Cell counter	Countess	Thermo Fisher, Waltham, MA, USA
Centrifuge	Rotina 420R	Hettich, Tuttlingen, Germany
	cooling Centrifuge5417R	Eppendorf, Hamburg, Germany
	Heraeus Megafuge8R	Thermo Fisher, Waltham, MA, USA
	for PCR plates Universal320	Hettich, Tuttlingen, Germany
Fluorometer	Qubit 3.0	Thermo Fisher, Waltham, MA, USA
Freezer -80°C	B 35-50	Fryka, Esslingen, Germany
	Froma -86°C ULT Freezer	Thermo Fisher, Waltham, MA, USA
Freezing container	Mr. Frosty	Hartenstein, Würzburg, Germany
Gel chambers		PEQLAB, Erlangen, Germany
Heating circulator	EO	Julabo, Seelbach, Deutschland
Ice machine	SPR 80	Nordcap, Bremen, Germany
Incubator	Forma Scientific CO2 Water Jacketed Incubator	Thermo Fisher, Waltham, MA, USA
Laminar Flow	HeraSafe Laminar Flow	Thermo Fisher, Waltham, MA, USA
	Safe2000	Thermo Fisher, Waltham, MA, USA

	Maxisafe2020	Thermo Fisher, Waltham, MA, USA
Microscope	Axiovert 200	Carl Zeiss, Oberkochen, Germany
	Axioplan2	Carl Zeiss, Oberkochen, Germany
	Primo Vert	Carl Zeiss, Oberkochen, Germany
Mini centrifuge		Labnet, Edison, NJ, USA
	DW-41	Qualiton, Korea
Multidispenser	HandyStep electronig	Brand, Wertheim, Germany
Nitrogen tank	Arpege 70	Air Liquide, Düsseldorf, Germany
Orbital shaker	Unimax 1010 DT	Heidolph Instruments, Schwabach, Germany
PCR Cycler	vapo.protect	Eppendorf, Hamburg, Germany
	T100 Thermal Cycler	Bio-Rad, Munich, Germany
pH meter	WTW pH-197	Xylem Analytics, Weilheim, Germany
Pipette Controller	Pipetboy2	Integra, Biebertal, Germany
	accu-jet pro	Brand, Wertheim, Germany
Pipettes	Pipetman 2, 10, 20, 100, 200, 1000 µl	Gilson, Middleton, WI, USA
	PipetmanG 2, 10, 20, 100, 200, 1000 µl	Gilson, Middleton, WI, USA
Plate reader	Varioskan	Thermo Fisher, Waltham, MA, USA
Power supply	PowerPac 200	Bio-Rad, Munich, Germany
qPCR Cycler	CFX Connect Real-Time System	Bio-Rad, Munich, Germany
Rotation wheel	SB3	Stuart, Staffordshire, UK
Scale	BL150	Sartorius, Göttingen, Germany
	Analytical Lab Balance	Sartorius, Göttingen, Germany
Sonicator	Bioruptor Plus	Diagenode, Seraing, Belgium
	VCX 130	Sonics, Newtown, CT, USA
Spectrophotometer	NanoDrop ND-1000 Spectophotometer	PEQLAB, Erlangen, Germany
Thermoblock	ThermoStat plus	Eppendorf, Hamburg, Germany
Thermomixer	Thermomixer compact	Eppendorf, Hamburg, Germany
Ultracentrifuge	4K15C	Sigma-Aldrich, Taufkirchen, Germany
UV working table	FT-20E/365	Thermo Fisher, Waltham, MA, USA
	MultiImage Light Cabinet	Alpha Innotech, Kasendorf, Germany
Vacuum manifold		Promega, Maddison, WI, USA
Vacuum pump	WOB-L	Welch, Fürstenfeldbruck, Germany
Vortexer	Vortex-Genie2	Scientific Industries, Bohemia, NY, USA
Water bath		Köttermann, Uetze, Germany
Western blot	Vertical Electrophoresis Cell	Thermo Fisher, Waltham, MA, USA

electrophoresis chamber		
Western blot imaging system	Odyssey	LI-COR, Homburg, Germany
Western blot transfer system	Mini Trans-Blot Cell	Thermo Fisher, Waltham, MA, USA

3.1.13. Software

Specific bioinformatic packages are cited in the respective Methods section.

Program	***Distributor***
Bio-Rad CFX Manager	Bio-Rad, Munich, Germany
GraphPad PRISM 8	GraphPad Software, San Diego, CA, USA
Microsoft Office 2010	Microsoft, Redmond, WA, USA
Transcriptome Analysis Console 4.0	Thermo Fisher, Waltham, MA, USA
Zotero 5.0	from www.zotero.org/download

3.2. Methods

3.2.1. Cell culture standards

3.2.1.1. Standard culture conditions

All cell lines were cultured in cell culture flasks and plates in RPMI 1640 full medium at 37°C in a fully humidified 5% CO2 atmosphere.

RPMI 1640 full medium: RPM 1640 with stable glutamine + 10% fetal calf serum (FCS) + 100 U/ml penicillin + 100 µg/ml streptomycin

When working with cell lines with doxycycline inducible constructs, FCS tested to be tetracycline-free was used exclusively.

Most cells were at least semi-adherent to the culture flasks. When culture dish coating was demanded for adherence, gelatin or collagen coating was used. Both coating strategies were interchangeable, but collagen coating appeared superior for long-term culture (e.g. single cell

cloning). Generally, culture dishes were coated with gelatin for CHLA25, EW3, and TC106, with collagen for CHLA10, EW24, and MIC.

For gelatin coating, 2% sterile collagen was melted in water bath at 37°C, 5 µl per cm² culture surface were filled into the culture flasks/dishes, and incubated for 30 min at 37°C. Excess liquid was aspirated and cell surface dried for 15 min before using the culture flasks/dishes.

For collagen coating, collagen from calf skin (3 mg/ml) was diluted 1:45 in phosphate-buffered saline (PBS) on ice, 40 µl per cm² culture surface were filled into the culture flasks/dishes. After 30 min incubation at 37°C, liquid was aspirated, and culture flasks/dishes were ready to use.

To exclude any mycoplasma contamination, which might affect assays and readouts, supernatant of all cells was frozen whenever they were used for an assay. In case of a detection of mycoplasma in the supernatant (see Mycoplasma PCR), infected cells were cultured with 1:3,000 Plasmocure in RPMI 1640 full medium for at least seven days and retested after expansion of the cells without further treatment. In case of a persistent contamination, the treatment was prolonged incrementally.

3.2.1.2. Subculturing/harvesting cells

Cells were subcultured every 2-5 days before total confluence. For subculturing/harvesting cells, supernatant was aspirated and cell surface washed with PBS to clean from debris and trypsin-blocking magnesium or calcium ions. Then, cells were detached by trypsin proteolysis for up to 5 min at 37°C, trypsin activity was stopped by addition of medium, cells were resuspended in trypsin-medium-mixture, and spinned down at 360 g. The supernatant was removed and the cell pellet was resuspended in full medium. Cells were subcultured in ratios 1:2 to 1:10.

3.2.1.3. Freezing cells

For cell storage, freshly detached and centrifugated cells were resuspended in freezing solution, transferred into cryo tubes, slowly cooled down by one degree per minute to -80°C in Mr. Frosty freezing containers and then transferred to liquid nitrogen.

Freezing solution: 45% vol RPMI full medium, 45% vol FCS, 10% vol DMSO

3.2.1.4. Counting cells

For counting cells, freshly harvested cell suspension was mixed with trypan blue and 10 µl were transferred into a Neubauer counting chamber. The average count of vital cells, meaning cells without trypan staining (trypan exclusion method), in all 4x4 squares was multiplied with the dilution factor (cells in trypan blue) and the chamber factor 10^4 to calculate the actual cell number per millilitre cell suspension.

For automated cell count, the Countess Automated Cell Counter was used with manufacturer's disposable counting chambers and a 1:1 dilution of cell suspension and trypan blue. The threshold between living and dead cells was adjusted for each cell line individually by adapting the parameters circularity and size.

3.2.2. Transient RNAi mediated protein knockdown

To reduce expression of the *EF1* type 2 fusion oncogene, RNA interference (RNAi) was employed: The posttranscriptional expression regulation via endogenous micro RNA (miRNA) was mimicked by synthetic small interfering RNA (siRNA) transferred into cells using a transfection reagent. In the transfected cells, the siRNA complexes with the RNA-induced-silencing complex (RISC), binds to complementary transcript finally leading to degradation of the bound mRNA and subsequent reduced translation and protein expression (Novina and Sharp, 2004).

siRNAs were prediluted to 20 µM. In RPMI medium without further supplements the transfection reagent and the siRNA were mixed (see Table 1), incubated for 12 min to enable complex formation of the siRNA with the transfection reagent, and then 313 µl mixture were transferred to cells freshly seeded in 10 cm² dishes in 1.687 ml full medium (fast forward transfection).

Table 1: Volumes for siRNA transfection mix. 313 µl mix are added to cells growing in a 10 cm² dish.

Final siRNA concentration	10 nM	25 nM	50 nM
Medium, supplement free (µl)	337.1	337.3	329.5
siRNA (20 µM stock) (µl)	6.9	5.18	10.35
HiPerFect transfection reagent (µl)	1.04	2.59	5.18

For lower concentrations of siRNA, lower amounts of cell toxic transfection reagent are needed. Hence, cells were first transfected with low siRNA concentrations (10 nM and

25 nM). To increase the RNAi effect, in a second experiment 25 nM and 50 nM were used and cells were retransfected according to this protocol after 48 h.

3.2.3. Lentiviral transduction of shRNA

While the RNAi effect of siRNAs is only intermediary, as the siRNA gets used up and degraded, stable RNAi can be achieved by genomic integration of a short hairpin RNA (shRNA) sequence. Transcribed shRNA acts like endogenous miRNA.

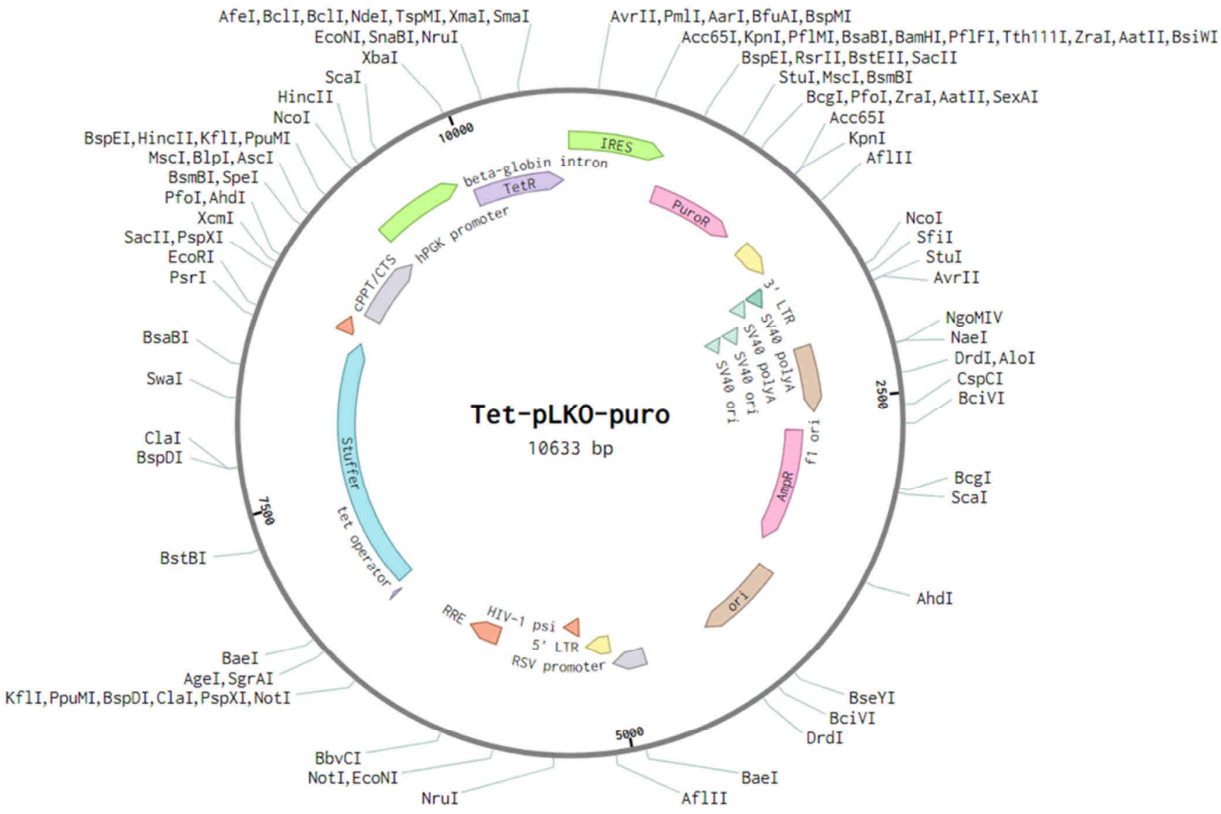

Figure 3: Vector map of the Tet-pLKO-puro vector with its components and restriction sites

For genomic integration of a shRNA sequence, together with a transcription activating promoter, a lentiviral system was employed. The shRNA integrated into the Tet-pLKO-puro vector was packed into a replication incompetent lentivirus, which infected the target cells, and utilizing its lentiviral enzymes reverse transcriptase and integrase, integrated stably the Tet-pLKO-puro sequence between two long terminal repeats (yellow in Figure 3) into the genome.

The Tet-pLKO-puro vector (Figure 3) comprises an ampicillin resistance cassette for selection of bacteria containing the vector, a puromycin resistance cassette which is integrated into the genome of the target cells for selection of successfully transduced cells, and a tet-responsive element, which enables tetracycline mediated activation of the transcription of the shRNA (Wiederschain et al., 2009). The latter feature was the main reason for selecting this vector system: While other vector systems result in constitutive expression of the shRNA construct, which might impair cell growth *in vitro* and *in vivo* or might even make it impossible, the shRNA integrated from the Tet-pLKO-puro vector is not expressed until induction by simple addition of doxycycline (Dox) to the culture medium or drinking water of mice.

shRNAs against different *EWSR1-ETS* fusion oncogenes were designed and their knockdown efficiency was predicted with the gpp web portal of the Broad Institute (Cambridge, MA, USA). Sequences were selected on the criteria 100% sequence match, highest adjusted score (predicted knockdown performance and cloneability adjusted for predicted off-target effects), and least predicted additional human target sequences (especially those, which are expressed in the target cell type EwS). Additionally, a non-targeting control shRNA was used. shRNA sequences (top and bottom) were ordered with the correct overhangs for ligation with the digested TET-pLKO-puro vector. Both sequences were reconstituted to 100 pmol/µl in nuclease-free water. For annealing of the sequences, 11.25 µl of top and bottom shRNA were mixed with 2.5 µl of 10X annealing buffer, the sample was heated up to 95°C and slowly cooled down by 1°C/min to 14°C. 1 µl annealed shRNA was diluted in 399 µl of 0.5X annealing buffer for further use.

10X annealing buffer: 100 mM Tris-HCl, 1 M NaCl, pH adjusted to 7.4

Two times 4 µg TET-pLKO-vector-puro (10,633 bp) was double digested with 2 µl EcoRI HF (20 U) and 1 µl AgeI HF (20 U) in 1x smart cut buffer for 15 min at 37°C, each. Digestion was inactivated at 65°C for 20 min. Both samples were pooled and DNA was precipitated by addition of 140 µl nuclease-free water, 20 µl sodium acetate (3 M), 440 µl of cold EtOH (100%) and freezing at -80°C for at least 45 min. The precipitated DNA was spinned down at 4°C and 20,000 g for 30 min. The pellet was washed with 1 ml 70% EtOH during 15 min further centrifugation. Then, the pellet was tried at 37°C and resuspended in 12 µl nuclease-free water.

Successful digestion was controlled by loading the DNA on a 1% agarose gel, where the 1,800 bp stuffer (light blue in Figure 3) and the 8,834 bp linear plasmid appeared nicely

separated. The lane at 8 kbp was cut out and cleaned up. DNA concentration was measured with NanoDrop spectrophotometer and adjusted to 25 ng/µl.

For ligation of shRNA and TET-pLKO-puro vector 1 µl vector DNA, 1 µl diluted and annealed shRNA, 1.5 µl 10X ligation buffer, 1 µl T4 DNA ligase and 10.5 µl nuclease-free water were mixed and incubated at room temperature for 30 min.

For expansion of the ligated vector, stellar competent cells were transformed. 50 µl of stellar competent cells thawed on ice were mixed with 3 µl of ligation mix, incubated on ice for 30 min, and then heat shocked for 60 sec at 42°C. After 2 further min of incubation on ice, 447 µl prewarmed SOC medium was added to the bacteria. The bacteria were incubated at 37°C for 1 h on a thermomixer with 300 rpm, concentrated by centrifugation at 8,700 g for 5 min and resuspension in 100 µl SOC medium, and half plated on agar plates with ampicillin (100 µg/ml). Agar plates were incubated at 37°C over night.

Table 2: Components of colony PCR

	Stock conc.	Amount (µl)	Final conc.
Nuclease-free water	NA	27.25	NA
GoTaq green buffer	5x	10	1x
$MgCl_2$	25 mM	6	3 mM
dNTPs	10 mM	1	200 µM
Primer mix (TetpLKO-FW/-Rv)	10 µM, each	2.5	0.5 µM, each
GoTaq Polymerase	5 U/µl	0.25	1.25 U/ 50 µl
Colony solution	NA	3	NA

Table 3: Thermal protocol for colony PCR

Step #	Description	Temperature (°C)	Time (s)	Comment
1	Initiation	95	600	
2	Denaturation	98	10	
3	Annealing	59-49	30	Decrease 1°C/2 cycles
4	Extension	72	60	Return to step 2, 20 times
5	Denaturation	98	10	
6	Annealing	56	30	
7	Extension	72	60	Return to step5, 20 times
8	Final extension	72	600	

A colony-polymerase chain reaction (PCR) was performed to test for successful integration of the vector into the bacteria. Single colonies were picked and incubated in 100 µl LB with 100 µg/ml ampicillin for 1.5 h, 37°C, 300 rpm. The PCR was run with components described in Table 2 and thermal conditions described in Table 3.

On a 1.5% agarose gel, the presence of an amplified shRNA insert (about 420 bp) was controlled.

The remaining colony material in LB was transferred into 50 ml LB medium with 100 µg/ml ampicillin for bacteria growth over night at 37°C and 180 rpm.

500 µl bacteria culture were mixed with 500 µl glycerol (50%) and frozen as stock. The plasmid was extracted from the bacteria with the PureYield Plasmid Midiprep System using manufacturer's protocol. Bacteria in LB were spinned down at 5,000 g for 10 min, pellet was resuspended in 3 ml Resuspension Solution, and lysed with additional 3 ml Lysis Solution. After 3 min incubation time, 5 ml Neutralization Solution were added. The mixture was centrifuged at 15,000 g for 20 min. The supernatant was transferred into a Clearing Column on top of a Binding Column on a vacuum manifold. After all liquid was aspirated, the Binding Column was washed with 5 ml Endotoxin Removal Wash and 20 ml Column Wash Solution. Then, plasmid was eluted with 600 µl nuclease-free water during centrifugation at 2,000 g for 5 min. DNA concentration was assessed with a spectrophotometer.

The shRNA sequence was controlled for correctness by sanger sequencing at Eurofins Genomics (Ebersberg, Germany) using forward (TetpLKO-Seq-Fw) or (if forward primer was not sufficient to generate a read over the shRNA) reverse sequencing primer (TetpLKO-Seq-Rv).

For packaging of the Tet-pLKO-puro vector with shRNA into lentivirus, $2x10^6$ HEK293T cells were plated at day 0 on T75 culture flask in 10 ml medium. At day 1, 4,600 µl OptiMEM, 10 µg plasmid DNA, 10 µg D8.9 and 3 µg VSVG were mixed with 46 µl Plus transfection reagent. After 10 min incubation, 60 µl Lipofectamine LTX transfection reagent were added. After 25 min incubation, the transfection mix was added to the HEK293T cells. Four hours later the medium was altered to normal medium; eight further hours later, the medium was changed to 5 ml medium with 30% FCS. At day 3, about 36 h after the last medium change, the virus containing supernatant of the HEK293T cells was filtered with 0.45 µm pore size to eliminate cellular components.

For infection of target cells, 1 ml of supernatant was added to about 25% confluent, actively proliferating cells in T25. After reaching confluence, the cells were transferred into a T75 culture flask supplemented with puromycin (lowest lethal dose (LLD) or maximum 1 µg/ml). A culture flask with non-infected cells was used to test for successful selection.

The LLD of puromycin for non-transfected cells was identified by seeding wild type cells in 12 well plates with medium supplemented with the puromycin concentrations 0, 0.5, 1, 2, 5, and 10 µg/ml for 48 h and microscopical assessment of the toxic effect. LLD was either 0.5 µg/ml (identified for A673, CHLA10, EW7, EW24, MIC, RH1, SKNMC, TC32, TC71) or 1 µg/ml.

3.2.4. Transient EWSR1-FLI1 overexpression

For transient EWSR1-FLI1 overexpression to test antibody specificity in western blot analysis, cells on 10 cm² culture dishes were transfected with pCMV_EF1_puro (carrying the cDNA for EWSR1-FLI1 type 1, from Didier Surdez, Institut Curie, Paris, France) using LTX/Plus transfection reagents: For three transfections 3 µg DNA in 600 µl OptiMEM were incubated with 6 µl Plus reagent for 10 min. 7.5 µl LTX reagent were added, after another 25 min incubation 191.1 µl transfection mix were added to the cells cultured in 1809 µl medium. The medium was changed to fresh full RPMI medium after 4.5 h. After 48 h the cells were harvested: One quarter was scratched from the culture dish and lysed in RA1 buffer for RNA isolation, the left cells were lysed in RIPA buffer for western blot analysis.

3.2.5. Single cell cloning

To receive isogenic clones of transduced cell lines for better reproducibility, cells were seeded at very low density. First, the supernatant of the cells was kept as conditioned medium, passed through a 0.45 µm filter to remove cellular components and filled up to 18 ml with fresh full medium. Then cells were harvested, resuspended in 5 ml medium and counted. The cell suspension volume corresponding to 96 cells was calculated. By 1:10 dilution the volume was increased until it was manageable. Then 96 cells were transferred into the 18 ml medium with condition medium. 150 µl of this suspension (0.8 cells/well) were seeded per well of a 96 well plate. The growth of single colonies was microscopically controlled and corresponding wells were marked. When colonies reached nearly confluence, supernatant was removed, cells were detached with Accutase at room temperature and transferred into 12-well culture plate, then subsequently transferred to larger culture dishes for testing of knockdown-inducibility, freezing and direct use.

3.2.6. Mice experiments and tissue microarray (TMA) generation

Mice experiments were conducted with allowance of the government of Upper Bavaria (ROB-2532.Vet_02-15-184) and in accordance with the 3R principle (replacement, reduction and

refinement). NOD/scid/gamma immunodeficient mice were kept in individually sterile ventilated cages in the certified mouse facility of the Institute of Pathology of the LMU Munich. Cages were only opened under a laminar flow. Mice at the age of at least eight weeks were xenotransplanted with EwS cells with Dox-inducible fusion oncogene knockdown by subcutaneous injection of $2.5x10^6$ EwS cells in 100 µl 1:1 mix PBS and Geltrex into the right flank after shaving of the injection site. Per cell line/experiment 12 mice were xenotransplanted. The welfare of the mice was controlled daily. Tumor size was controlled every two days with a caliper. Additionally, each mouse was controlled for any changes in body condition, social behavior, appearance, breathing, or dehydration, diarrhoe/rectum prolapse, tumor ulceration/invasive tumor growth/pain/high tumor volume (average diameter > 15 mm), weight loss, increased abdominal volume and abnormal grimace (Mouse Grimace Scale). Whenever symptoms in any category were observed, a score up to 7 (most severe) was given for each category. When the sum score was at least 7, the stress for the mouse was assessed as too high and the mouse was sacrified by cervical translocation. When tumors reached an average diameter of 10 mm, mice were alternately assigned to treated/control group. The treated group received 2 mg/ml Dox (for induction of shRNA integrated into the genome of xenotransplanted cells) in 50 mg/ml sucrose drinking water. The control group received 50 mg/ml sucrose drinking water. Mice were marked with ear punches. 96 h after assignment to any group, mice were sacrified by cervical dislocation and tumors were immediately isolated. Two fractions per tumor were snap-frozen in liquid nitrogen for RNA isolation. The remaining tumor was put into formalin for fixation. After fixation, tumors were dehydrated and embedded into paraffin blocks. From the snap frozen samples, RNA was isolated, reverse transcribed and a quantitative real-time PCR was performed to identify the tumors with the highest and lowest expression of the fusion oncogene. The three tumors with the highest and lowest mRNA expression of the fusion oncogene were used for construction of TMAs. Three cores (1 mm diameter) per tumor were transferred into a TMA scaffold. 4 µm sections were cut, stained with haematoxylin and eosin and via immunohistochemistry for several potentially regulated proteins.

3.2.7. Immunohistochemistry (IHC)

Sections of the TMA generated from EwS xenotransplanted in mice were stained for FLI1 and fusion oncogene targets at the immunohistochemical facility of the Pathology of the LMU Munich. For antigen retrieval, microwave treatment was performed with the antigen retrieval AR-10 solution for FLI1, the antigen retrieval ProTaqs I and V Antigen-Enhancer for pMYBL2 and PAX7, and the Target Retrieval Solution S1699 for SOX6. Endogenous peroxidases were blocked with 7.5% H_2O_2 and blocking serum for 20 min. As primary antibodies, anti-

FLI1 (254M, 1:120), anti-pMYBL2 (ab76009, 1:100), anti-PAX7 (PAX7-c, 1:180) and anti-SOX6 (HPA003908, 1:1,600) were used. As secondary antibody, peroxidase-conjugated anti-rabbit/-mouse IgG antibody were used: MP-7401 for FLI1, pMYBL2 and SOX6, and PK6200 for PAX7. Slides were stained with AEC+ (SOX6) and DAB+chromogen and counterstained with haematoxylin Gill's Formula. Immunoreactivity was semi-quantified with a modified immunoreactivity score (IRS) as employed for hormone receptor status (Remmele and Stegner, 1987): Staining intensity was scored in four categories, from no to strong staining, represented by score 0 to 3. The stained portion of cells was scored in five categories from 0-19% in steps of 20% to 80-100%, represented by score 0-4. The product of both scores resembled the final IRS.

3.2.8. DNA isolation

Genomic DNA was isolated using the NucleoSpin Tissue kit. About 10^7 cells were pre-lysed in 200 µl Buffer T1, mixed with 25 µl Proteinase K (proteinase from the kit diluted in proteinase buffer) and 200 µl Buffer B3. After protein digestion at 70°C for 15 min, 210 µl ethanol (96-100%) were added for DNA precipitation and the sample was loaded on a collection tube. After centrifugation for 1 min at 11,000 g, the tube's membrane was washed with 500 µl Buffer BW, and 600 µl Buffer B5, each washing step performed with 1 min centrifugation at 11,000 g. The membrane was dried by two further minutes centrifugation and the DNA was eluted from the membrane with 40 µl nuclease-free water. DNA was quantified using 1 µl eluted DNA on a NanoDrop spectrophotometer.

3.2.9. RNA isolation, reverse transcription and qRT-PCR

3.2.9.1. RNA isolation

RNA was isolated using the NucleoSpin RNA kit with a slightly modified manufacturer's protocol: Cells were lysed in 350 µl RA1 lysis buffer and roughly mixed with 350 µl ethanol (70%) to enable membrane binding. The mixture was transferred into a collection tube. RNA was bound to the membrane during centrifugation at 11,000 g at 4°C for 1 min. The membrane was desalted with 350 µl MDB. Potentially contaminating genomic DNA was digested with 95 µl 1:10 in rDNAse buffer diluted rDNAse for 15 min at room temperature. rDNAse was stopped by washing with 200 µl RAW2 buffer. The membrane was washed twice with RA3 buffer, 600 µl and 250 µl. For the last washing step, the centrifugation time was increased to 2 min to dry the membrane. RNA was eluted in 40 µl nuclease-free water. RNA was quantified using 1 µl eluted RNA on a NanoDrop spectrophotometer.

3.2.9.2. Reverse transcription

RNA was reverse transcribed to cDNA using the High Capacity cDNA Reverse Transcription Kit. Per reaction 2 µl 10x RT Buffer, 2 µl 10x RT Random Primers, 0.7 µl 25x dNTPs and 0.7 µl of MultiScribe Reverse Transcriptase were filled up to 20 µl volume with 1 µg RNA and nuclease-free water. The thermal conditions indicated in Table 4 were applied.

Table 4: Thermal protocol for reverse transcription

Step #	Description	Temperature (°C)	Time (s)
1	Primer annealing	25	10
2	Reverse transcription	37	120
3	Inactivation of the reverse transcriptase	85	5

Resulting cDNA was diluted 1:10 in nuclease-free water and stored at -20°C.

3.2.9.3. Primer design and test

As primer design tool the online Realtime PCR Tool from Integrated DNA Technologies (Coralville, IA, USA) was used with mRNA sequences retrieved from NCBI (Bethesda, MD, USA). Primer pairs were tested in an *in silico* PCR with the UCSC browser and with BLAT for their specificity. Exon-exon spanning primers were preferred, as eventually contaminating genomic DNA would not be sufficiently amplified in a qRT-PCR protocol with such primers to cause any bias on the results. When several isoforms of a gene were known, primers binding all isoforms were preferred.

For testing the primers, qRT-PCR was performed with cDNA of cells with known high expression of the target sequence. The cDNA was used undiluted and in six subsequent 1:2 dilutions (maximum dilution 1:64). When displaying the dilution on an x-axis in log2-transformation (0 to -6) and the Cq values on the y-axis, the optimal slope would be 1 (meaning for each duplication of the PCR product one additional cycle is needed). The average actually observed slope x was used in the formula $2^{\frac{-1}{x}-1}$ to calculate the primer specificity (optimum 1). Additionally, melting curves were checked to show only one peak per primer pair.

3.2.9.4. Quantitative real-time polymerase chain reaction (qRT-PCR)

qRT-PCR combines PCR and quantification. DNA gets denaturated at high temperature, then primers complementary to a specific sequence can bind to the single-stranded DNA (annealing), which builds the starting point for DNA polymerisation. By cyclic repetition of these steps, DNA flanked by the primers gets exponentially amplified. A fluorescent dye binding to double stranded DNA indicates the amplification. The cycle number, when the fluorescence signal passes a threshold, is depicted as Cq (quantification cycle) value.

Per reaction 7.5 µl SYBR Select Master Mix (including buffer, polymerase, nucleotides and fluorescent dye) were mixed with 6.75 µl of the 1:10 pre-diluted cDNA and 0.75 µl of forward-/reverse-primer mix (conc. 10 nmol/ml, each). 96 well PCR plates were used. Standard PCR plates were clear. When high Cq values were expected (>30), white plates were utilized to avoid any risk of cross-signaling. qRT-PCR was performed in a CFX Connect Real-Time System cycler using the protocol indicated in Table 5.

Table 5: Thermal protocol for qRT-PCR

Step #	Description	Temperature (°C)	Time (s)	Comment
1	Polymerase heat activation	95	120	
2	Denaturation	95	10	
3	Annealing and elongation	60	30	Plate read, go to step 2 for 49 times
4	Final denaturation	95	30	
5	Cooling before melting curve	65	30	
6	Melting curve	65-95	5	Plate read and increment of 0.5°C, repeat

To quantify the qRT-PCR products, the ΔΔCq method was used. First, the Cq value of each sample and primer pair was normalized to the Cq value of the same sample and the primer pair for the housekeeping gene *RPLP0*. Then, the resulting ΔCq values were compared to the ΔCq value of a control sample. As with each PCR cycle the product is doubled, a high ΔΔCq means that more PCR product doubling was necessary for signal detection, hence less cDNA and less transcript was present in the sample.

$$\Delta Cq_{sample+primer\ x} = Cq_{sample+primer\ x} - Cq_{sample+primer\ RPLP0}$$

$$\Delta\Delta Cq_{sample+primer\ x} = \Delta Cq_{sample+primer\ x} - \Delta Cq_{control\ sample+primer\ x}$$

$$Relative\ expression_{primer\ x\ target\ in\ sample}[\%] = 2^{-\Delta\Delta Cq_{sample+primer\ x}} \cdot 100$$

3.2.10. Western blot

3.2.10.1. Sample preparation

For quantification of proteins in cell lines, subconfluent grown cells were lysed in 100 µl cold RIPA buffer per 10 cm² culture surface, after washing off the proteins from the medium with PBS. All cell material was scratched into a collection tube, shaken for 30 min on ice, and spinned down at 11,000 g for 5 min. The supernatant was kept as cell debris free protein sample.

RIPA buffer: 150 mM NaCl, 1% Triton X-100, 0.5% sodium deoxycholate, 0.1% sodium dodecyl sulfate (SDS), 50 mM Tris-HCl, pH adjusted to 8; directly before use add 15 µl protease inhibitor cocktail and 5 µl Na_3VO_4 (200 mM) to 980 µl RIPA buffer

Na_3VO_4: 200 mM in ddH_2O, pH adjusted to 10, boiled until clear

Total protein per sample was estimated with the colorimetric Bradford protein assay. The Bradford assay was performed in a 96-well culture dish. For each control and sample 100 µl of 1:5 Bradford solution diluted with ddH_2O were added to different wells. For controls and samples 1 µl volume was used. As controls BSA dilutions containing 0.125 µg, 0.25 µg, 0.5 µg, 0.75 µg, 1 µg, 1.5 µg, and 2 µg BSA per microlitre were applied. After 5 min incubation on an orbital shaker, absorption at 595 nm was measured on a plate reader. Using the controls, a linear function between protein concentration and absorption was calculated in Microsoft Office Excel as standard curve. Based on this standard curve, the protein concentration of the samples was calculated.

3.2.10.2. SDS polyacrylamide gel electrophoresis (SDS-PAGE)

Per sample, 15-30 µg protein were used for SDS-PAGE. To the corresponding volume of the sample, 1/4 volume of loading dye was added and the sample was denaturated at 95°C for 5 min for disruption of tertiary structures.

Loading dye: 312.5 mM Tris-HCl, 10% SDS, 50% glycerol, bromphenol blue; filtered; before use add 250 µl of 1 M DTT to 1 ml loading dye

For all proteins tested here, a 10% resolving gel was suitable. All ingredients of the polyacrylamide gel (Table 6) were mixed and filled into the gel cassette up to 5 mm below the end of the comb. For a clear horizontal line between the resolving and stacking gel, the resolving gel was covered with isopropanol during polymerisation. After polymerisation of the resolving gel, isopropanol was removed, the stacking gel ingredients were mixed and added on top of the resolving gel, and the comb was stuck on top of the gel cassette.

Table 6: Composition of the polyacrylamide gel

	Resolving gel (10%) (µl)	Stacking gel (µl)
ddH_2O	2,925	3,543
1.5 M TrisHCl (pH 8.8)	1,950	NA
1 M TrisHCl (pH 6.8)	NA	860
30% Acrylamide/Bisacrylamide	2,500	500
10% SDS	75	47.5
10% APS	40	40
TEMED	10	10

1.5 M TrisHCl: Add 45.4 g Tris to 250 ml ddH_2O, adjust pH to 8.8

1 M TrisHCl: Add 30.3 g Tris to 250 ml ddH_2O, adjust pH to 6.8

10% APS: dissolve 10 g APS in 100 ml ddH_2O

10% SDS: dissolve 10 g SDS in 100 ml ddH_2O

After polymerisation of the stacking gel, the gel cassette was assembled with the electrode chamber in a tank. Running buffer was added to the upper chamber, which is faced by the upper and open part of the gel cassette, and to the lower part. The comb was removed and in the thus built gel pockets the sample or a protein ladder (2 µl) were added. Electrophoresis run for 90 min at 120 V. The running front was controlled to reach the end of the gel at this time. Afterwards, the gel cassette was opened and the wet blotting system assembled.

Running buffer: 3 g/L Tris base, 14.4 g/L glycine, 0.1% SDS

3.2.10.3. Electroblotting

A wet transfer system was applied. Whatman paper was soaked in transfer buffer. The gel was placed with orientation front-downside on the paper, then the buffer-soaked nitrocellulose membrane was placed on top, carefully avoiding any air bubble, before another Whatman paper was laid on top. This stack was placed into a holding system between two foams and put into the transfer chamber filled with transfer buffer. The membrane faced the anode, while the gel faced the cathode. This way, the negatively charged proteins will follow the electric current in direction to the anode, getting immobilised in the nitrocellulose membrane. Transfer was performed for 90 min at constant 2 A. To avoid overheating, the system was continuously cooled.

Transfer buffer: 3 g/L Tris base, 14.4 g/L glycine, 20% vol methanol

Afterwards, the transfer system was disassembled. For control of equal loading and successful transfer of the proteins on the membrane, the membrane was shortly stained with Ponceau S. Excess staining was washed off with ddH_2O. By three times short washing of the membrane in TBS-T on an orbital shaker, the membrane was decolorized again.

10x TBS: Add 24 g Tris base and 88 g NaCl to 1 L ddH_2O, adjust pH to 7.3

1x TBS: dilute 10x TBS 1:10 in ddH_2O (final conc. 20 mM Tris, 150 mM NaCl)

TBS-T: Add 1 ml Tween-20 to 1L 1x TBS

3.2.10.4. Protein detection

To avoid unspecific antibody binding, the membrane was blocked with 5% milk or 5% BSA (filtered, depending on the preference for the respective antibody) in TBS-T for 1 h. The membrane was then incubated with the protein specific antibody (Table 7) in milk or BSA at 4°C over night.

After washing the membrane and eliminating unbound antibodies with TBS-T (three times, 10 min, each), the membrane was incubated with a horseradish peroxidase (HRP) coupled antibody specific for the species of the primary antibody for 1 h.

Table 7: Antibodies for western blot

Target	Antibody	Species	Dilution	In BSA/milk	Primary/secondary
GAPDH	Santa Cruz 32233	Mouse	1:2,000	milk	primary
FLI1	Cell Marque 254M	Mouse	1:1,000	BSA	primary
ERG	Cell Marque 434R	Rabbit	1:2,000	milk	primary
Anti-mouse	Promega W402B	Goat	1:3,000	milk	secondary
Anti-rabbit	OriGene EU R1364HRP	Goat	1:5,000	milk	secondary

The unbound secondary antibody was removed by three further washing steps with TBS-T, before detection of chemiluminescence. An electrochemiluminescence reagent was directly pipetted on the membrane and the signal generated by the cleavage of this reagent by the HRP was detected and quantified with the LI-COR Odyssey imaging system.

3.2.11. Mycoplasma PCR

To rule out mycoplasma contamination in cell culture, a nested PCR protocol was used. To this end, supernatant was taken from cultured cells after at least three days of cell growth close to confluency. The supernatant was denaturated at 95°C for 5 min and then spinned down to avoid any cellular components in the PCR.

The primers are listed in Table 8. Each primer was diluted to 100 µM according to manufacturers' protocol. In one volume primer mix each primer was diluted 1:10 (hence each primer constitutes to 10% of the total volume of the primer mix). In the primer mix for the first PCR the forward primers were diluted 1:5, (constituting 20% of final volume, each), to balance the concentration of forward and reverse primers.

In the first PCR, 1 µl of cell supernatant and the corresponding primer mix were combined with the remaining components of the PCR mix (see Table 9). PCR was run with the thermal program depicted in Table 10.

Table 8: Primers for mycoplasma PCR

	Oligoname	Sequence 5'-3'
1st PCR	Myco-F1	ACACCATGGGAGCTGGTAAT
	Myco-F1t	ACACCATGGGAGTTGGTAAT
	Myco-R1	CTTCATCGACTTTCAGACCCAAGGCAT
	Myco-R1tt	CTTCTTCGACTTTCAGACCCAAGGCAT
	Myco-R1cat	CCTCATCGACTTTCAGACCCAAGGCAT
	Myco-R1ac	CTTCATCGACTTCCAGACCCAAGGCAT
2nd PCR	Myco-F2	GTTCTTTGAAAACTGAAT
	Myco-F2a	ATTCTTTGAAAACTGAAT
	Myco-F2cc	GCTCTTTCAAAACTGAAT
	Myco-R2	GCATCCACCAAAAACTCT
	Myco-R2ca	GCATCCACCACAAACTCT
	Myco-R2at	GCATCCACCAAATACTCT

Table 9: Components of the mycoplasma PCR. SN: supernatant

	Stock conc.	Amount (µl)	Final conc.
Nuclease-free water	NA	1st 18.75/2nd 18.55	NA
Taq Buffer ($MgCl_2$ free)	10x	2.5	1x
$MgCl_2$	25 mM	1.5	1.25 mM
dNTPs	10 mM	0.5	200 µM
Primer mix	80 µM/ 60 µM	1st 0.625/2nd 0.825	2 µM
Taq Polymerase	5 U/µl	0.125	0.625 U/25 µl
SN/1st PCR product	NA	1	NA

Table 10: Thermal protocol of the mycoplasma PCR

Step #	Description	Temperature (°C)	Time (s)	Comment
1	Initiation	95	120	
2	Denaturation	95	30	
3	Annealing	52 (1st), 45 (2nd)	60	
4	Elongation	68	60	Go to step 2, 39 times
5	Final elongation	68	420	

Then the PCR was repeated, using the product of the first PCR and the second primer mix and the annealing temperature of 45°C (see Table 8-10).

The product of the second PCR run was mixed with 5 µl 6x gel loading buffer and run with a positive and negative control in a 1.5% agarose gel with 100 bp DNA ladder.

3.2.12. PCR for fusion oncogene

To test, which fusion type is present in EwS cell lines, primers for both EWSR1 and ETS transcription factor were used in a PCR with cDNA of the cells. The components are listed in Table 11. A thermal protocol was applied adjusted to the melting temperature of the primers, and the product length (Table 12).

Table 11: Components of the fusion oncogene PCR

	Stock conc.	Amount (µl)	Final conc.
Nuclease-free water	NA	20.375	NA
Taq standard buffer	10x	2.5	1x
dNTPs	10 mM	0.5	200 µM
Primer mix	10 µM, each	0.5	200 nM
Taq DNA Polymerase	5 U/µl	0.125	0.625 U/25 µl
Template cDNA	< 1µg/µl	1	NA

Table 12: Thermal protocol of the fusion oncogene PCR

Step #	Description	Temperature (°C)	Time (s)	Comment
1	Initiation	95	30	
2	Denaturation	95	30	
3	Annealing	52	60	
4	Elongation	68	30	Go to step 2, 39 times
5	Final elongation	68	300	

The product was loaded with 5 µl 6x loading dye buffer on a 2% agarose gel with 100 bp ladder. Type 1 and 2 *EWSR1-FLI1* fusion were easily to distinguish due to a 66 bp upwards-shift of the *EWSR1-FLI1* type 2 PCR product due to an additionally included and amplified exon.

In case of unclear results in the gel electrophoresis, the product was cut out of the gel, cleaned-up and sent for sanger sequencing at Eurofins Genomics with one of the PCR primers as sequencing primer.

To test for the expression of a reciprocal fusion in *EF1* positive EwS cell lines, which is frequently present when a reciprocal translocation resulted in *EF1* instead of more complex genetic events (Anderson et al., 2018), primers were applied as described by Elzi et al. (Elzi et al., 2015). The components and the thermal condition of the PCR are listed in Table 13-14.

Table 13: Components of the PCR for *FLI1-EWSR1* fusion

	Stock conc.	Amount (µl)	Final conc.
Nuclease-free water	NA	14.125	NA
Taq standard buffer	5x	5	1x
$MgCl_2$	25 mM	3	3 mM
dNTPs	10 mM	0.5	200 µM
Primer mix	10 µM, each	1.25	500 nM
Taq DNA Polymerase	5 U/µl	0.125	0.625 U/25 µl
Template cDNA	< 1µg/µl	1	NA

Table 14: Thermal protocol of the PCR for *FLI1-EWSR1* fusion

Step #	Description	Temperature (°C)	Time (s)	Comment
1	Initiation	95	120	
2	Denaturation	98	25	
3	Annealing	63-53	30	-0.5°C increment/ cycle
4	Elongation	72	30	Go to step 2, 19 times
5	Denaturation	98	25	
6	Annealing	56	30	
7	Elongation	72	30	Go to step 5, 19 times
8	Final elongation	72	300	

For reciprocal type 1 fusion, products were expected at 100 bp (primer mix#1) and 180 bp (primer mix#2), for type 2 only at 149 bp (primer mix#2).

3.2.13. Gel electrophoresis and clean-up

For size separation of DNA and detection of PCR products of distinct sizes, agarose gel electrophoresis was employed. To this end, 1-2% (higher percentage for smaller DNA fragments) agarose were mixed with Tris-acetate-EDTA (TAE) electrophoresis buffer, boiled

in a microwave until agarose was completely dissolved, and then cooled to about 60°C before either ethidium bromide, or in later experiments SYBR Safe as DNA stainer, was added and the gel was casted. Electrophoresis with samples in loading dye and DNA ladder was run in a TAE filled chamber at 90-120 V. Stained DNA was made visible with UV light.

10X TAE: 48.4 g Tris base, 11.4 ml of glacial acetic acid (17.4 M), 3.7 g EDTA in 800 ml deionized water, filled up to 1 L

To extract the DNA of a specific lane in the gel electrophoresis, the lane was cut out of the gel and DNA was cleaned-up using the Macherey Nagel NucleoSpin Gel and PCR Clean-up kit according to the manufacturer's protocol. In short, the gel was dissolved in 200 µl NTI buffer at 50°C, then the DNA was bound to a column's membrane during centrifugation at 11,000 g for 30 s, the membrane was washed with 700 µl NT3 buffer by centrifugation, then it was tried by one further minute of centrifugation, before DNA was eluted with 20 µl nuclease-free water.

3.2.14. Whole genome sequencing analyses

3.2.14.1. Sample preparation and sequencing

To generate a multi-dimensional comprehensive dataset characterizing EwS cell line models, whole genome sequencing (WGS) was performed. EwS cell lines planned for WGS were tested for cell line identity by STR profiling. Then, from cells grown on 25 cm^2 culture surface, genomic DNA was extracted. The quality of the DNA was first controlled with a NanoDrop spectrophotometer (high 260 nm:280 nm absorbance ratios, 260 nm:230 nm absorbance ratio around 1.8) and quantified with Qubit.

50 µl of DNA with 50 ng/µl concentration was sent for sequencing at the High Throughput Sequencing Unit (W190) of the Core Facilities Genomics and Proteomics at the German Cancer Research Center (DKFZ, Heidelberg, Germany). The genomic DNA integrity was assessed as DNA Integrity Number (DIN) at the DKFZ, before library preparation with a PCR-free protocol. Sequencing was performed on an Illumina HiSeq X ten platform (Illumina, San Diego, CA, USA) with 150 bp paired-end reads.

3.2.14.2. Alignment

Alignment of the resulting FASTQ files from the sequencing was performed according to a protocol by Wright *et al.* (Wright et al., 2017), which is optimised for the applied sequencer.

In the first step, accounting for 1% PhiX spike-in per lane, FASTQ files were aligned with Burrows Wheeler aligner (Li and Durbin, 2009) (v.0.7.15, bwa mem) to the PhiX reference genome provided by Illumina, simultaneously assigning the read groups read group identifier, platform unit, sample, platform, and library identifier. A SAM file with unaligned reads was generated and directly converted to BAM format with samtools (Li et al., 2009) (version 1.9). With picard (Picard Toolkit, 2019, Broad Institute, GitHub Repository, http://broadinstitute.github.io/picard/) RevertSam the BAM file was converted into an unaligned UBAM file, clearing alignment tags and additionally removing not paired reads. With picard MarkIlluminaAdapters sequenced library adapters were marked with an adapter-trimming tag (XT). The UBAM file was then converted into FASTQ format again with picard SamToFastq, which did additionally clip adapter sequences. FASTQ files were aligned to the hg19 reference genome from the UCSC genome browser with bwa mem. Read information and metadata lost during the conversion of the UBAM file to FASTQ were enclosed to the newly generated BAM files by merging with the UBAM files using picard MergeBamAlignment. Duplicates were marked and removed with picard MarkDuplicate. Base quality scores were recalibrated for variant calling with GATK (Van der Auwera et al., 2013) (version 4.1.0.2) from the Broad Institute, building first a statistical model for known sites from the 1000 Genomes Project (1000 Genomes Project Consortium et al., 2010) and the dbSNP database (Sherry, 2001) with BaseRecalibrator and then adjusting the scores with ApplyBQSR. Finally, quality was controlled with FastQC (Andrews, 2019) (v.0.11.5) and displayed merged with MultiQC (Ewels et al., 2016). Coverage per position was count with samtools depth and displayed as average of 100 kb bins.

3.2.14.3. Variant calling and annotation

Single nucleotide variants were called with bcftools, using mpileup (Li, 2011) (version 1.9) followed by call process and position normalization, and GATK Mutect2, run for somatic mutations. ANNOVAR (Wang et al., 2010) and SnpEff (Cingolani et al., 2012) were employed for annotation. To control SNPs at specific positions of interest, the predicted SNPs were cross-checked with calls from VCMM (Shigemizu et al., 2013), which takes the pileup file of samtools as input, and freebayes (Garrison and Marth, 2012; Liu et al., 2019).

Copy number variation (CNV) was estimated from the WGS data using CNVnator (Abyzov et al., 2011), extracting reads, building histograms, statistics, generating 1000 bp bins for

segments and 100 bp bins for specific sequences, before call and genotyping process. Additionally, CNVkit (Talevich et al., 2016) with 300 bp bins and exclusion of not accessible chromosomal regions was applied on the data.

Genomic transchromosomal fusions were assessed with BreakDancer (Chen et al., 2009).

Further structural variants were called with LUMPY (Layer et al., 2014), giving split and discordant reads as input and using a "blacklist" of non-accessible chromosomal regions, and GRIDSS (Cameron et al., 2017, 2020).

Potential rearrangements by chromoplexy were extrapolated from copy number (CNVnator) and structural variation (GRIDSS) data using ChainFinder (Baca et al., 2013).

Ploidy was estimated with an algorithm from the ploidyNGS tool (Augusto Corrêa dos Santos et al., 2017) comparing the most and second most frequent allele of heterozygous loci for read numbers, expecting a nearly 1:1 distribution in case of euploidy. Due to high ram consumption, the tool was not directly applied on the WGS data, but the method was utilized on read counts already generated by VCMM.

When the tools were not capable of processing all samples together, they were processed independently, but in parallel with GNU parallel (Tange, Ole, 2020).

3.2.14.4. GGAA-mSat genotyping

For genotyping of GGAA mSats in EwS cell line WGS data, first a mSat library was built. The hg19 reference genome was used as input for the tool Tandem Repeats Finder (Benson, 1999) (version 4.09) with parameters match weight 2 (not overrating any match), mismatch weight 5 (medium penality for mismatches, three correct bases in the GGAA motif can compensate one mismatch), InDel (insertion or deletion) penality 7 (strong penality, one full GGAA motif can compensate the penality for an InDel), match probability 80 (maximum), InDel probability 10 (low), minscore 32 (the product of minimal required correct sequence length multiplied with the match weight), and max period size 4 (as four bases GGAA motif is searched).
From the resulting data file, entries with four consecutive GGAA or TTCC motifs were selected and filtered for high rates (39% each) G and A or T and C in the respective mSat. Overlapping sequences were matched with bedtools intersect (Quinlan and Hall, 2010). Thus, the created data file should comprise all GGAA-mSats with at least five repeats, shown as minimal number for enhancer activity (Gangwal et al., 2008), and GGAA-mSats with four repeats, but otherwise only small deviation(s) from the motif, so that small genetic variations

might form even longer GGAA stretches in the EwS cell lines. The list was converted into BED format.
Using HipSTR (Willems et al., 2017) on the aligned WGS data of the EwS cell lines and the GGAA-mSat library in BED format, the mSat positions were genotyped for each cell line. A PCR stutter model was not given as input, as library preparation for sequencing was done with a PCR-free protocol. Readouts of poor quality were removed with the tool dumpSTR with minimal call rate for each locus of 30%, minimal call depth per locus of 10 and minimal supporting reads per allele of 3, maximum InDel probability in mSat flanking regions of 0.15 and minimal call quality of 90%. Synoptic statistics were calculated with qcSTR. Both tools are included in the TRTools package (Mousavi et al., 2020).

3.2.14.5. Motif calling

Enrichment for transcription factor binding sites was investigated using HOMER (Heinz et al., 2010), giving random sequences from the reference genome as control when no comparison group was present, or mSat-flanking regions of not fusion bound mSats and promoters of not fusion regulated genes when probing the fusion bound mSats and promoters of fusion regulated genes. Predicted transcription factor motifs were inspected manually for alignment with enriched sequences.

3.2.15. Chromatin immunoprecipitation and high-throughput sequencing (ChIP-Seq)

3.2.15.1. ChIP

To assess interactions of the EwS fusion oncogenes with the genome and the chromatin state of EwS cell lines, ChIP-Seq was performed. The following ChIP protocol is based on the iDeal ChIP-Seq kit for Transcription Factor from Diagenode. It was refined for EwS by Didier Surdez (Institut Curie, Paris, France) and taught to the author of this thesis. The author established this protocol at the Institute of Pathology of the LMU Munich.

For chromatin immunoprecipitation (ChIP), $4x10^6$ cells for transcription factors and 10^6 cells for histone marks were needed. Cells were seeded 48 h before ChIP in duplicates and controlled to be adherent, not fully confluent, and to grow in still fresh medium before use to avoid any loss or bias due to metabolic changes. One culture flask per cell line was harvested for counting the cell number. The other flask was used for ChIP. First, medium was changed to fresh medium with 1% methanol-free formaldehyde for crosslinking of protein and DNA. After 10 min incubation at room temperature, formaldehyde was quenched with 1 ml gylcine (2 M) per 10 ml medium for 5 min.

Culture surface was washed three times with 20 ml ice-cold PBS. Cells were lysed with as many millilitres lysis buffer IL1B as million cells were counted for the complete culture flask. Lysate and cells were scratched from the culture surface and incubated on ice for 20 min. The lysate was spinned down at 4°C and 500 g for 5 min, supernatant was decanted, and pellet was resuspended in 6 ml IL2 per 10^6 cells to break the nuclear membranes. After 10 min incubation on ice and 5 min centrifugation at 4°C and 500 g, supernatant was decanted and the pellet of lysed cells was resuspended in 333 µl IS1b shearing buffer with 1:200 protease inhibitor per $5x10^6$ cells. 333 µl lysate were sonicated in a Bioruptor Plus, high power, 30 sec on/off. The optimal number of sonication cycles was tested before ChIP for each cell line (see Table 15) to achieve genomic fragments between 200 and 600 bp, hence long enough for sequencing, but not too long for pulldown or reducing information in the sequencing. Sonicated DNA was centrifuged for 10 min at 4°C and 16,000 g and the supernatant was transferred into new reaction tubes.

Table 15: Sonication cycles tested and applied for optimal DNA fragmentation of EwS cell lines for ChIP

Cell line	Sonication cycles		Cell line	Sonication cycles
A673	20		MIC	23
CHLA10	17		POE	16
CHLA25	9		RDES	20
EW1	23		RH1	20
EW22	20		SKES1	20
EW24	25		SK-N-MC	20
EW3	10		TC106	15
EW7	23		TC32	20
MHHES1	17		TC71	16

Magnetic beads were washed three times for 5 min with 1 ml washing buffer (3.2 ml ChIP-Seq-grade water, 800 µl IC1b, 80 µl BSA) on ice, and then resuspended in the initial volume. 30 µl beads were coupled to antibodies by addition of 1.8 µl proteinase inhibitor cocktail, 6 µl BSA, 20 µl of 5x buffer IC1b and maximum 2 µg of the antibody (see Table 16). The bead mix was filled up to 100 µl with ChIPseq-grade water and incubated 4 h at 4°C on a rotation wheel.

For transcription factor binding antibodies, 250 µl sheared chromatin were added to the magnetic beads, for histone marks 66 µl sheared chromatin and 174 µl shearing buffer. For binding of the antibodies, the mixture was incubated at 4°C over night.

Table 16: Antibodies for ChIP

Target	Antibody	Concentration (µg/µl)	Volume (µl)	Amount in ChIP (µg)
EWSR1-FLI1	Abcam 15289	0.2	10	2
EWSR1-ERG	Abcam 92513	0.361	4.16	1.5
H3K4me3	Diagenode C15410003	1.4	1	1.4
H3K27ac	Abcam 4729	1	1	1
H3K27me3	Diagenode C15410069	1.45 µg/ml	2	2.9

The bead-antibody-protein-DNA suspension was washed with 350 µl of iW1, iW2, iW3, and iW4 for 5 min on the rotation wheel at 4°C, each. Then, DNA was eluted from the beads with 100 µl iE1 during 30 min incubation at room temperature on a rotation wheel.

DNA was decrosslinked with 4 µl IE2 at 65°C for 4 h. As control for unspecific bindings, 2.5 µl of sheared chromatin (input) was filled up with iE1 to 100 µl and processed from the decrosslinking step on in parallel with the other samples.

DNA was purified by addition of 2 µl Carrier, 100 µl isopropanol (100%) and 10 µl iPure beads v2. After 10 min incubation on a rotation wheel to enable binding of the DNA to the beads, the beads were washed with 100 µl wash buffer 1 and 2 for 3 min on a rotation wheel, each.

Lastly, chromatin was eluted from the iPure beads with 25 µl of elution buffer C for 15 min on a rotation wheel. The elution step was repeated and the eluate was pooled.

10 µl of the eluate were diluted with 190 µl nuclease-free H_2O and used for qRT-PCR, following the standard qRT-PCR protocol using primers for CCND1-CTCF (CTCF binding site), CCND1-EF1 (EWSR1-ETS binding site), CDKN1A, CUL1, GATA2, IGF2, SEC14L2. A relatively high amount of DNA template for the CCND1-EF1 and SEC14L2 primers was expected for FLI1/ERG and H3K27ac ChIP products. Additionally, GATA2, CDKN1A, and CUL1 primers were designed as positive control for H3K27ac ChIP. CCND1-CTCF served for control of the promoter mark H3K4me3 ChIP. IGF2 is normally repressed (also by CTCF) and was hence used for control of the repressive histone mark H3K27me3 ChIP. As the input control was 1% of the DNA used for FLI1/ERG ChIP, the same Cq value for input and FLI1/ERG could be interpreted as 1% of the in qRT-PCR detected sequence was actually immunoprecipitated. Knowing, that 66/250 sheared chromatin was used for histone marks, the same Cq for input and histone marks could be interpreted as 3.79% immunoprecipitation. In this way, the immunoprecipitated DNA sequences were quantified relative to input.

To test for sufficient chromatin shearing, 50 µl sheared chromatin were mixed with 150 µl PBS and decrosslinked with 1.5 µl proteinase K (20 mg/ml) for 4 h at 65°C. DNA was cleaned up with 201.5 µl phenol/chloroform/isoamlyalcohol 25/24/1. After 10 min centrifugation at high speed, the aqueous phase was mixed with 200 µl chloroform/isoamylalcohol 24/1. After another 10 min of centrifugation, DNA was precipitated with 2 µl glycogen (20 mg/ml), 20 µl sodium acetate (3 M), and 400 µl ethanol (100%) at -20°C for at least 30 min. Precipitated DNA was spinned down at highest speed for 20 min at 4°C, washed with 500 µl ethanol (70%), and spinned down again for 10 min. Supernatant was removed carefully. After the pellet was dried, DNA was resuspended in 20 µl nuclease-free water. Potential RNA decontamination was digested with 2 µl of RNAse cocktail at 37°C for 30 min. DNA was quantified on a spectrophotometer and 2.5 µg were run on a 1.3% agarose gel.

The ChIP product was quantified with Qubit. 10 ng DNA in 50 µl volume were sent for library preparation and sequencing at the Division for NGS of the Institut Curie, Paris, France.

3.2.15.2. Sequencing, alignment and peak calling

Libraries of the ChIP product were prepared at the Institut Curie and sequencing was performed on the Illumina HiSeq 2500 platform with 100 bp single end reads. For CHLA25, EW3, RDES, SKES1, and TC106 the read length was increased to 150 bp. EW24 ChIP was sequenced with NovaSeq 100 bp paired end reads.

As for A673, CHLA10, EW1, EW7, MHHES1, MIC, POE, and TC71 ChIP products from Didier Surdez were already sequenced, the corresponding ChIP products of the author of this thesis were not additionally sequenced. The data were kindly provided for collaboration.

Alignment was performed by Sandrine Grossetête-Lalami (Institut Curie). The FASTQ files from sequencing were aligned with Bowtie 2 (Langmead and Salzberg, 2012) to the hg19 reference genome. Data were converted into BAM format and sorted with samtools. Peaks were called with MACS2 (Zhang et al., 2008), using input as control and the broad-flag option for histone ChIP. BedGraph files were normalized with the corresponding input file, and sorted and converted into bigWig format with the UCSC tools.

3.2.15.3. Downstream analysis of ChIP-Seq peaks

Genes nearby peaks, overlaps of peaks between samples and overlaps with specific genomic regions were identified with HOMER and bedtools (v.2.27) closest and intersect

(Heinz et al., 2010; Quinlan and Hall, 2010). So called super-enhancers (SE), regions with high H3K27ac described as relevant for cell identity, were identified by the ROSE algorithm, stitching regions with H3K27ac histone modification together to potential enhancer regions and ranking them by their coverage in ChIP BAM files, thereby separating super-enhancers of other enhancers (Lovén et al., 2013; Whyte et al., 2013). For cross sample normalisation of peak height, bigWig files were normalized to each other with the R tool ChIPIN (Institut Curie, not yet published), which uses across samples constantly expressed genes as markers for likely similar ChIP signals. The list of constantly expressed genes was extracted from microarray data (see below) of the same cell lines as used for ChIP, by sorting genes by their expression standard deviation among the cell lines and selecting those 10% with lowest standard deviation. The data were converted into bedgraph files with UCSC bigWigToWig to calculate cross-sample normalized read densities relative to transcriptional start sites and other ChIP peaks using HOMER. Additionally EnrichedHeatmap and circlize were employed for heatmap creation (Gu et al., 2014, 2018).

3.2.16. Transcriptome analyses

3.2.16.1. Sample preparation

In order to investigate the effect of EwS fusion oncogenes on gene expression, DNA-microarray analyses were performed. EwS cell lines with inducible fusion oncogene knockdown were plated in 6-well plates (10 cm^2) with the ratios described in Table 17. The ratios were optimized to reach around 65% cell confluence at the time of cell harvest. Three wells were treated with 1 µg/ml Dox for fusion knockdown induction, three others served as control. After 48 h, the medium was refreshed and filled up to 4 ml. After 96 h, the samples were harvested in RA1 lysis buffer.

For all samples, RNA was isolated, quantified with a NanoDrop spectrophotometer, partly reverse transcribed for quantification of fusion knockdown in qRT-PCR. The requirements for DNA microarrays were 1) volume higher/at least 10 µl, 2) amount 1 µg, concentration 50-200 ng/µl. Hence, 11 µl of 100 ng/µl RNA was sent for DNA-microarray analysis.

Table 17: Cell splitting ratios for EwS cell lines to reach 65% confluence after 96 h

Cell line	Ratio -Dox condition	Ratio +Dox condition		Cell line	Ratio -Dox condition	Ratio +Dox condition
A673shEF1	1:10	1:10		MICshEF1	1:8	1:8
CHLA10shEF1	1:5	1:5		POEshEF1	1:10	1:10
CHLA25shEE	1:10	1:10		RDESshEF1	1:16	1:8
EW1shEF1	1:24	1:8		RH1shEF1	1:27	1:19
EW22shEF1	1:8	1:8		SKES1shEF1	1:9	1:4
EW24shEF1	1:5	1:5		SKNMCshEF1	1:12	1:10
EW3shEE	1:8	1:6		TC106shEE	1:19	1:19
EW7shEF1	1:19	1:10		TC32shEF1	1:19	1:19
MHHES1shEF1	1:13	1:8		TC71shEF1	1:38	1:13

3.2.16.2. DNA-microarray and data normalisation

DNA-microarray analysis was performed at IMGM Laboratories (Martinsried, Germany) on human Clariom D microarrays (Affymetrix, Santa Clara, CA, USA). The CEL files were processed with the Transcriptome Analysis Console (TAC 4.0, Thermo Fisher Scientific, Waltham, MA, USA) and the Affymetrix human Clariom D annotation file (version 2) applying the Signal Space Transformation-Robust Multiarray Analysis (SST-RMA) method (including background reduction and intensity normalization). The readout comprises 135,754 transcript IDs. As the focus of this work is on genes transcribed and translated to potentially clinically relevant proteins, all IDs annotated with RNA, noncoding or pseudogene were removed. Furthermore, unmapped genes (hence not applicable for comparison with ChIP/WGS data), those only predicted by AceView and those not following the guidelines for naming genes from HUGE Gene Nomenclature Committee (only capital letters and numbers, no punctuation except for hyphens) were removed. The remaining genes were labelled with Entrez Gene IDs using DAVID (Huang et al., 2009a, 2009b). This resulted in a list of 25,962 unique transcript IDs annotated with 25,237 unique genes. When several transcript-IDs represented one gene, the mean expression value was calculated and used for further investigations.

3.2.16.3. Identification of regulated genes

To define regulated genes, but avoid arbitrary cut-offs, the algorithm of SE identification from ROSE was adopted: Genes were displayed ranked by their fold change from highest negative to highest positive fold change upon fusion oncogene knockdown (x-axis) together with their fold-changes (y-axis). When normalizing fold changes and x-axis, all

downregulated genes, where the slope did not drop beneath one, and all upregulated genes, from when the slope reached 1, were defined as regulated. This corresponds to building a linear function between 0:0 and the most extreme fold change and then finding the gene, where a linear function with the same slope is tangential.

3.2.16.4. Comparison of transcriptomes

As comparison dataset to the cell lines, previously published expression data (Baldauf et al., 2018b) on 21 cancers and 71 normal tissues were used. To eliminate batch effects, expression values of the different cohorts were normalized to each other using ComBat (Johnson et al., 2007). Similarities in and between the cohorts were assessed with t-Distributed Stochastic Neighbor Embedding (t-SNE) in R.

3.2.17. Proteome analyses

3.2.17.1. Sample preparation

Complementary to the transcriptome, the proteome of EwS cell lines was assessed. Cells were grown to about 60-70% confluence on 10 cm² culture dishes for 96 h with and without Dox treatment, as done for transcriptome analysis. Relatively high medium volume was applied to avoid any decrease in pH and accompanying potential metabolic effects. Afterwards, cells were washed to remove any proteins from the medium: First 4 ml medium without any supplements was added and directly aspirated, then cells were incubated with 2 ml empty medium for 15-30 min at 37°C, then again washed with 2 ml empty medium.

Cells were lysed in Nonidet-P40 buffer for 15 min at RT. Lysed cells were scratched into protein low-bind tubes, sonicated with 60% amplitude, 30 sec on, 6 cycles. Lysates were quantified with Bradford assay (see Western blot).

Nonidet-P40 buffer: 1% Nonidet P40 in 10 mM NaCl, 10 mM Tris-HCl, pH 7.6; one complete protease mini tablette added per 10 ml lysis buffer immediately before use

10 µg protein lysate were sent for mass spectrometry and protein quantification to the research unit Protein Science of the German Research Center for Environmental Health (Munich, Germany).

3.2.17.2. Imputation and interpretation

Quantification data on 7,242 proteins were received. Quantification data were complete for 3,248 proteins. Using the IterativeSVD algorithm of fancyimpute (https://pypi.org/project/fancyimpute/) for proteins with quantification data in at least 70% of samples, missing values were imputed for 1,336 proteins, leading to a total number of 4,584 fully quantified proteins. Regulated proteins were identified with the same algorithm as regulated transcripts. Data were interpreted using gene set enrichment analysis.

3.2.18. Methylation analyses

3.2.18.1. Sample preparation

To assess the effect of fusion oncogene knockdown on methylome, samples were prepared as described for transcriptome analysis, but were lysed in T1 buffer and genomic DNA was isolated. Concentration and purity were assessed using a NanoDrop spectrophotometer. 1 µg of DNA in 17 µl volume were sent for analysis.

3.2.18.2. Analysis and data normalisation

Methylation analysis at CpGs was performed at the Genotyping Platform of the Genome Analysis Center at the German Research Center for Environmental Health (Neuherberg, Germany) using Illumina Infinium MethylationEPIC BeadChip arrays. Methylation levels were calculated from the raw data files (IDAT) using Illumina Genome viewer.

3.2.18.3. Differential methylation analysis and comparative analysis

For identification of differentially methylated regions in samples with fusion oncogene knockdown versus without knockdown, each cell line was analyzed using the R tool minfi (Aryee et al., 2014). For comparison of EwS cell line methylation profile with the methylation profile of various primary cancers, one IDAT file per sample was uploaded to the classification tool of MolecularNeuropathology.org (Capper et al., 2018). The same data were integrated into a t-SNE plot provided by a senior bioinformatician of the respective group (Dr. Martin Sill).

3.2.19. Gene set enrichment analyses and gene ontology

The gene set enrichment analysis (GSEA) tool is a knowledge based predictor of phenotypes associated with a specific gene expression profile (Subramanian et al., 2005). Previously described gene sets annotated with phenotypes were retrieved from the Molecular Signature Database (Broad Institute, c2.all.v6.2). Genes or proteins assessed in the DNA-microarrays or by mass spectrometry were ranked by their fold changes in different fusion states. The ranked list was compared with the database by the GSEA tool with 1,000 permutations. As the enrichment score increases, when a gene from the ranked list is present in a gene set, but decreases in the opposite case, only gene sets with many "hits" at the top or bottom of the ranked list reach a high absolute enrichment score. This value is further weighted by the position of the hits in the list, resulting in the normalised enrichment score (NES).

The Gene Ontology Consortium curates gene annotations in the three ontologies molecular function, cellular component and biological process (Ashburner et al., 2000). By giving a gene list of interest to an overrepresentation tool based on gene ontology data, the genes get annotated, the number of genes expected in the list versus the observed number is compared and over- or underrepresentation documented. Here, PANTHER was employed as classification system (Mi et al., 2019).

3.2.20. Survival analysis

Survival analyses were performed with an in-house generated tool (GenEx) to identify potentially clinically relevant genes. This tool automatically assesses from gene expression data and corresponding clinical data survival association and significance level for each gene in short time. Significance calculations are based on the Mantel-Haenszel test. The option "best percentile" finds in a range between 20th and 80th percentile the optimal cut-off between the two groups with low and high expression of an individual gene based on maximum distance of the survival curves, thereby avoiding multiple testing. The survival data were published before (Sannino et al., 2019) and comprise 196 individual EwS cases from three cohorts with expression data on 13,253 genes. Cross-cohort normalization was performed with Combat. The data fulfill TCGA standards with at least 60% tumor purity. Survival analyses of GenEx were partly validated using Graph Pad Prism.

3.2.21. Statistical analyses and graphical presentation

Statistical analyses were performed with Graph Pad Prism, R and Microsoft Office Excel. The standard tests for assessing statistical significance were the two-sided independent

Student's t-test (between two groups comparison, normal data distribution assumed, based on the Kolmogorov-Smirnow test) and the Mann-Whitney-(U)-Test (between two groups comparison, no normal data distribution). When comparing two groups for two discrete categories, Fisher's exact test or Chi² test (for higher sample number than 5 per group/categorie) were applied.

Data were represented using Graph Prad Prism, Microsoft Office, R ggplot2 (Wickham, 2016), Circos (0.69) (Krzywinski et al., 2009), IGV (Thorvaldsdóttir et al., 2013), Benchling (2020, retrieved from https://benchling.com), BioVenn (Hulsen et al., 2008) or the respective data analysing software described in the respective Methods section.

4. Results

4.1. A EwS model cohort for studying the effects of *EWSR1-ETS* fusion oncogenes was generated

4.1.1. EwS cell lines selection

A relatively large cohort of EwS cell lines needed to be selected for fusion oncogene knockdown experiments and knockdown effect evaluation, as with only two or three cell lines the inter-patient heterogeneity cannot be modelled. One major selection criterion, among the 30 cell lines available at the LMU Munich, was the respective fusion oncogene. As 85% of EwS tumors harbor an *EWSR1-FLI1* (*EF1*) fusion, more *EF1* type 1 than 2, and about 510% *EWSR1-ERG* (*EErg*), these fusions should also be represented by the cell lines. For EErg knockdown, all selected *EErg* positive cell lines should comprise the identical fusion transcript.

To assure the presence of the fusion types described for the different cell lines, thereby performing a first identity check and exclusion of cross-contamination, PCRs for the fusion transcript were done on cDNA. Gel electrophoresis of the fusion transcript showed a clear up-shift of the lane for *EF1* type 2 versus 1 (from 110 bp to 176 bp), reasoned by an additional *FLI1* exon present in the fusion (Figure 4A). Whenever this method was not informative, the presence and type of fusion for each cell line was successfully validated by sanger sequencing of the transcript. This was especially important for the *EErg* positive cell lines to ensure identical transcript sequences. Chromatogramms of the sequencing of CHLA25, EW3 and TC106 showed identical fusions between exon 7 of *EWSR1* and exon 6 of *ERG* (here reverse complement due to sequencing with a reverse *ERG* primer, Figure 4B). Other fusion types identified in the 30 EwS cell lines were *EWSR1*(exon7)-*ERG*(exon7) (COGE352), *EWSR1*(exon9)-*FLI1*(exon7) (EW16), *EWSR1*(exon10)-*ERG*(exon6) (EW18), *EWSR1*(exon7)-*ERG*(exon9) (SKNPLI), *EWSR1*(exon7)-*FEV*(exon3) (TC205) and *EWSR1*(exon10)-*FLI1*(exon5) (ES7, ORS). As these fusion types were only observed once or twice, the available number of cell lines might be too less to test for the comparability of EwS positive for those fusions versus others and/or to test for fusion specific biology. Hence, cell lines with these rare fusion types were not further used.

Next, cell lines were selected based on their growth under the same standard cell culture conditions, as different culture conditions might result in bias in later analyses. Similar doubling times, preferentially under 48 h were desired to achieve quickly sufficient cell numbers for later experiments.

Using these criteria, the 18 cell lines depicted in Table 18 were selected, 11 with *EF1* type 1 fusion, 4 with *EF1* type 2 fusion and 3 with *EErg* fusion.

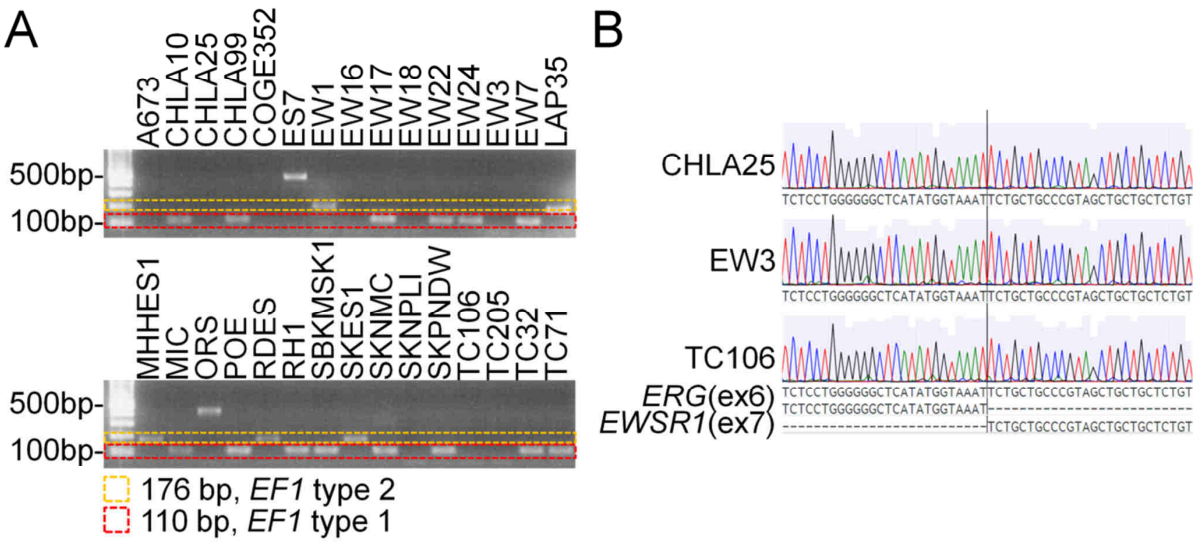

Figure 4: PCR and sanger sequencing confirm the presence of pathognomonic *EWSR1-ETS* fusions in tested cell lines. A) Gel electrophoresis of PCR amplified *EF1* fusion transcript; red dotted: expected size of *EF1* type 1 transcript; orange dotted: expected size of *EF1* type 2 transcript. B) Chromatogramm of reverse Sanger sequencing of the transcribed fusion between *EWSR1* (exon 7) and *ERG* (exon 6) in three cell lines.

Table 18: EwS cell lines selected for fusion oncogene knockdown and generation of the Ewing Sarcoma Cell Line Atlas

EF1 type 1	*EF1* type 2	*EErg*
A673	EW1	CHLA25
CHLA10	MHHES1	EW3
EW7	RDES	TC106
EW22	SKES1	
EW24		
MIC		
POE		
RH1		
SKNMC		
TC32		
TC71		

For these cell lines, genomic DNA was sent for STR profiling. The achieved results were cross-checked with databases of cell line suppliers (ATCC and DSMZ (Dirks et al., 2010),

depicted as A and D in Table 19, respectively) and four publications comprising STR profile data on several EwS cell lines (Brohl *et al.*, Deamen *et al.*, May *et al.*, and Ottaviano *et al.* (Brohl et al., 2014; Daemen et al., 2015; May et al., 2013; Ottaviano et al., 2010), depicted as 1-4 in Table 19). Noteworthy, STR profiles from Ottaviano *et al.* did not comprise the here tested STR loci D13S317, D7S820 and CSF1PO. For the remaining cell lines without any reference, the COSMIC cell line project (Tate et al., 2019) was searched for STR results via cellosaurus (https://web.expasy.org/cellosaurus/, indicated as C in Table 19).

For 16 cell lines, reference or comparison data were found. For A673, all references showed 11 and 12 repeats at the D5S818 locus, but the here generated STR profile did only show one peak at 11 repeats. Interestingly, Ottaviano *et al.* listed this STR profile variant in their results, indicating that biallelic 11 repeats at D5S818 might be a common drift for A673. For SKNMC, the 11-repeat allele at the TPOX locus was not observed in the own data, but was consistently present in the reference data. For EW22, a 12-repeat allele was observed for the D16S239 locus, which was not described in the only reference (COSMIC). Interestingly, data for MHHES1 and TC71 were inconsistent. While Daemen *et al.* described, as here observed, biallelic 8 repeats at the D13S317 locus of MHHES1, DSMZ and COSMIC report one allele with 13 repeats. For TC71, three different references described three different loci as homozygous, while all others described this locus (as observed here) as heterozygous: Daement *et al.* missed seemingly the Y peak for amelogenin, Brohl *et al.* the 11 repeats at the CSF1PO locus, and Ottaviano *et al.* the 18 repeats at the vWA locus.

Only for two cell lines, no reference or comparison data were found, namely MIC and POE. While there were no cell lines in the databases with >80% identity for the here observed POE STR profile (which comprised an 11-repeat allele at locus D13S317 in wildtype cells, which was later in a single cell clone not present, anymore), for MIC there was 89% overlap to the cell lines Calu-6 (anaplastic carcinoma) and NCI-H1838 (lung carcinoma), and 82% with NCI-H661 (lung carcinoma) and HCC1599 (breast carcinoma) at ATCC. As for MIC the pathognomic *EF1* fusion was verified, confusion of this cell line with one of the hits from ATCC can be excluded.

Thus, for 16 cell lines identity could be confirmed, only for three cell lines small drifts in one STR locus, each, were observed, and diversity between the cell lines was proven. Table 19 can be used as a reference, and when working with the here generated cell lines, to control for genomic stability and sample purity.

Table 19: STR profiles for the selected EwS cell lines. The datasets used for comparison of own results are indicated with A (ATCC), D (DSMZ), 1 (Brohl *et al.*), 2 (Daemen *et al.*), 3 (May *et al.*), 4 (Ottaviano *et al.*) and C (COSMIC). Alleles not detected in own data, but described in references, are marked with "-"; alleles observed here but not described in references, are marked with "+".

	D5S818	D13S317	D7S820	D16S539	vWA	TH01	AMEL	TPOX	CSF1PO
A673[A,D,1,2,3,4]	11	8	10	11	15	9.3	X	8	11
	-12	13	12	11	18	9.3	X	8	12
CHLA10[3]	12	12	9	11	16	7	X	11	12
	12	12	10	11	16	7	X	11	12
CHLA25[1,3]	10	10	8	11	15	7	X	8	10
	10	12	10	12	17	7	X	8	10
EW1[C]	11	13	11	12	19	6	X	8	11
	15	14	12	12	19	9.3	Y	11	11
EW3[4,C]	12	12	8	11	15	9.3	X	8	11
	12	12	11	12	16	10	Y	9	12
EW7[C]	10	8	9	12	16	6	X	8	12
	13	8	9	12	18	6	X	9	12
EW22[C]	11	12	10	+12	17	6	X	9	11
	11	12	11	13	17	6	X	11	11
EW24[C]	12	14	12	11	15	6	X	10	11
	12	14	13	11	16	9.3	X	12	12
MHHES1[D,2,C]	13	8	9	11	16	8	X	8	11
	13	8/-13[DC]	11	11	17	9	X	8	11
MIC	10	11	8	12	17	7	X	8	10
	12	11	10	13	20	9	Y	8	12
POE	12	9	11	10	14	8	X	11	10
	13	-11	12	10	17	9	X	11	10
RDES[A,D,1,4]	11	11	10	9	17	7	X	9	11
	11	12	10	11	17	7	Y	11	11
RH1[D]	11	10	8	13	17	9.3	X	8	11
	11	10	12	13	17	9.3	Y	8	11
SKES1[A,D,1,4]	12	8	10	11	14	6	X	8	11
	12	9	11	11	17	9.3	Y	8	11
SKNMC[A,D,1,3,4]	11	11	8	12	17	9.3	X	9	10
	11	11	8	12	18	9.3	X	-11	10
TC32[1,3]	12	10	8	13	15	6	X	9	11
	13	12	11	14	18	9.3	X	11	13
TC71[D,(1),(2),3,(4)]	10	11	10	11	17	9.3	X	8	10
	10	12	10	14	17	9.3	Y	9	11
TC106[1]	12	10	9	13	16	9.3	X	8	11
	13	11	10	13	18	9.3	X	10	13

4.1.2. Generation of EwS cell lines with inducible EWSR1-ETS knockdown

To assess the effects of the fusion oncogene, knockdown experiments were planned. To this end, the cells were stably lentivirally transduced with doxycycline (Dox) inducible shRNA sequences.

For *EF1* type 1 several knockdown experiments have been described. But, to specifically target the fusion transcript, the potential target sequences are limited. Here, we tested target sequences similar to those described by Tirode *et al.* and Carrillo *et al.* (Carrillo et al., 2007; Tirode et al., 2007).

In two experiments, the sequence of Carrillo *et al.* resulted in better knockdown 48 h after shRNA induction with 1 µg/ml Dox in three cell lines tested (Figure 5). Hence, this target sequence was used for all *EF1* type 1 positive cell lines.

Figure 5: shRNA sequences targeting *EF1* type 1 reduce fusion transcription. Left: The *EF1* type 1 transcript sequence around the fusion point is depicted aligned with *EWSR1* and *FLI1*, additionally the target sequence oriented to the work of Tirode *et al.* and adopted from Carrillo *et al.* are shown. Right: Dot plot indicating *EF1* rest expression after knockdown induction in five tested cell lines in up to two experiments, bar indicates mean.

For *EF1* type 2 an attempt to target specifically the fusion has been described, but knockdown was only achieved by targeting *FLI1* C-terminally (Chansky et al., 2004). The design of a specific shRNA against the fusion transcript is highly aggravated as the first exon from *FLI1* in the fusion starts with GTTCA, as does the next *EWSR1* exon, exon 8. The end of *EWSR1* exon 7 again shares some bases with *FLI1* exon 4 (Figure 6A). Two target sequences were tested as siRNAs. As no knockdown was achieved in two pilot experiments, even at 50 nM siRNA concentration and retransfection after 48 h, the possibility to target *FLI1* instead was evaluated. To this end, high expression of wildtype *FLI1* had to be excluded, as done by qRT-PCR (Figure 6B). Two target sequences predicted with good performance were tested, one targeting exon 5, the other exon 9 of *FLI1*. The latter clearly outperformed the other and was used for all *EF1* type 2 positive cell lines (Figure 6C).

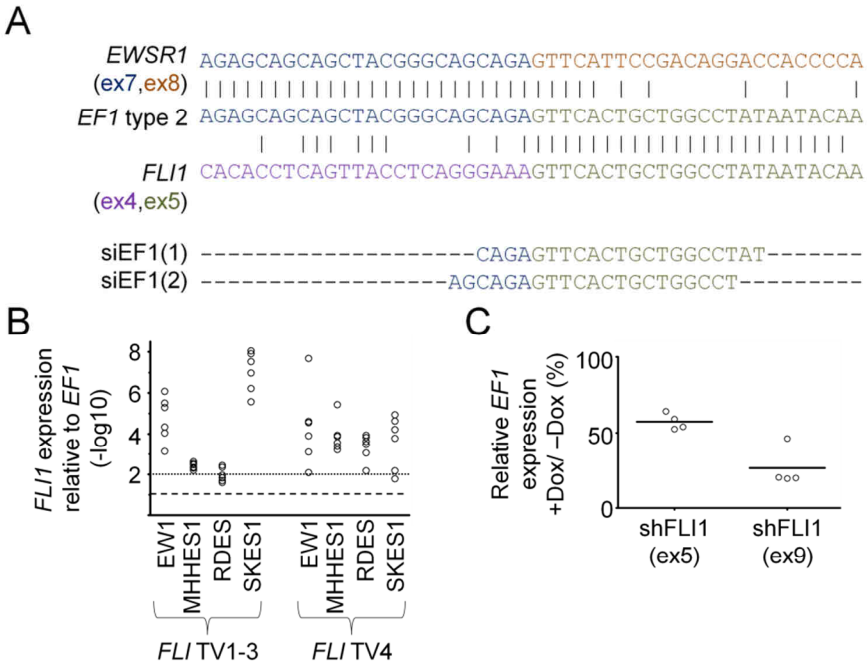

Figure 6: Targeting the *EF1* type 2 transcript at fusion point does not sufficiently reduce *EF1* expression, while C-terminal *FLI1* targeting is reasonable and effective. A) Alignment of the *EF1* type 2 transcript at the fusion point with *EWSR1*, *FLI1* and two potential targeting sequences. B) Dot plot indicating the wildtype (not fused) *FLI1* expression in *EF1* type 2 positive cell lines compared to *EF1* expression represented as -log10, interspaced line indicates 10%, dotted line indicates 1% relative expression, qRT-PCR, *n*=6. C) Dot plot indicating *EF1* knockdown with two different shRNAs, targeting exon 5 and exon 9 of *FLI1*, in a pilot experiment with all four *EF1* type 2 positive cell lines, horizontal bars indicate mean, qRT-PCR.

For *EErg* a target sequence overlapping the fusion was selected with relatively high predicted knockdown efficiency (Figure 7).

In the further text, the cell lines transduced with shRNA against EF1 and EErg are referred to with suffix "shEF1" and "shEE", respectively, whenever potential confusion with wildtype cells might arise.

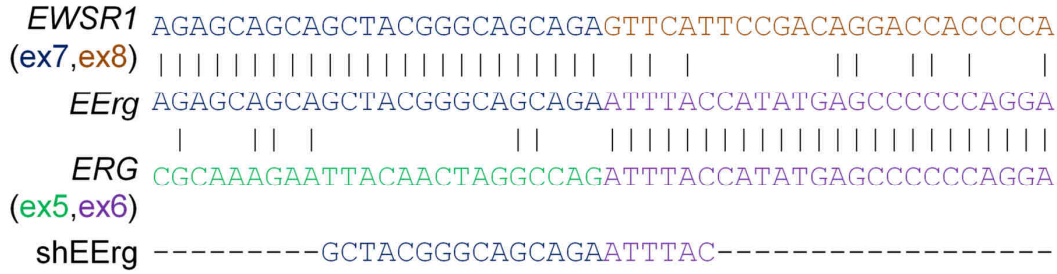

Figure 7: Target sequence of the shRNA against *EErg*. The target sequence is depicted together with the *EErg* transcript at the fusion point aligned to *EWSR1* and *ERG*.

To avoid any sample heterogeneity which might result over time in subclone overgrowth and less reproducible results, isogenic clones of the cell lines were generated. Only for EW1, RDES, EW24 and EW3 no clone was established. While for the first two cell lines clones were growing, but not with sufficient knockdown of the fusion oncogene, the last two did not form any clones, hinting that these cell lines might not be suitable for clonogenic growth assays.

The knockdown over time was assessed to select a time point, when readouts of the knockdown effects should be generated. This experiment in four cell lines showed that knockdown of the *EWSR1-ETS* fusions (*EEts*) reaches its maximum after 48 h and stays stable, at least until 120 h, when Dox (1 µg/ml) was removed (Figure 8). As downregulation of the fusion oncogene took 48 h and the downregulation of its targets should be investigated, knockdown experiments were henceforth carried out with 96 h knockdown induction.

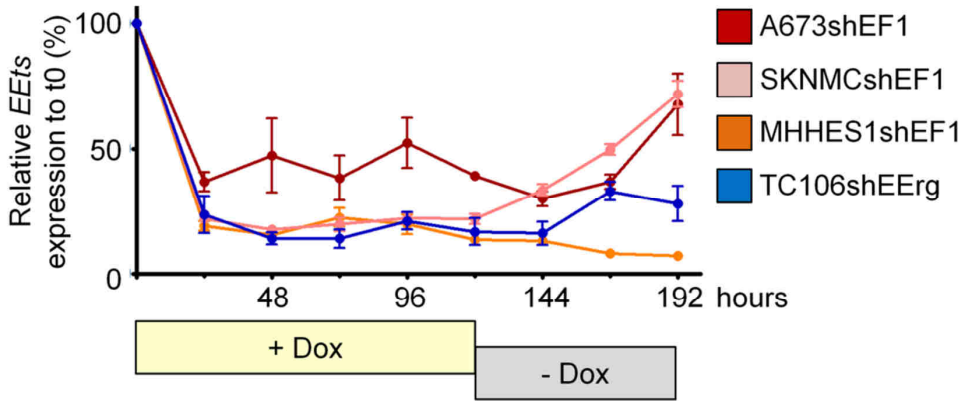

Figure 8: *EEts* is stably downregulated in shRNA transduced cell lines 48-72 h upon Dox treatment. Time curve of the *EEts* transcript level relative to time point 0, from 0-120 h Dox treatment, plus 72 h without Dox treatment. Cell lines are indicated by colors. Dots represent mean transcript level in *n*=3 experiments, whiskers indicate SEM, qRT-PCR.

Four independent experiments showed successful knockdown of the respective fusion oncogene in all generated cell lines on the transcriptional level in qRT-PCRs (Figure 9, upper panel). The PCR results could be confirmed on the protein level by western blot analysis (Figure 9, middel panel). As the downregulation of the fusion oncogene upon Dox addition to the lentiviral transduced cells was consistent, those Dox treated cells are henceforth referred to as EEts-low cells, while those cells without Dox treatment are referred to as EEts-high.

The applied FLI1 antibody was successfully tested for representing the dynamics of EF1 expression with knockdown in A673shEF1 and overexpression in the same cell line and

HEK293T (which does normally not express EF1, see Figure 9, lower panel). The EF1 lane appears at around 80 kDa. Additionally, overexpression of EF1 in the not EF1 expressing osteosarcoma cell lines SAOS-2 und U2OS was tried for antibody validation, but the achieved expression levels were compared to that in HEK293T far less on the transcriptome level (factor 27 and 87 less, respectively), and the 80 kDa lane appeared only very weakly for U2OS. For EErg an ERG antibody described as suitable for the prostate cancer fusion TMPRSS2-ERG has been applied and showed EErg downregulation upon shRNA induction with lanes around 70 kDa.

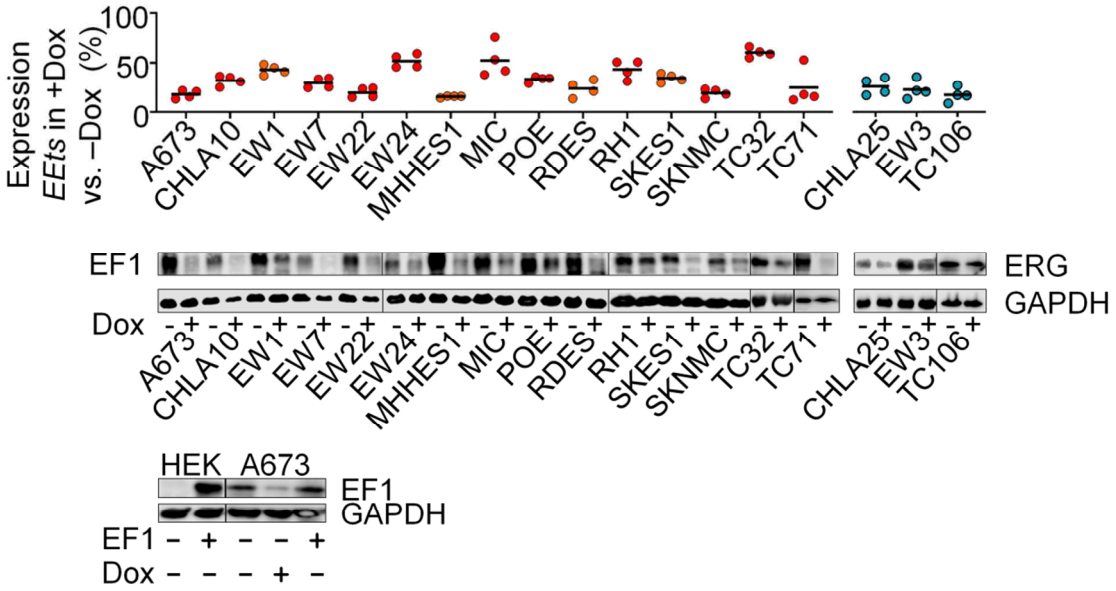

Figure 9: qRT-PCRs and western blot demonstrate successful knockdown induction of the fusion oncogene. Upper panel: Expression of the fusion in EEts-low state (fusion knockdown) vs. -high in all 18 stably transduced cell lines, each experiment indicated as dot (*n*=4), bar indicates mean. Middle panel: Representative western blots of the indicated cell lines without and with (-/+) fusion oncogene knockdown induction, GAPDH is shown as loading control. Lower panel: Western blot of HEK293T (HEK) cells without and with overexpression of EF1, and of A673shEF1 without treatment, with fusion knockdown induction, and with EF1 overexpression.

Lastly, the effect of the shRNAs on the fusion partners was controlled in four experiments on the transcriptional level. The ETS transcription factors were often relatively lowly expressed when compared to the fusion (<0.1% in EEts-high condition) and therefore considered as not abundant when such low expression was detected in three out of four experiments. Conversely, EWSR1 expression was even higher than the fusion expression level. In six of seven cell lines with abundant FLI1 expression (only detectable with primers for transcript variant 1-3, but not for variant 4), the *FLI1* expression level increased upon fusion knockdown in each replicate. Strikingly, even in two of the *EF*1 type 2 positive cell lines, in which the shRNA targets *FLI1*, knockdown of the fusion oncogene leads to massive increase

in FLI1 expression. This effect was not observed for *ERG*. The shRNA against *EErg* showed interference with *EWSR1*, but the knockdown efficiency on the fusion oncogene was on average 46% stronger (Figure 10).

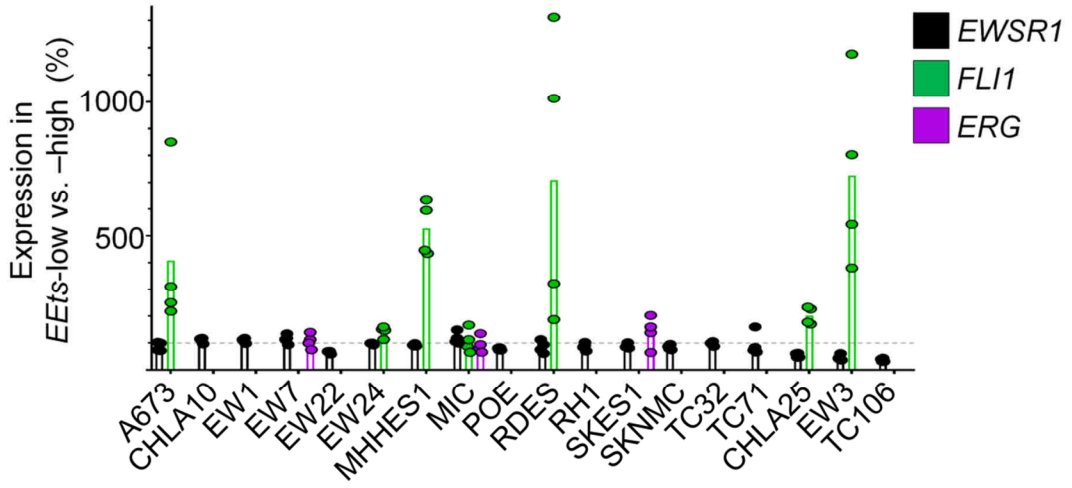

Figure 10: Expression of wildtype fusion partners is mainly unaffected by fusion knockdown. The expression of not fused *EWSR1* (black), *FLI1* (green, TV1-3), *ERG* (purple) in *EEts*-low vs. high cells is depicted; when a gene was less than 0.1% expressed in *EEts*-high condition in three replicates, it was assumed as not expressed and not presented in the graph. Dots represent single measurements of biological replicates (*n*=4), columns mean. Dashed lines indicate 100%, meaning unaltered expression, qRT-PCR.

To possibly test for unspecific effects of Dox treatment and lentiviral transduction, all used cell lines were stably transduced with inducible non-targeting shRNA (shControl), too, if the shControl cell line was not generated in the working group before. Namely, the following shControl cell lines were generated: CHLA10shControl, CHLA25shControl, EW1shControl, EW3shControl, EW22shControl, EW24shControl, EW7shControl, MHHES1shControl, MICshControl, RH1shControl, SKES1shControl, TC71shControl, TC106shControl.

4.1.3. Tissue microarrays of the EwS cell line models enable fusion oncogene target regulation evaluation *in situ*

Six cell lines were expanded *in vivo* to generate model tumors. For half of the tumors, knockdown was induced for 96 h. From the tumor material, two tissue microarrays (TMAs) were generated. These TMAs enabled the *in situ* validation of the effect of EF1 on described target genes on protein level, and can additionally be used for evaluating new antibodies against EF1 targets for immunohistochemistry. Interestingly, the fusion knockdown was even stronger *in vivo* than observed *in vitro* in qRT-PCR (Figure 11).

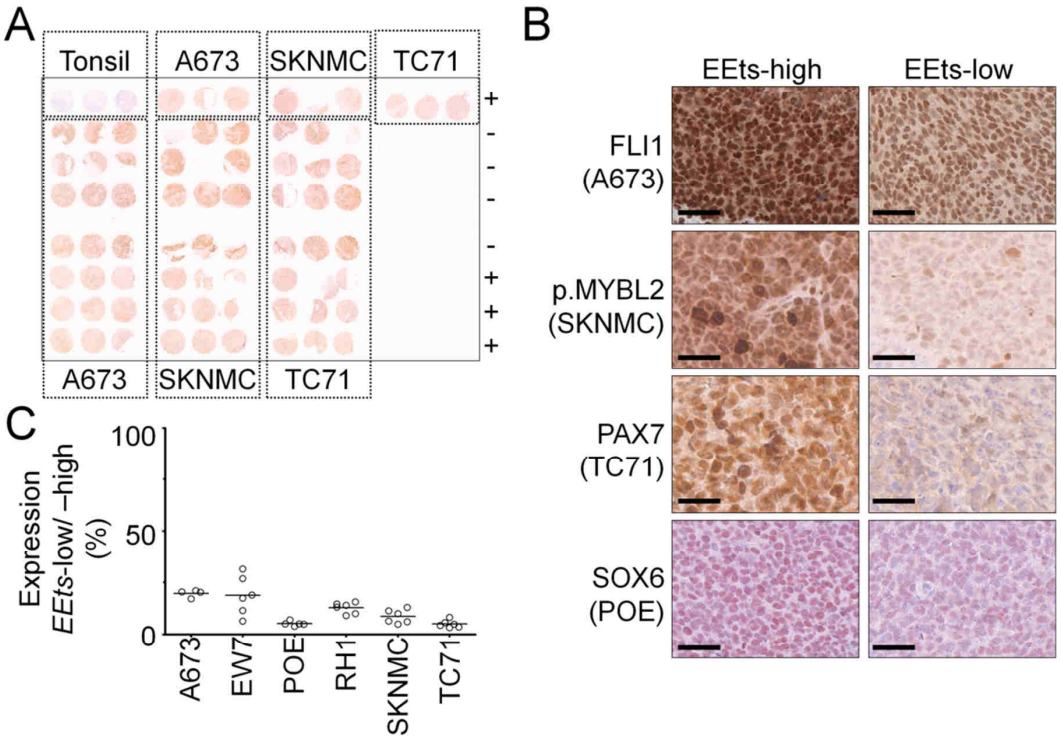

Figure 11: *In vivo* expansion of EwS cell lines with inducible fusion knockdown enables confirmation of regulatory effects *in situ*. A) Scan of one HE stained TMA comprising three cell lines, three tumors per condition (-/+ Dox), three cores per tumor. B) Exemplary micrographs from TMAs stained for FLI1, phospho-MYBL2, PAX7 and SOX6, the cell line is depicted in brackets, scale bar indicates 50 µm. C) *EF1* in EEts-low vs. EEts-high tumors, horizontal bars indicate mean, qRT-PCR.

4.2. The Ewing sarcoma cell line atlas (ESCLA)

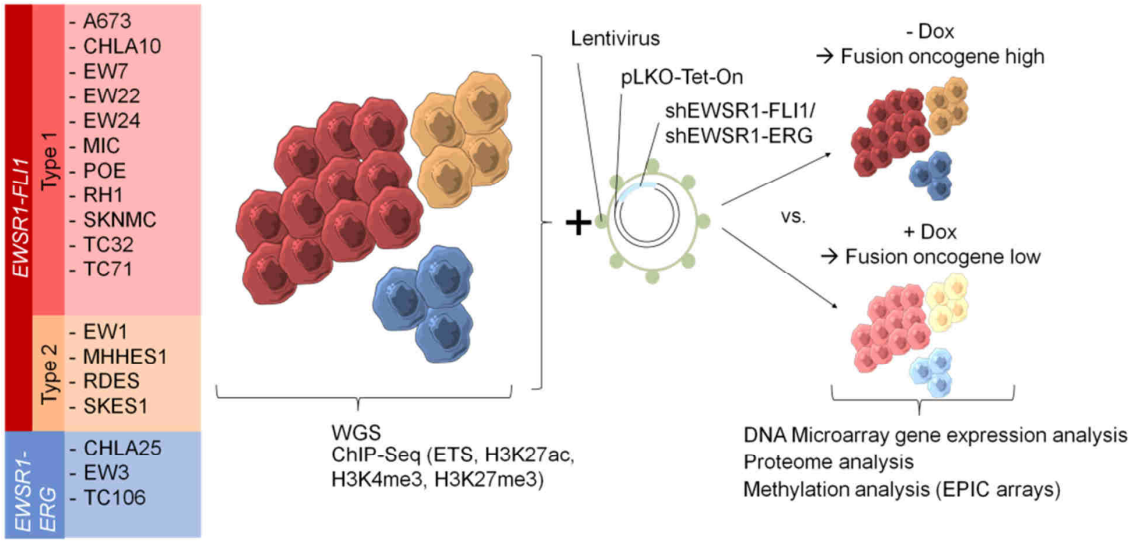

Figure 12: Components of the ESCLA

To achieve a comprehensive multi-dimensional dataset on the cell lines, which were used and generated here, whole genome sequencing (WGS), chromatin immunoprecipitation and sequencing (ChIP-Seq), DNA-micrarray transcriptional, proteome and methylome analyses were performed (Figure 12).

4.2.1. WGS

WGS data were generated on an Illumina HiSeq X ten platform with 150 bp paired end reads. The estimated coverage was 37.8x, and when testing in bins showed a rather equal coverage across the genome, except for spikes, which were mainly observed nearby centromeres, and dips in regions of common copy number losses (Figure 13A). The initial FASTQ files already showed good quality parameters in FastQC (Andrews, 2019), but successive alignment and quality control steps further improved the parameters, as seen when controlling the final aligned BAM files. Especially adapter sequences were reduced, GC content adapted and overrepresented sequences were not detectable anymore (Figure 13B). Noteworthy, the final quality parameters were, in contrast to the initial phred scores, not estimates by the sequencers, but the result of base quality score recalibration based on known genomic variants. For variant calling, this recalibaration results in scores that are more representative.

Based on the WGS data, rather low counts of single nucleotide variants (SNVs) and InDels, structural variants and copy number alterations (CNV) were observed (Figure 14A, left panel). The copy number variations comprised already described chromosome 1q, 8, and 12 gain in 5, 14 and 7 cell lines, respectively (27.8%, 77.8%, 38.9; Figure 14A, middle panel, exemplary karyogram indicating CNVs for EW24 harboring all the recurrent variations in Figure 14B). In addition, the known frequent non-silent *TP53* and *STAG2* mutations were observed (Figure 14A, right panel). To test for euploidy, the method of ploidyNGS (Augusto Corrêa dos Santos et al., 2017) was adopted, which is the comparison of frequencies of the most common and second most common allele at heterozygous loci. Outputs from VCMM applied on WGS data indicated InDels and the number of reads per genotype. The rates of reads per allele at the first 100,000 heterozygous loci were plotted in a stacked histogram. Mainly equal distribution of the first and second most common allele at heterozygous loci indicated diploidy (Figure 14C). In case of higher numbers of chromosome sets (n), different ratios would have occurred (2:1 in 3n, 3:1 in 4n cells).

As GGAA microsatellites (mSats) are known to be highly relevant in the interaction of EWSR1-ETS with the genome, potential GGAA mSats were searched in the reference

genome (yielding 8,311 potential loci), and were genotyped for all cell lines using HipSTR (Willems et al., 2017). 3,831 mSats (46.1%) were partially (≥6 cell lines), and 1,908 (22.3%) fully genotyped. As expected, high divergence in repeat numbers compared to the reference genome was observed. Noteworthy, mSats with 30-60 bp length appeared more often expanded than those with greater length, which might indicate limitations of HipSTR for genotyping mSats with more than 15 GGAA repeats. Nevertheless, mSats with maximum 24 consecutive GGAA motifs were genotyped (Figure 14D).

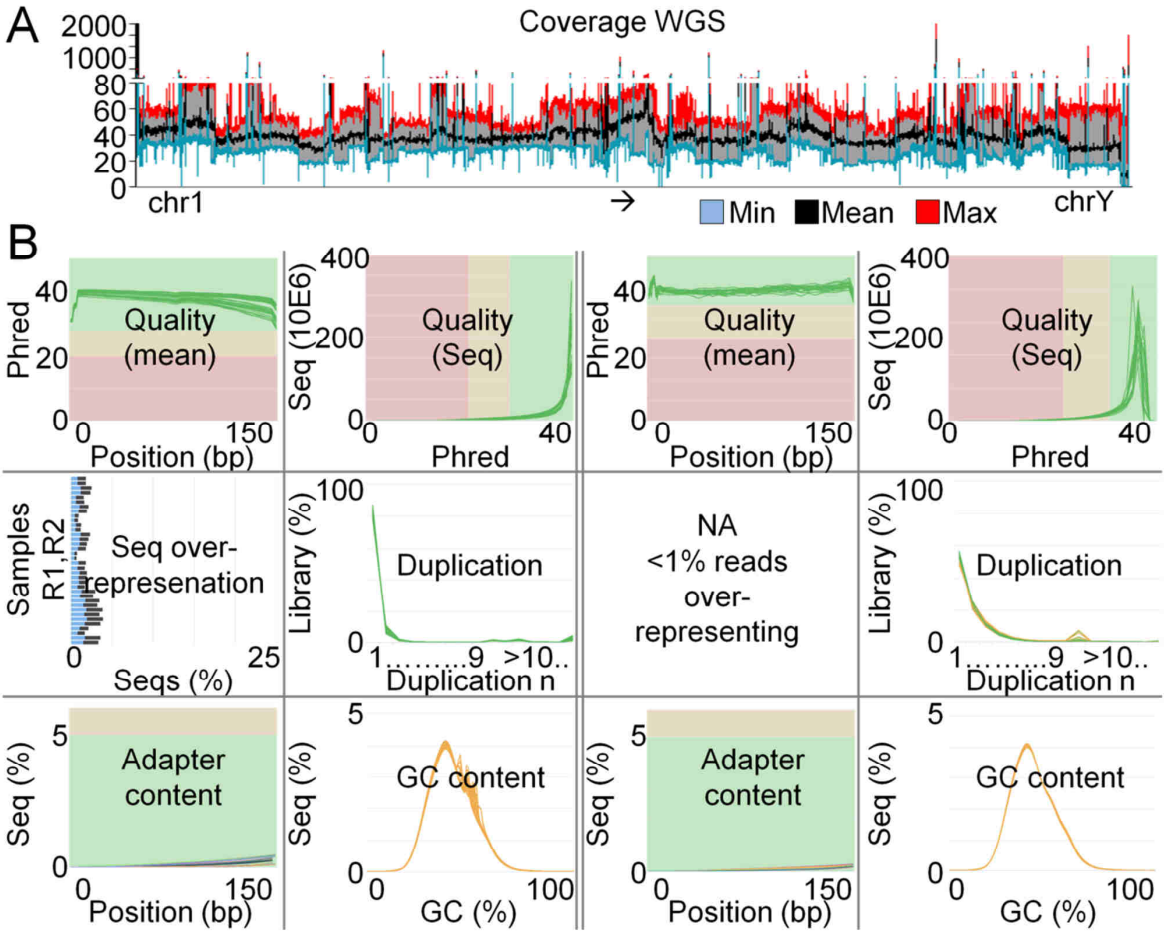

Figure 13: WGS resulted in >30x coverage and high-quality data. A) Coverage calculated across 100 kbp bins, depicted is the minimal, mean and maximum coverage per bin in all 18 cell lines. B) FastQC quality check of WGS data before alignment (left) and after alignment and further processing (right); Seq: sequences; R1, R2: first and second in paired reads.

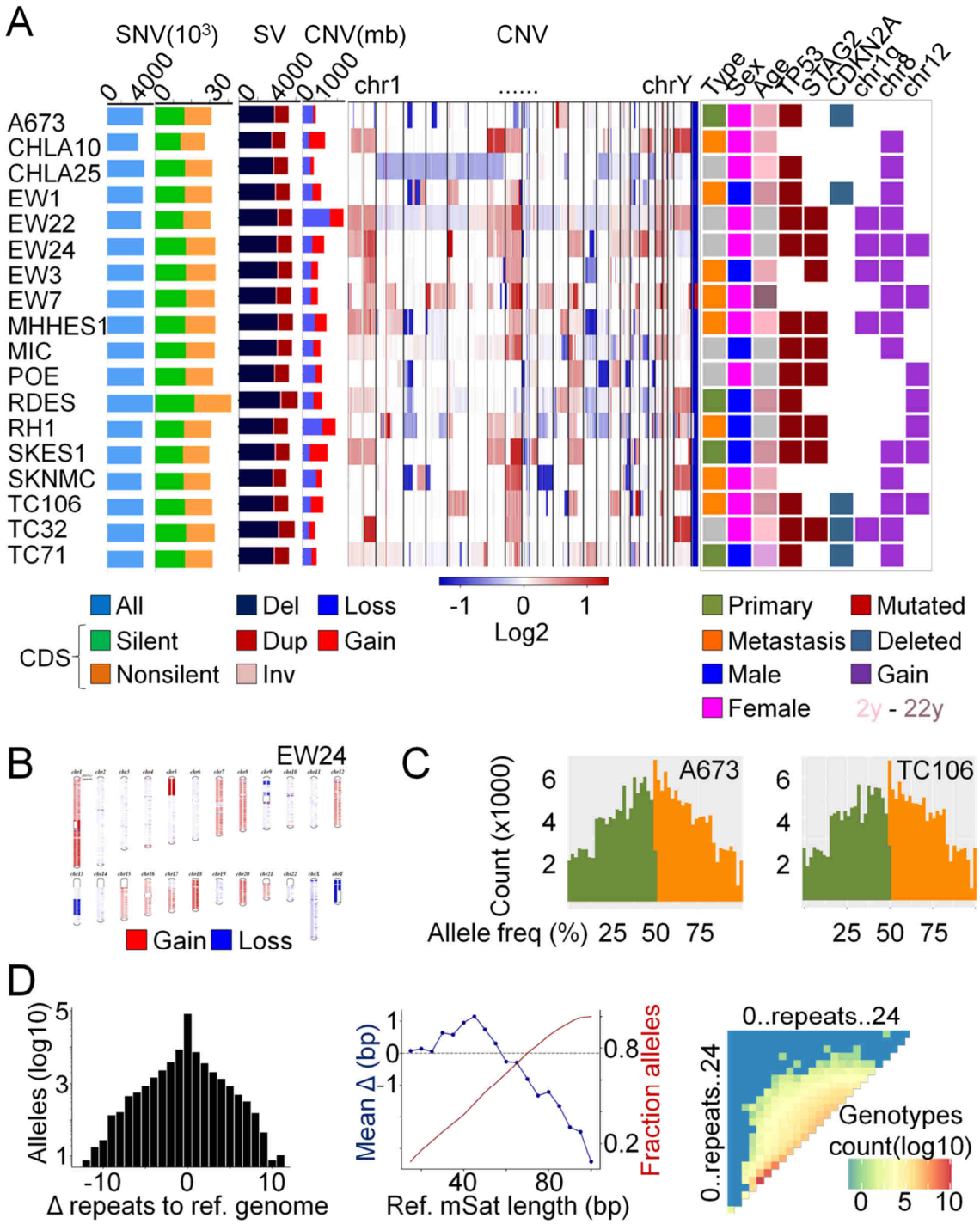

Figure 14: WGS data on EwS cell lines reveal the presence of known genetic variations and genetic silence in EwS. A) Genetic variations in EwS cell lines. Left: Bar plots indicating the number of SNVs, structural variants (SV) and CNVs (mbp deleted/duplicated) for each cell line; middle: heatmap indicating log2 copy number ratios across the genome for each cell line; right: tile plot indicating basic characteristics of the cell lines and summing up common genetic variations. B) Exemplary karyogram with colour coded report of CNVs in EW24. C) Exemplary histograms of A673 and TC106 indicating the number of reads supporting the most and second most common allele at heterozygous loci for InDels, indicating a nearly 1:1 distribution and thus euploidy. D) Quality indicators of HipSTR mSat genotyping. Left: Bar plot indicating number of alleles called with specific repeat number difference to the reference genome; middle: line graphs indicating the cumulative fraction of alleles (red) over reference mSat lengths and the mean length deviation of called mSats relative to the reference (blue); right: heatmap indicating counts of all observed genotypes described as numbers of consecutive GGAA repeats in the longest (vertical axis) and shortest (horizontal axis) allele.

4.2.2. ChIP-Seq

As no antibodies are available, which are specific for the fusion oncogene, FLI1 and ERG antibodies were employed to test for the fusion oncogene binding to the genome. Hence, the data on not-fused *FLI1*/*ERG* expression displayed in Figure 10 were reinspected, showing only for EW24 two-digit percentage of *FLI1* expression compared to the fusion (28.3%), and only for two other cell lines relative expression over 1% (A673 2.3%, RDES 1.7%), hinting to maximum small bias by the wildtype FLI1 in ChIP. For EWSR1-ERG positive cell lines, ERG expression ranges relatively to the fusion between 0.05% and <0.001%. The FLI1 antibody has been extensively tested for specificity by a collaborator (Didier Surdez, Paris, France), who saw strong decrease of ChIP product upon EF1 downregulation. As the ERG antibody was not yet described for this fusion oncogene, ChIP-PCR with and without knockdown of EErg in TC106shEE was performed, showing clearly a decrease in the fusion target site CCND1-EF1 and in the H3K27ac positive control sites CUL1, CDKN1A (Figure 15A), thus hinting towards antibody specificity. Interestingly, immunoprecipitation of the H3K4me3 promoter mark control site (CCND1-CTCF), which was normally only observed at low level in PCR for FLI1/ERG-ChIP, dropped by factor 10 upon EErg knockdown.

The number of ChIP peaks for the fusion per cell line ranged between 4,642 and 38,811 with one outlier on each extreme (1,253 peaks for TC32, 110,885 peaks for EW24, the only cell line, for which sequencing was performed with paired end reads). In sum, 156,092 distinct peaks for the fusion were detected (merging peaks with 200 bp width around the peak center). 280 consensus peaks for all cell lines were identified (Figure 15B). 82.1% of the consensus peaks overlapped with the previously published consensus EF1-binding sites for two EwS cell lines published by Riggi *et al* (Riggi et al., 2014). Also, nearly all other binding sites previously published in the consensus set were detected (1,499 of 1,555, 96.4%), but could not be confirmed as actual consensus sites, rather as "common" sites with on average 12 cell lines (67%) yielding peaks on those sites. In contrast, 50 new consensus sites were detected (Figure 15C). Described GGAA mSats known to interact with *NR0B1* and *MYBL2* (Beck et al., 2012; Musa et al., 2019) were covered in most cell lines. Interestingly, even after cross-sample peak height normalization, differential binding was observed between the cell lines (Figure 15D).

Most binding sites were located in introns and intergenic regions (Figure 15E); 2.1% (3,241 of 156,092) of sites were positioned at the locus of potential GGAA-mSat positions, for the consensus set even 65% (182/280), indicating a strong enrichment of GGAA mSats at the consensus versus all binding sites (P<0.0001).

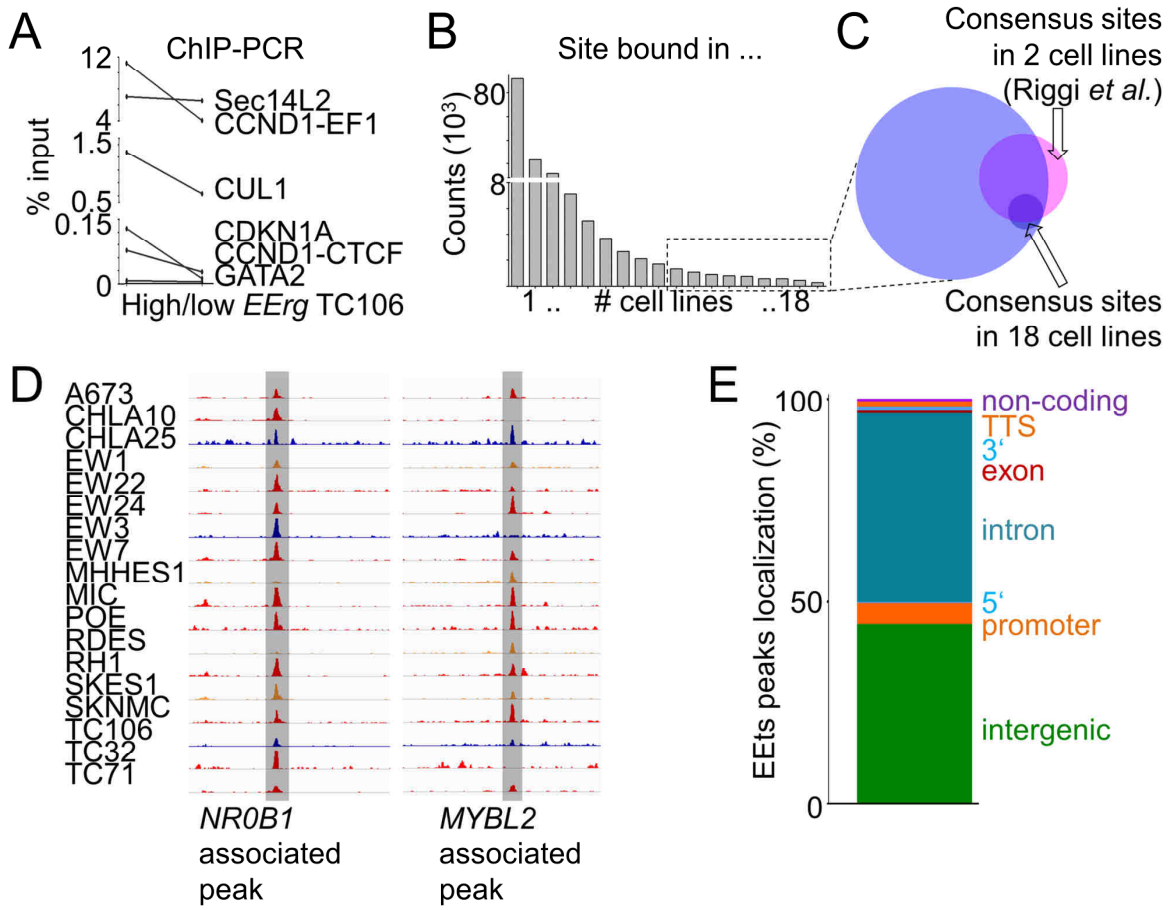

Figure 15: ChIP for fusion oncogenes supports previous reports and yields a new consensus binding set. A) ChIP-PCR results of TC106 cells in fusion high and low state, given as % of input, names of the sites targeted by primer pairs are indicated. B) Bar plot indicating the number of identified EEts-binding sites found to be actually bound in # cell lines. C) Venn-Diagram displaying the overlap of EEts-binding sites observed as actually bound in at least 50% of cells, the consensus binding site set described by Riggi *et al.* based on two cell lines, and the here generated consensus set. D) Exemplary plots of peaks at published EEts binding sites, NR0B1 and MYBL2 associated loci, for all cell lines of the ESCLA. E) Distribution of the localization of EEts-binding sites in context of genes; TTS: transcriptional termination site.

The median and average distance to the next transcriptional start sites (TSS) were 47 and 106 kbp, respectively. Fusion binding was most enriched closely upstream of TSS and was colocalized with H3K27ac histone mark (Figure 16A,B).

H3K27ac ChIP-Seq enabled super-enhancer calling (exemplary plot for TC71 in Figure 16C with highlighted super-enhancer sites nearby known *EEts*-regulated genes). 571 to 1,596 super-enhancers were called per cell line, and, when merging their positions over cell lines, 4,339 super-enhancer sites or clusters were detected. Only 58 of these 4,339 sites comprised super-enhancers of all cell lines. Noteworthy, 99.3% of super-enhancer sites overlapped with at least one EEts binding site, 53% with at least one site, where EEts binding was observed in half the cell lines or more. For the 58 super-enhancer sites shared

across the cell line, the overlap rate with those frequently EEts-bound sites was even 84.5% (significant enrichment, *P*<0.0001). Moreover, genes closest to and/or overlapping these 58 super-enhancers were significantly overrepresenting gene ontology terms in context of development, transcription and metabolism.

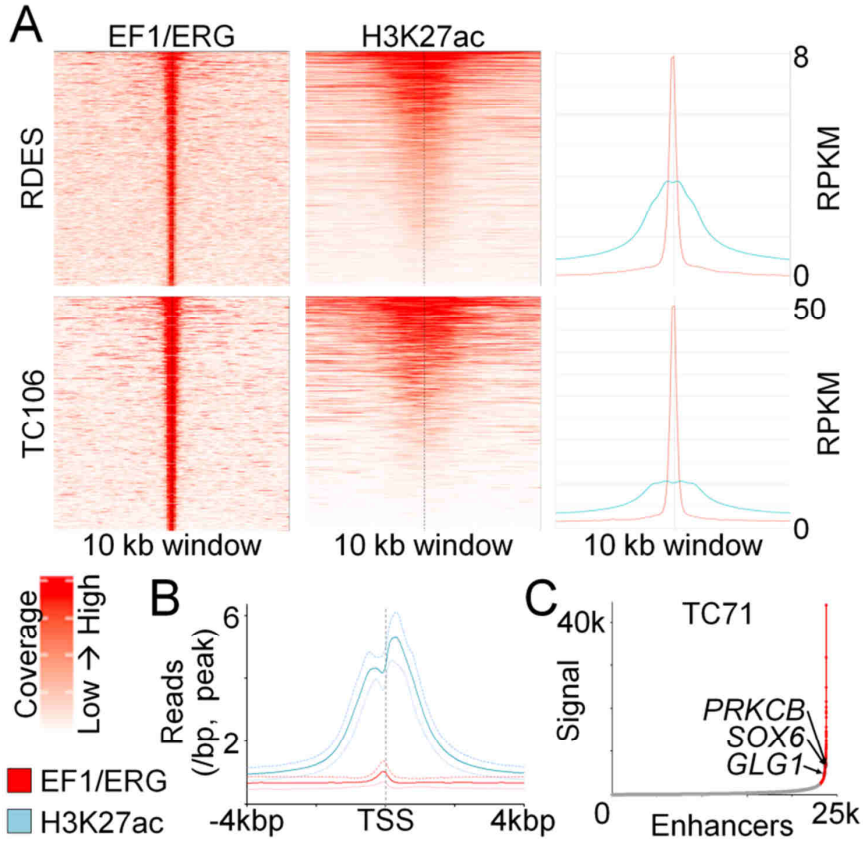

Figure 16: H3K27ac is observed around EEts binding and enables super-enhancer calling. A) Heatmaps for the coverage of EEts and H3K27ac ChIP in a 10 kb window around EEts binding sites (sorted from top to bottom: sites with peaks in 18 cell lines to 9 cell lines) for RDES and TC106, and metaplots combining both heatmaps in line graphs. B) Histogram of read depth for EEts and H3K27ac relative to TSS, displayed is the mean depth for all cell lines with the solid line, dashed lines indicate read depth plus and minus standard deviation. C) Plot of stitched enhancers and their signal value calculated by ROSE for TC71, super-enhancers are displayed in red, arrows point to super-enhancers nearby the listed *EEts*-regulated genes.

4.2.3. DNA-microarray transcriptome profiling

More than 25,000 transcripts coding for proteins were quantified before and after fusion knockdown in 18 EwS cell lines. Fusion knockdown resulted in strong deregulation of the transciptome (Figure 17A), when looking on the average expression fold changes among all cell lines. In line with previous publications, described EEts targets were affected by fusion knockdown in most or even all cell lines (Figure 17B). Gene set enrichment analysis (Subramanian et al., 2005) on a ranked list (average rank calculated for each gene across

cell lines) computed as proof-of-principle the "Kinsey_Targets_of_EWSR1-FLI1_Fusion_Dn" set as the strongest enriched set among genes upregulated upon fusion knockdown. Several gene sets associated with proliferation and cancer were enriched among genes downregulated upon fusion knockdown (Figure 17C). Overt EEts independent effects of Dox treatment or the transduction process were not observed so far in previous projects on EwS in the working group. t-SNE analysis of shControl transduced cell lines (same dose Dox for the same time; analysed on ClariomD DNA microarray) confirmed the detectable treatment effect on the transcriptome as minor compared to the effect achieved with fusion targeting shRNA (Figure 17D).

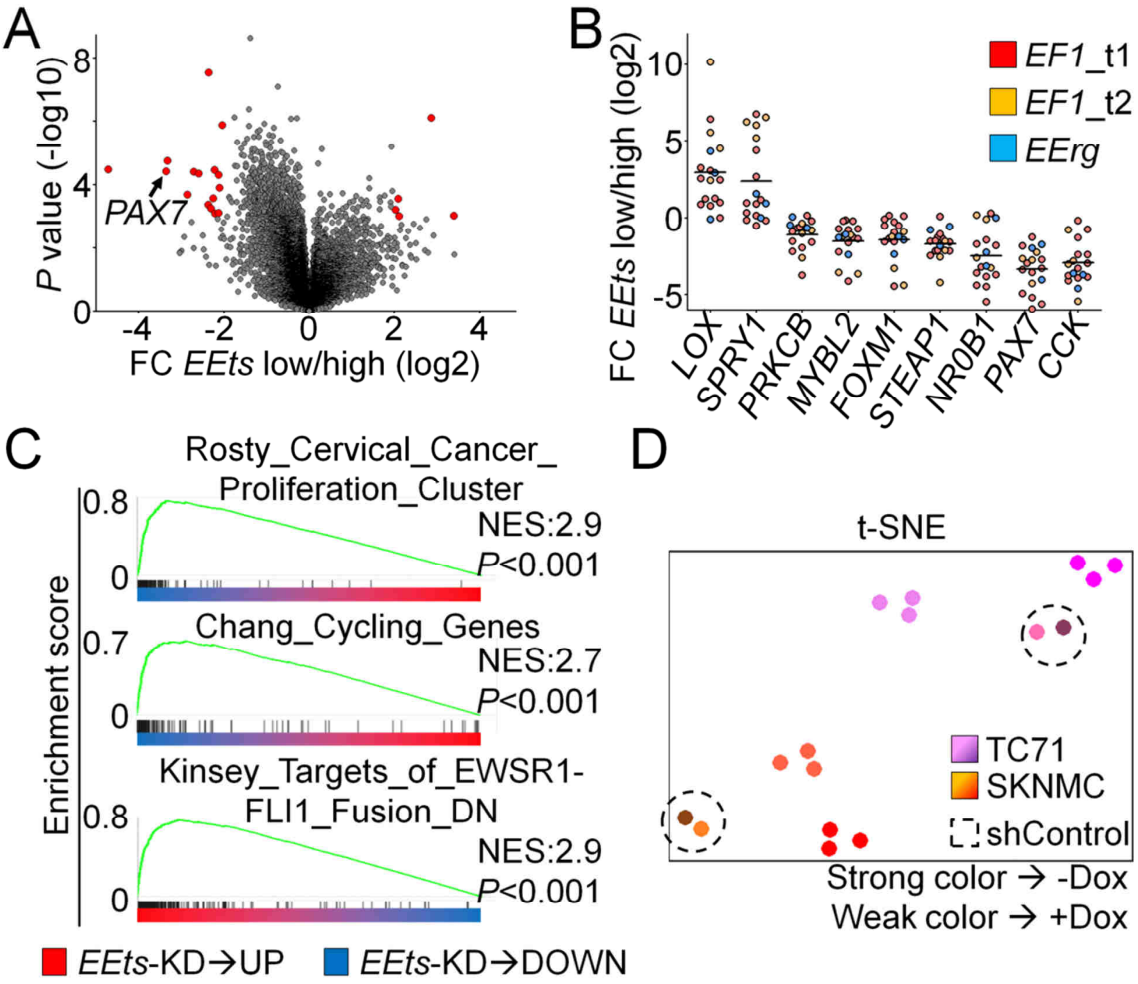

Figure 17: *EEts* knockdown leads to strong transcriptome alterations. A) Volcano plot indicating log2 fold change (FC) of individual genes (dots) upon *EEts* knockdown versus -log10 transformed *P* value for differential expression; genes with absolute FC>2 and *P*<0.001 are highlighted in red. B) Dot plot indicating log2 FC of the expression of listed genes upon *EEts* knockdown, each dot represents one cell line, the fusion type is color-coded. C) Exemplary GSEA graphs; NES: normalized enrichment score. D) t-SNE plot of transcriptome data from TC71shEF1, TC71shControl, SKNMCshEF1, and SKNMCshControl with (weak color) and without (strong color) Dox treatment, the shControl cells are circled and are much closer to each other than the treated versus untreated shEF1 cells.

To avoid an arbitrary cut-off for gene regulation, especially as different effect sizes of the fusion knockdown were expected due to different knockdown efficiency, a semi-quantitative approach was adopted from the ROSE super-enhancer call algorithm (Lovén et al., 2013; Whyte et al., 2013). In short, for each cell line all genes with either positive or negative fold-change were listed on the x-axis with their fold-changes on the y-axis, both axes were normalized to each other. From the point, where the slope equals 1, all genes were denoted as regulated. This resulted in 5,925 genes denoted as downregulated and 4,686 genes as upregulated in at least one cell line (605 to 1,455 downregulated, and 324 to 1,124 upregulated per cell line). Gene ontology analysis for downregulated genes yielded significant overrepresentations of several gene sets annotated with biological processes in context of DNA replication and cell cycle progression. Only four genes were downregulated in all *EEts*-low cells, indicating a uniform upregulation of these genes in EwS by the fusion oncogene, thus bearing potential as biomarkers for EwS. These genes were namely *FEZF1*, *TRPM4*, *PAX7*, and *PPP1R1A*. On the other hand, not any coding gene was upregulated in all *EEts*-low cell lines, only a long non-coding RNA, *DNM3OS*, known in the context of epithelial to mesenchymal transition (Mitra et al., 2017). However, already four genes (*ATP9A*, *FSTL1*, *NRCAM*, *SRGAP1*) were upregulated in 17 of the 18 cell lines, and another 11 in 16 cell lines, indicating a rather uniform suppression in EwS of these genes, which might hint towards incompatibility with EwS growth/survival and could give inroads into new therapeutic targets.

To test for an actual relevance of the apparent transcriptome rearrangement by the fusion oncogene, survival data of 196 EwS patients were compared with the transcriptome data. 12,812 genes were represented in both datasets. Without correction for multiple testing and applying best-percentile method, either high or low expression of 5,723 genes was associated with survival (P<0.05; 44.7%). When only looking on the 6,132 genes, which were in both datasets and denoted as fusion regulated, 3,003 genes were associated with survival (49%), meaning a significant enrichment of survival associated genes among targets of the fusion knockdown (P<0.0001). When comparing the here generated transcriptome data with data from the DepMap project at the Broad Institute (Tsherniak et al., 2017) for seven EwS cell lines present in the ESCLA, again a significant enrichment of likely essential (score at least -1 for all cell lines) genes was observed in the *EEts* regulated genes (P=0.01).

As the pathognomonic *EEts* fusion is the major driver of EwS, it was tempting to speculate, that *EEts* kockdown leads to a change in the transcriptome profile bringing the cells closer to the cell of origin. However, t-SNE analysis with 71 normal tissues (Baldauf et al., 2018b) showed EwS cells in *EEts*-high and -low state clearly clustered with each other and

separated from the normal tissues (see overlap of *EEts*-high (dark red) and *EEts*-low (dark blue) in center of Figure 18).

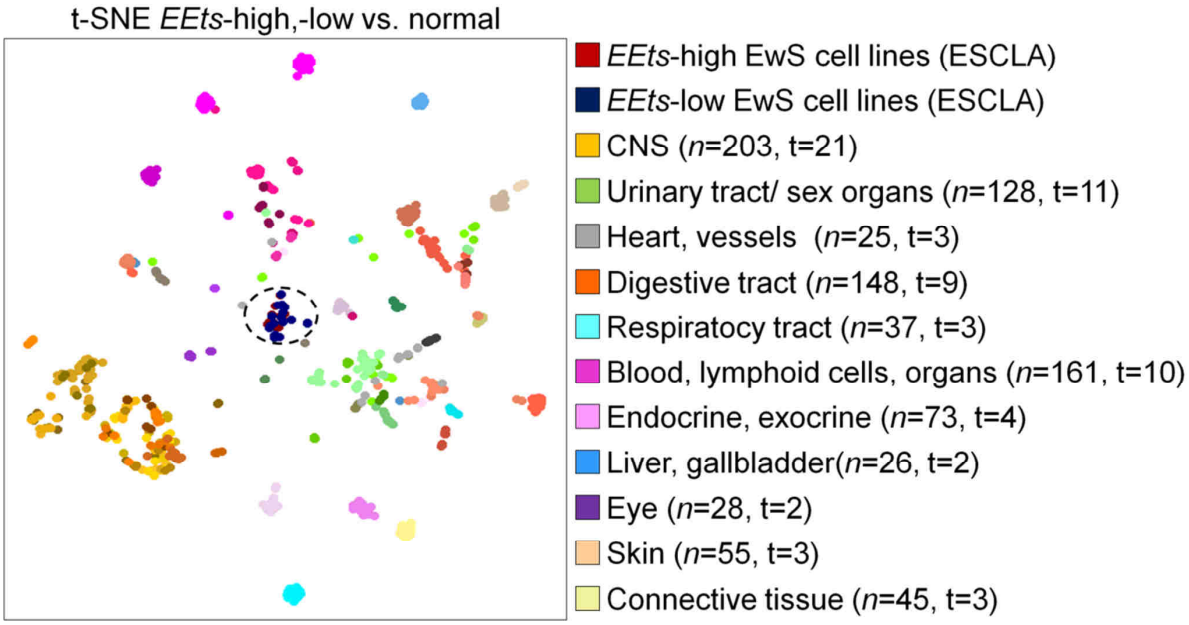

Figure 18: t-SNE plot of EwS cell lines (center, circled) in *EEts*-high and -low state does not show any alignment of the EwS transcriptome with any of 71 normal tissues. 71 normal tissues are represented by 11 groups. CNS: central nervous system; *n*: number of samples; t: number of tissue types per group.

4.2.4. Proteome analysis

Protein quantification by mass spectrometry in all cell lines was successfully performed for 3,248 proteins. Additional 1,336 proteins with scarce data across the samples were imputed, resulting in 4,584 proteins quantified for each cell line.

The proteins showed, as the transcriptome, heterogeneous dynamics by fusion oncogene knockdown, even if less pronounced (Figure 19A). Actually, the protein levels were strongly and significantly correlated with the expression data on transcriptome level (Pearson's r=0.58, P<0.0001, Figure 19B).

In line, GSEA on the proteins ranked by their fold-changes (rank calculated for each protein in each cell line, mean rank chosen for building the ranked list) yielded association with translation and, again, cell cycle (Figure 19C). When defining regulated proteins as done for transcriptome data (see above), cell cycle genes were again overrepresented among the downregulated proteins in gene ontology analysis.

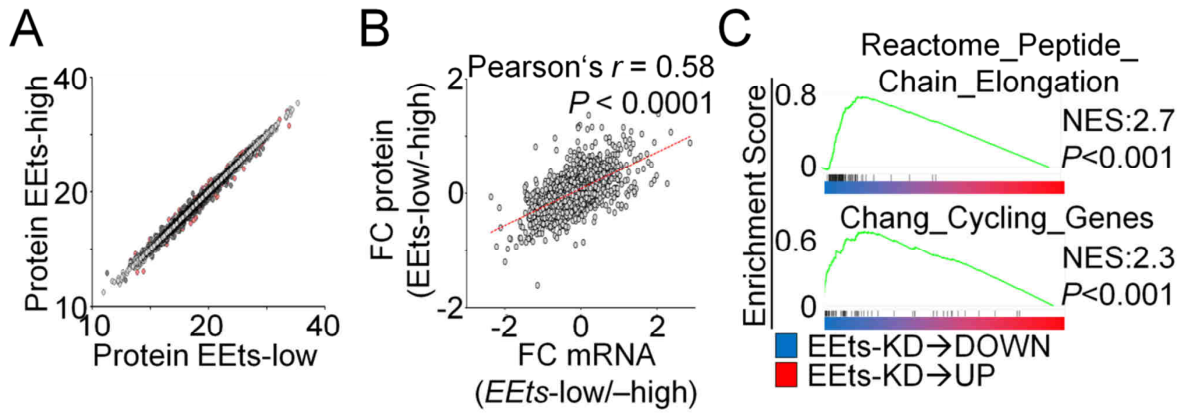

Figure 19: The proteome is similarly affected by fusion knockdown compared to the transcriptome. A) Dot plot indicating protein level in EEts-high versus -low cells, red dots highlight significantly differentially expressed proteins. B) Dot plot indicating correlation of fold changes observed on the transcriptome leven and on the protein level for all genes, whose expression was quantified in both data dimensions; dotted red line indicates linear regression curve. C) Exemplary GSEA graphs; NES: normalized enrichment score.

All these data hint towards a strong association of the transcriptome and proteome, and thus to transcriptional regulation as key factor for protein expression in EwS rather than post-transcriptional processes.

4.2.5. Methylome analysis

Differential methylation of CpG island in EEts-high versus -low state was assessed with Illumina EpicArrays and data processing with the minfi package (bumphunter algorithm) (Aryee et al., 2014). While for each cell line regions of differential methylation upon fusion knockdown were identified (383 to 18,127 regions, median 1102.5; 27 to1,322 denoted as significant; Figure 20), calling failed when processing all cell lines together. In line, intersection testing for the differentially methylated regions (DMRs) yielded not any region modulated in all cell lines by EEts knockdown, only one modulated in 12 cell lines, and 26 modulated in at least 50% of cells. For 23 of these 26 regions transcriptome data of the closest gene were available and revealed EEts-dependent expression regulation for 11 genes. However, when testing for coincidence of expression regulation and identification of a nearby DMR in individual cell lines, no significance existed. Nevertheless, when testing EwS cell lines transduced with shControl construct and with/ without Dox treatment, less differentially methylated regions were called (78.3%, and 10.4%, for A673 and TC32, respectively), indicating that some detected sites are due to a real effect of the fusion oncogene, which is seemingly highly dependent on cell context.

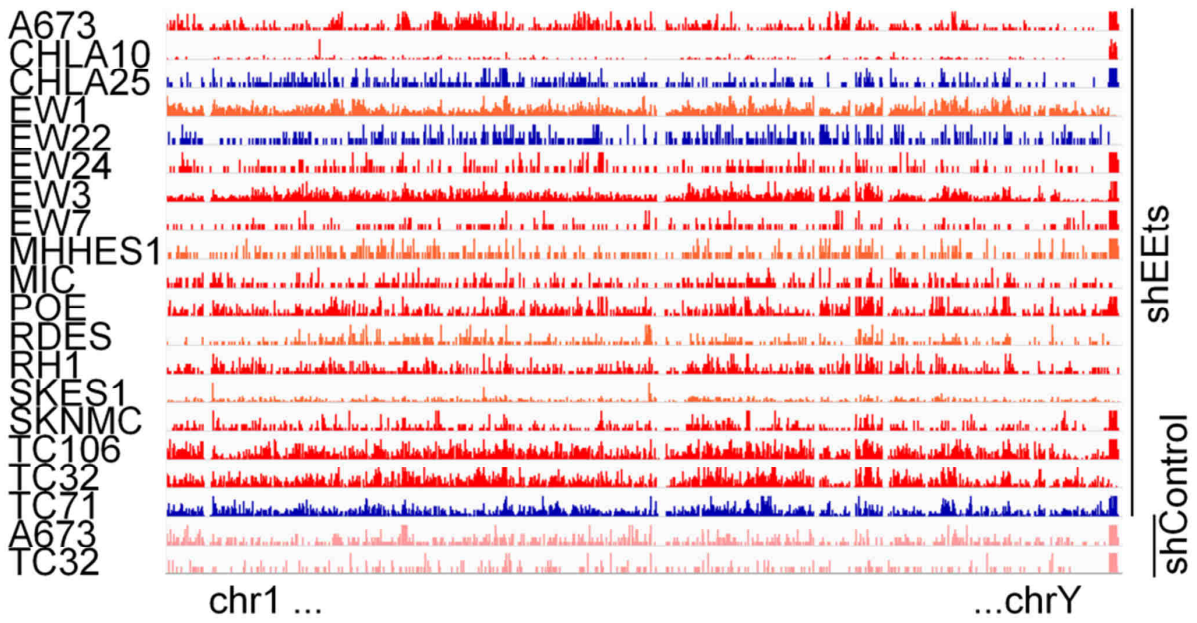

Figure 20: Differentially methylated regions were identified across the genome, but without inter-cell line consistence. IGV track plot. Different fusion types and shControl cells are represented by different colors. Methylation-differences were called between conditions EEts-high and -low.

In sum, the ESCLA includes data on a relatively large number of cell line models, is multidimensional and comprehensive for all cell lines, of good quality in each dimension, and displays genome wide alterations, fusion binding and histone marks, and transcriptome, proteome and methylome rearrangement upon fusion knockdown (Figure 21).

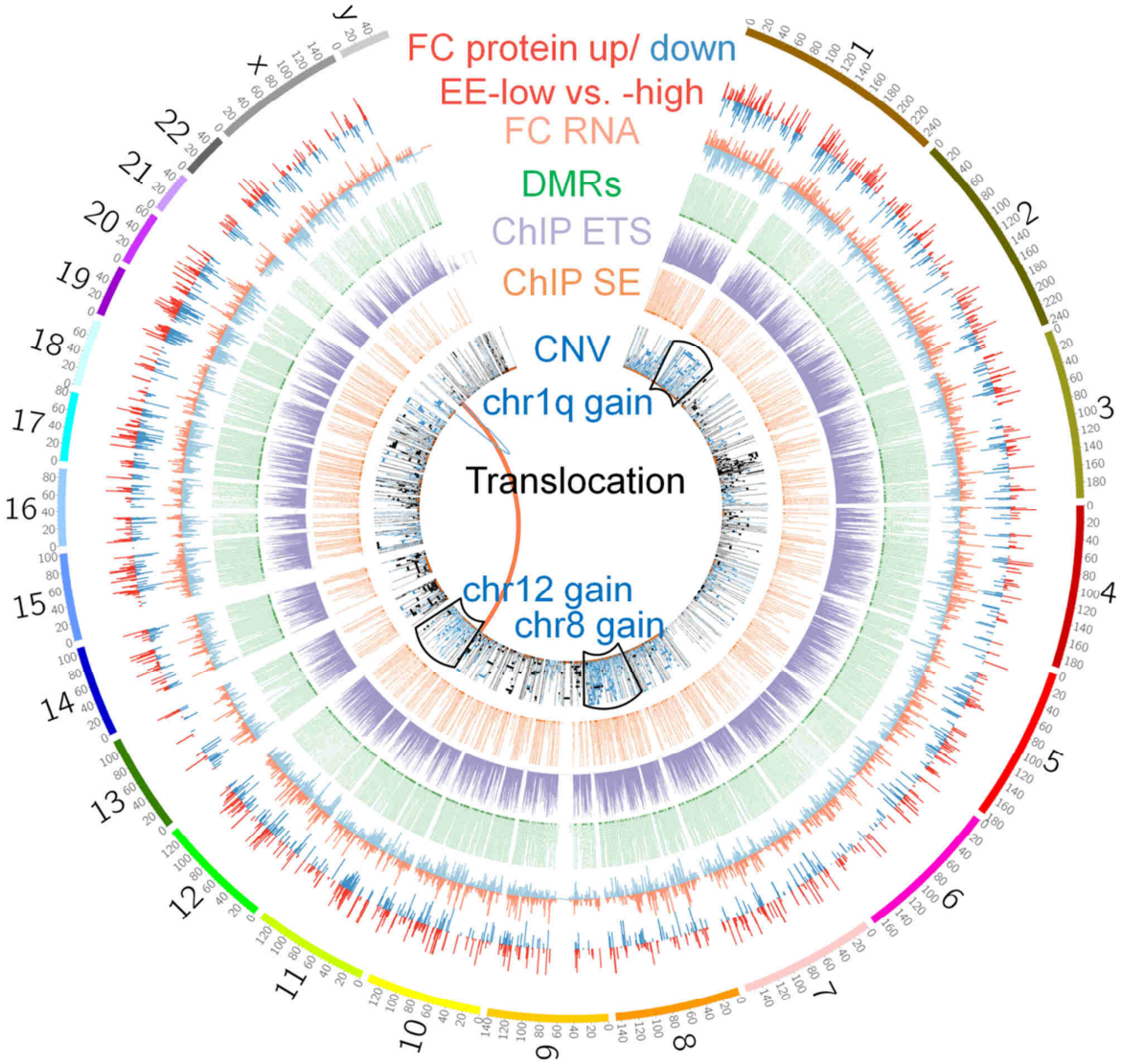

Figure 21: Circos plot displaying various aspects of the ESCLA, as genome wide proteome, transcriptome and methylome rearrangement upon EEts modulation, EEts binding sites and super-enhancers, and CNVs. DMR: Differentially methylated region; SE: super-enhancer; CNV: positions of gains are represented in blue, losses in black.

4.3. Applicability of the ESCLA as model for EwS

Research on cellular mechanisms and functional experiments are based on experimental models like immortalized cell lines. The quality of each model has to be carefully considered before translating any conclusion made on a model to clinics.

Thus, similarities of the cell line models in the ESCLA to primary EwS tumors were assessed. Additionally, potential differences between cell line models with distinct fusion

types were investigated as indicator whether the data can be used in conjunct analyses or demand subgrouping.

4.3.1. ESCLA transcriptional and methylation profiles are close to, but not overlapping with primary EwS

Curated transcriptome data of 50 primary EwS and 11 further morphological mimics were present (Baldauf et al., 2018b). The data were corrected for batch effects together with the ESCLA data and t-SNE was performed. Interestingly, EwS cell lines did not merge with the primary tumors, but were also clearly separated from other (differential) diagnosis (Figure 22).

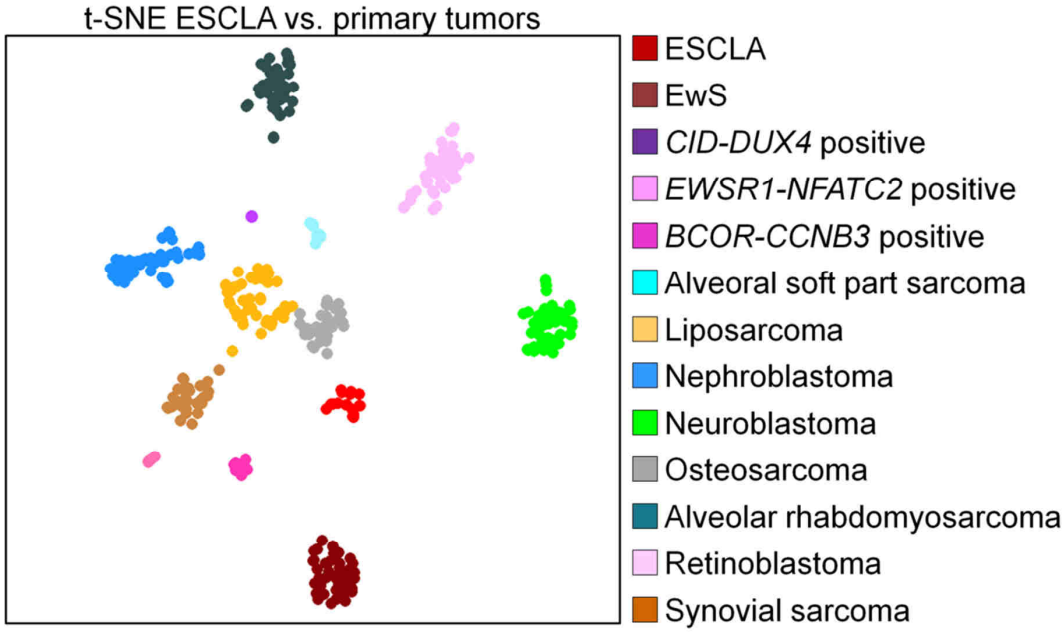

Figure 22: t-SNE plot of EwS cell lines (ESCLA), primary EwS and other sarcomas does not show transcriptional EwS identity for cell lines.

The MolecularNeuropathology.org platform provides prediction of sample identity by the methylation profile (Capper et al., 2018). For six cell lines, the methylation profile was calculated as being closest to EwS (CHLA10, CHLA25, MHHES1, MIC, RDES, TC32), while for the remaining cell lines the methylation profile was assigned to a bin of unclear identities. Noteworthy, for the cell lines with methylation profile closest to EwS, the metric was smaller for EEts-low samples compared to EEts-high samples. Additionally, a t-SNE plot with the same data, kindly provided by Dr. Martin Sill (DKFZ, Heidelberg, Germany), displayed close vicinity of the cell line methylation profiles with EwS profiles, but no overlap (Figure 23).

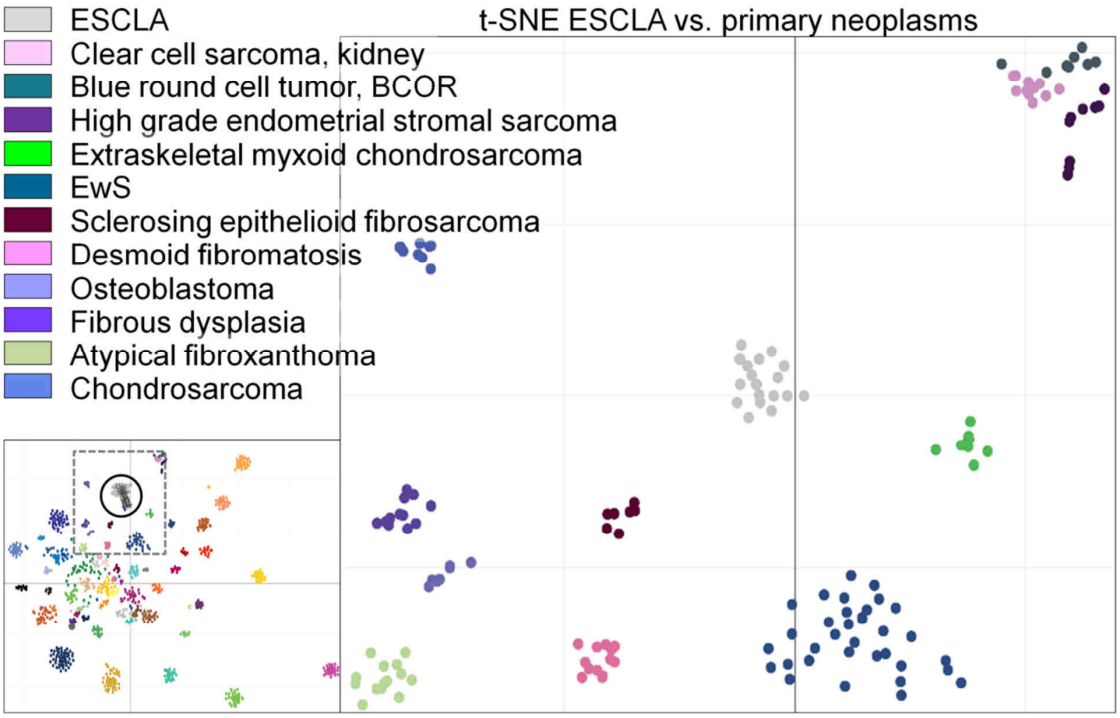

Figure 23: t-SNE of the ESCLA and other neoplasms indicates distinct methylation profile of EwS cell lines. Left-down: overview of the complete t-SNE, ESCLA data in circle, dashed lines are borders for the excerpt. Right: excerpt of the t-SNE plot around ESCLA cell lines.

In sum, cell lines resemble genomic variations observed in primary tumors (see Figure 14A), but depict an only similar, not identical transcriptional and methylation profile. Hence, any results drawn from the ESCLA are likely to resemble actuality in EwS, but require extensive validation in further models and potentially other models closer to the actual EwS biology (e.g. patient derived xenografts).

4.3.2. Distinct fusion types do not strongly differ in gene expression, expresion regulation, super-enhancers and methylation

Differences of EwS with distinct fusion types were often discussed in the context of prognosis, but poorly investigated in cell line models. Actually, high similarities in the biology and effects of the different fusions are a prerequisite to transfer results of studies based on only few fusion types, most often solely *EF1* type 1, to all EwS, and to perform joint analyses as done here.

Therefore, EwS cell lines with different fusion types were compared as to their transcriptional profile of all regulated genes in t-SNE analysis, which displayed no separation or clustering of any fusion type from others (Figure 24A). Next, the regulated genes of each individual cell line were compared with those of all other cell lines. As a minimum 605 and 324 genes were

down- and upregulated, respectively, per cell line, for each cell line the top 603 and 321 down- and upregulated genes (in total 924) were selected as the comparison dataset. To avoid low overlap rates due to false positives, the 33% of the selected genes with highest fold changes (most likely truly EEts regulated, *n*=308) of one cell line were compared to the the 924 top-regulated genes of each other cell line. All comparisons showed far higher overlap than the rate expected by chance (3.7%). Only *EF1* type 2 and *EErg* positive cells showed a higher overlap of regulated genes in their respective cohort than compared to cells of other fusion types (Figure 24B). As *EF1* type1 positive cells had nearly the same overlap with other type 1 positive cells as with type 2 positive cells, and type 2 positives shared similar numbers of genes with type 1 and *EErg* positive cells, likely all cell lines comprise a consensus set of often and strongly regulated genes, but *EF1* type 2 and even more *EErg* fusion seemingly results in additional fusion-type private regulated genes.

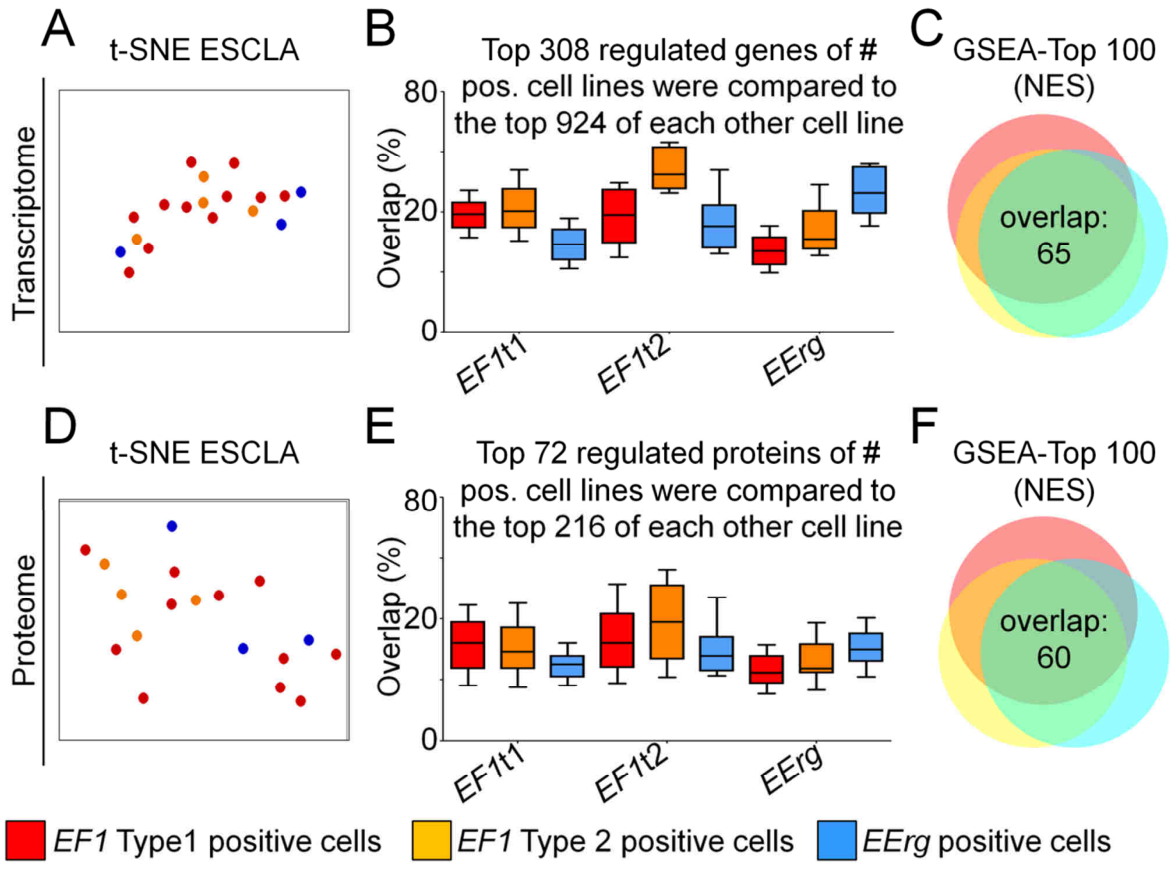

Figure 24: EwS cell lines with distinct fusion types do not differ in expression profile and regulation on the transcriptome and proteome level. Upper panel: results for transcriptome. Lower panel: results for proteome. A/D) t-SNE of the expression profile of EEts-high cell lines, distinct fusions are color-coded. B/E) Box plot indicating overlap of top-regulated genes in cell lines of specified fusion type with regulated genes in other cell lines, line represents median, box interquartile range, whiskers 10th-90th percentile; *EF1t1/2 EF1* type 1/2. C/F) Venn diagram of top 100 (NES as metric) gene sets correlated with genes downregulated upon EEts knockdown.

In line, among the top 100 enriched gene sets correlated with genes downregulated upon EEts knockdown (likely EEts driven) in GSEA, separated by fusion type again applying ranked lists with mean rank for the regulation of each gene, an overlap of 65% was observed (69-85% overlap for two individual fusion types; Figure 24C). Leading edge analysis of the fusion-specific called gene sets did not reveal any genes, which were present in more than 50% of the fusion-specific sets and exclusively regulated by the respective fusion type, except for two proteasome subunits in *EF1* type1 cells (*PSMA2*, *PSMD8*).

Highly similar results were achieved on the protein level (overlaps for GSEA 60%, all fusion types, 65-72% for two individual fusion types; Figure 24D,E,F).

As super-enhancers are considered to define cell identity, commonalities in called super-enhancers were also investigated. Because of high diversity in the number of called super-enhancers, and the lowest number being 571, only the top 500 super-enhancer regions were tested for overlap. It has to be noted, that super-enhancers span several thousand base pairs and were identified individually per cell line. Hence, there might be a relatively high rate for random overlaps. This is why here all 500 top super-enhancers were compared to all others (in contrast to similar analyses for transcriptome and proteome) to avoid that a few random hits might be sufficient to corrupt the results. Similar to the results on regulated genes, nearly uniform rates of commonality between the cell lines independent of fusion type were observed (Figure 25A).

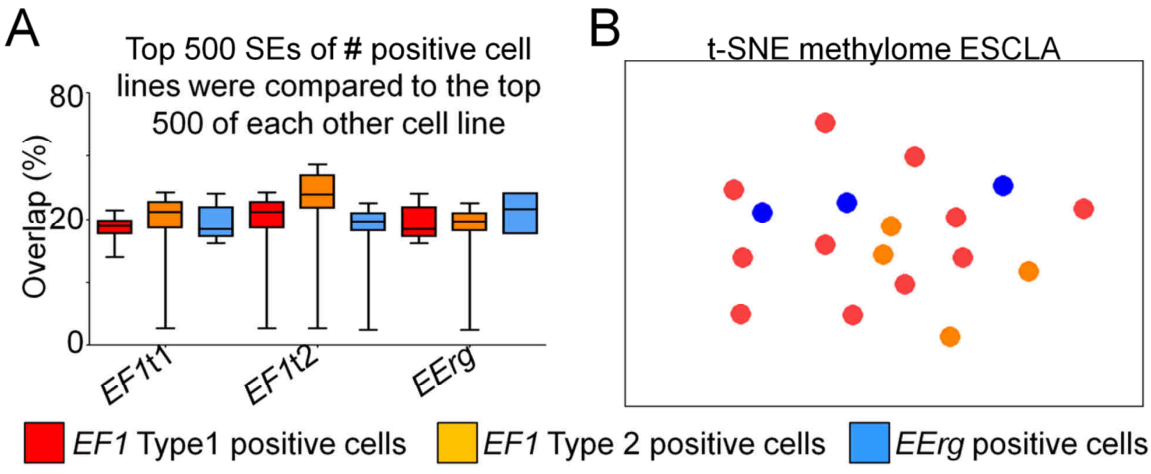

Figure 25: EwS cell lines with distinct fusions share several super-enhancers and similar methylome profiles. A) Box plot indicating the overlap of super-enhancer (SE) regions in one cell line compared to all others, separated by fusion; box indicates interquartile range, whiskers 10th to 90th percentile; EF1t1/2 EF1 type 1/2. B) t-SNE for the CpG methylation in the ESCLA; cell lines harboring different fusion types are color-coded.

Moreover, when comparing the CpG methylation profiles (more than 800,000 sites) of the different cell lines, again no clustering and separation of any fusion type was observed (Figure 25B).

A possible explanation for the observed differences of *EErg* positive cell lines might be that the fusion is always generated by chromoplexy, which affects also other genes. Other fusion types can be the result of simple reciprocal translocation. Under this aspect, the number of *EF1* type 1 and 2 positive cell lines with fusions by reciprocal translocation versus chromoplexy was assessed. As 52% of *EF1* positive tumors with reciprocal translocation express the second fusion product (Anderson et al., 2018), *FLI1-EWSR1*, this reverse fusion was probed with two primer pairs (Elzi et al., 2015) in PCR on cell-line cDNA. For *EF1* type 1 positive cell lines both primer pairs were expected to bind, for type 2 positive cells only one. In line, two lanes were found for A673 and POE, single lanes for MHHES1 and RDES (Figure 26A). BreakDancer (Chen et al., 2009) readout of the whole genome sequencing detected reads supporting the *EF1* fusion on the plus and minus strand of *EWSR1* in these cell lines, what can only be observed in case of reciprocal fusion (Figure 26B). The same was seen for EW1, EW7, MIC, SKES1, and TC32. For RH1 and TC71 the highest confidence score for the one-sited detected fusion hinted toward non-reciprocal translocation. ChainFinder (Baca et al., 2013) for detection of chromoplexy revealed rearrangement loops forming the fusion in EW24 and SKNMC (Figure 26C) and an *EWSR1* rearrangement in EW22. Although for the remaining cell lines the fusion development was not displayed by chainfinder, it is likely that once chromoplexy occurred in these cells as other rearrangement loops were detected. In consideration of standard usage of chainfinder with multi-step healthy control filtered samples, the achieved output once more corroborated the data quality of the ESCLA. Considering also CHLA10, the only cell line without any high confidence results for chromoplexy or reciprocal translocation, as product of chromoplexy, the rates of chromoplexy as developmental process of the fusion oncogenes in the ESCLA were 55, 0 and 100% for *EF1* type 1, 2 and *EErg*, respectively (Figure 26D).

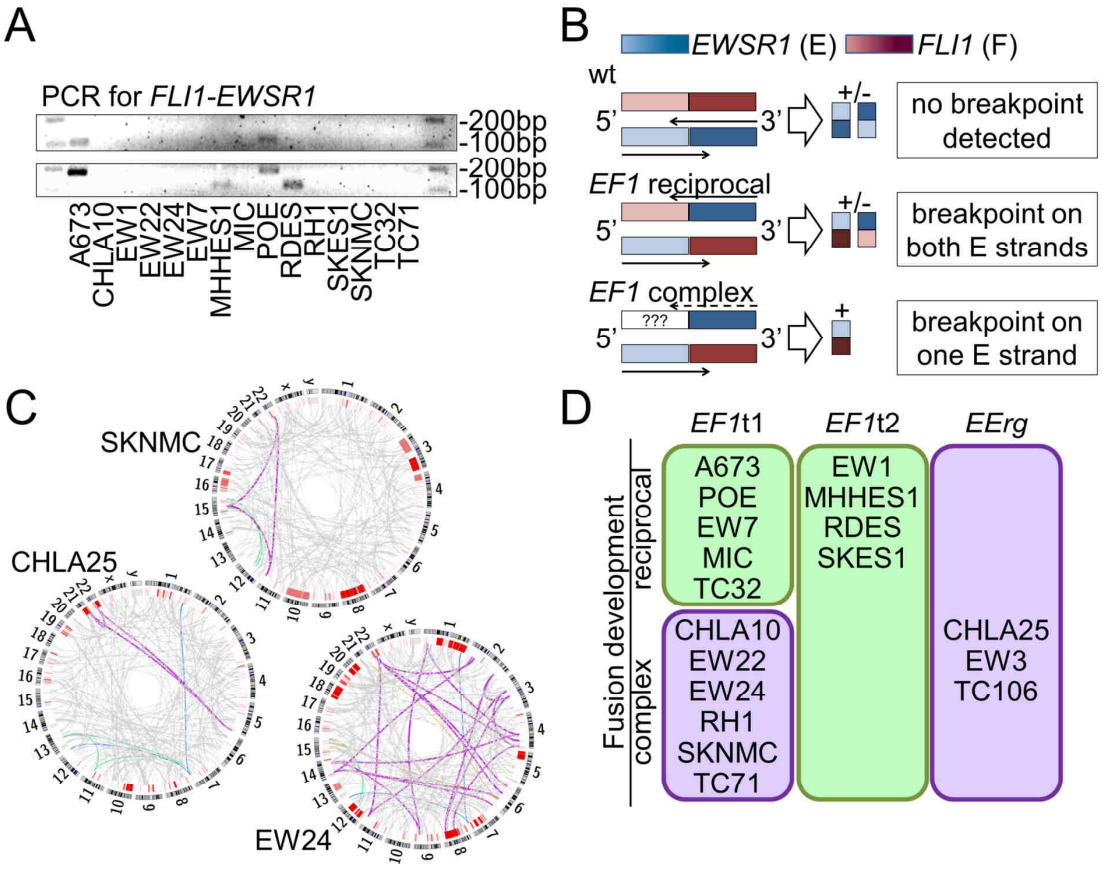

Figure 26: Cell lines with distinct fusion types in the ESCLA differ in their rate of chromoplexy. A) PCR product gel electrophoresis for the reverse *EF1* fusion. B) Scheme explaining the identification of reciprocal *EF1* fusions by sequencing reads on the plus and minus strand of *EWSR1* extracted by BreakDancer. C) Circos plots indicating rearrangement loops in EwS cell lines, loops affecting or generating the fusion oncogene are highlighted in purple. D) Overview of cell lines with reciprocal and complex fusion development per fusion type; *EF1t1/2*: *EF1* type 1/2.

4.4 Parameters affecting gene expression regulation by EWSR1-ETS

As the proteomics data were strongly correlated with the transcriptome data, but the latter were by far more extensive, proteomics data were neglected for further analyses. Equally, the missing concordance of differential CpG island methylation upon fusion knockdown suggested not any strong uniform effect of those methylation islands and the data were, hence, not integrated in the further analyses.

The number of GGAA repeats in GGAA mSats has often been described as determining factor of EEts-mediated enhancer activity, but has only been tested at a few loci and experimentally validated in an artificial mono-allelic approach (reporter assays). Of 8,311 potential GGAA-mSat loci, 3,377 loci were partially, 251 of them even always bound by EEts in ChIP-Seq analysis. Of note, this numbers differ from the previously mentioned count of

EEts-binding sites harboring GGAA mSats, as one binding site can cover more than one mSat when they are closely adjacent, and as the binding sites were inferred from peaks in all cell lines with 200 bp width around peak center, while here each cell line and full-sized peak was probed individually. As mentioned before, HipSTR genotyped 3,831 mSats, including 844 of the partially and 152 of the always EEts-bound mSats.

When combining the ChIP-Seq and HipSTR data, the average and maximum number of consecutive GGAAs of both alleles per locus correlated, indeed, with the rate of observed EEts binding (Pearson's r^2=0.98 and 0.99, P<0.0001; Figure 27A). As expected, the increase of the EEts-binding rate over the count of maximum consecutive GGAA repeats of both alleles had a slight latency when compared with the average count, indicating that long consecutive GGAA mSats on both alleles increase the probability of EEts-binding detection in ChIP compared to the presence of only one allele with such a long mSat. Only for mSats with up to 18 consecutive repeats, mean or maximum per locus, more than 100 genotypes were called. Hence, for longer mSats the number of observations was rather scarce and no conclusion were made for those. Interestingly, although the absolute number of genotyped GGAA mSats dropped at first dramatically with increasing consecutive-motif numbers, there was a second maximum at about 12 to 14 motifs (mean and maximum allele length), which might represent an enrichment of such longer and more often bound mSats in EwS.

However, as the results were not binary and not any steep increase of binding rates was observed at a specific GGAA-repeat length, likely additional factors influence, whether EEts binds or not. The tested mSat regions were selected by an algorithm with Tandem Repeats Finder. Of course, these mSats had recurrent motifs, but these could also be interspersed with other bases (Figure 27B). To account for the number of additional GGAA motifs and interspersed bases besides the longest GGAA stretch within the called mSat-region, henceforth referred to as mSat architecture, their potential as enhancers or suppressors of EEts binding was tested. As metric the rate of finding a mSat with a specific architecture bound versus not bound was used, normalized by the same metric of a mSats with the identical number of consecutive GGAA motifs, but without any further bases called. As for heterozygous loci an assignment of binding to one of the alleles was not possible, the analyses were carried out solely on homozygous mSats. Furthermore, to build ratios as metric, only mSats at least once observed as bound and not bound were investigated. Additionally, mSats with a specific architecture observed less than 8 times (50% of the average count of all bound and not bound mSats with distinct architectures) were not included to avoid strong effects of seldom and possibly random observations. The metric was normalized to 1 as value for no preference for EEts binding. Applying all these filters to reduce bias, mSat architectures were investigated for the number of bases flanking the

longest GGAA stretch before another GGAA motif, the average number of interspersed bases between additional GGAA motifs, the number of additional GGAA motifs, the longest consecutive repeat number among them, and the number of GGAA stretches.

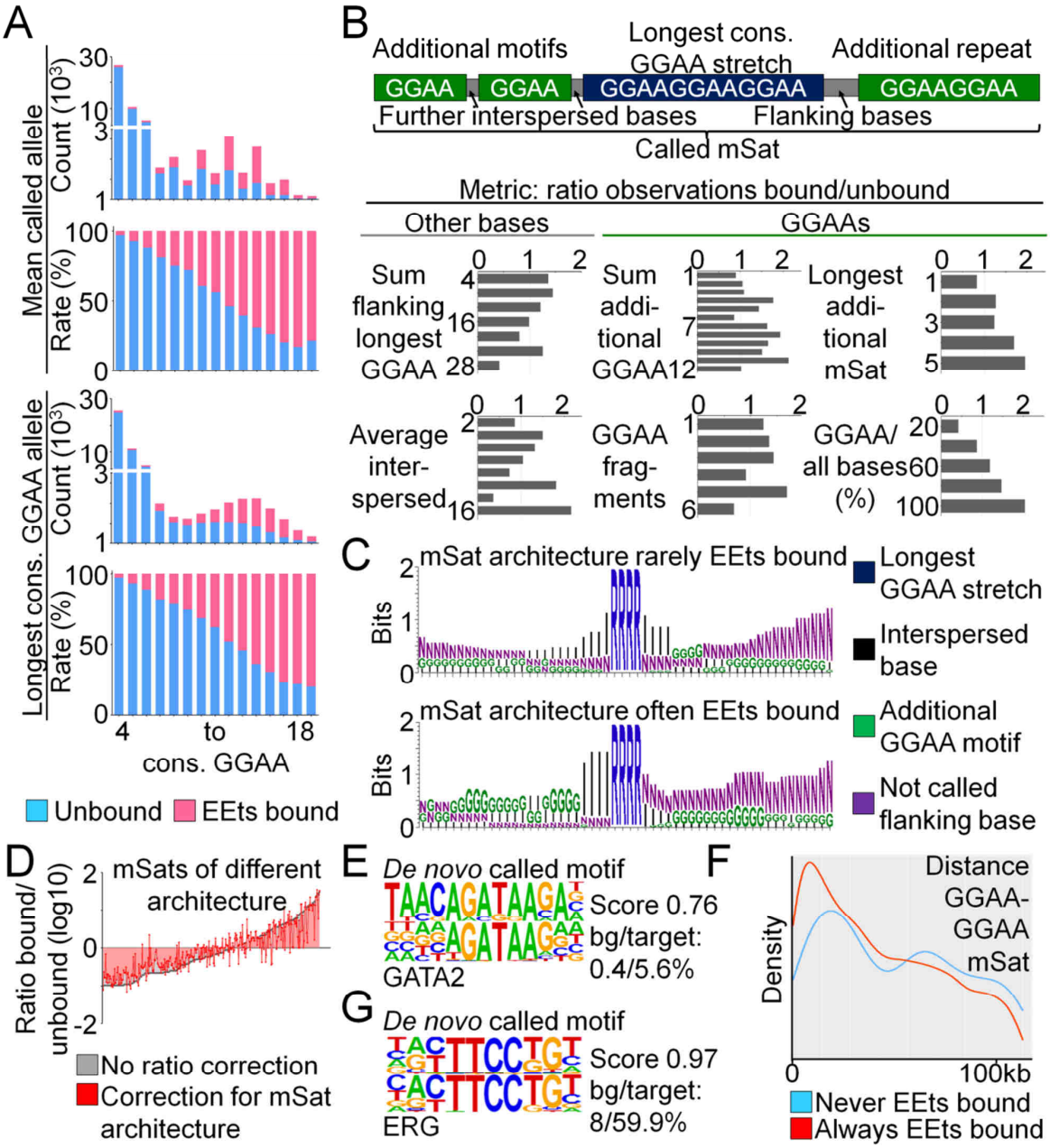

Figure 27: The number of consecutive GGAA motifs, additional nearby motifs and nearby transcription factor binding enhance EETs binding probability. A) Bar plots of alternately absolute count and relative count of GGAA mSats EETs-bound in ChIP versus unbound, stratified by the mean (top panel) and maximum (lower panel) number of consecutive GGAAs of both alleles per locus; cons.: consecutive. B) Top: Scheme of the called mSat "architecture": Besides the longest consecutive GGAA stretch additional single motifs and repeats with interspersed bases were called. Below: Subsetting of all genotyped GGAA mSats for different characteristics of the mSat architecture, bars indicate value of the metric ratio of observing the respective mSat EETs-bound or -unbound in ChIP, normalized to 1 and here additionally to the effect of the longest consecutive GGAA stretch; 1 indicates no effect, >1 favoring EETs binding, <1 EETs-binding impairment; cons.: consecutive. C)

Those 10% mSat architectures with the lowest metric (see B) and highest metric were transformed to a stacked motif, illustrating differences between rarely and often bound EEts mSats. D) Plot of the metric (see B) for each mSat architecture, in grey without any correction for any observed effect of the mSat architecture, red after correction for the longest consecutive GGAA stretch, flanking alternate bases, additional GGAAs and further interspersed bases, each dot represents one mSat architecture. The area under the curve is filled with light color and indicates how much determinants of EEts binding were calculated out of the metric. E) Exemplary stacked motif of *de novo* calling for enriched motifs nearby EEts-bound mSats and the matched GATA2 binding motif; bg: background. F) Density plot indicating the distance distribution between an EEts-bound GGAA mSat to any other (red) versus an unbound mSat (blue). G) Stacked motif of the most significantly enriched *de novo* called motif at EEts binding sites externally of GGAA mSats; bg: background.

All characteristics of the mSat architecture indicated complementary to each other, that additional GGAA motifs increase the probability of binding, especially when only a few interspersed bases were present, leading to more consecutive GGAAs. Moreover, closer vicinity of the additional GGAA motifs to the longest stretch seemingly increase the EEts-binding probability (Figure 27B,C). However, although values for characteristics with rather low overall counts of observations (<50) were not included in the Figures, several outliers were evident. Nevertheless, combination of the characteristics as percentage of GGAA bases of all bases in the called mSat, despite the longest consecutive GGAA stretch, led to a significant correlation with EEts binding (Pearson's r^2=0.99, P=0.0007; Figure 27B). Obvious differences between the mSat architecture up- and downstream of the longest consecutive GGAA stretch were not observed.

As the applied metric (rate of EEts-binding observation) was normalized to 1 (log10 transformed zero), adjusting the initial metric for all factors affecting binding probability should reduce the area under the curve of all observed metric values, optimally to zero. However, when successively normalizing the applied metric for the differences observed for the longest consecutive GGAA stretch, the number of flanking bases before another motif, the number of additional motifs, and the rate of interspersed further bases, the area under the curve decreased by 22.2% only - with the strongest effect already observed by the highest number of consecutive GGAA motifs (18.9%; Figure 27D).

As another potential influencing factor on EEts binding, nearby co-bound transcription factors can be considered. Therefore, mSats always bound in EEts ChIP were compared with those never bound for transcription factor binding motif enrichment 1 kb (standard promoter size) up- and/or downstream. No motif was found to be present nearby most bound mSats, but not nearby the unbound. The only significantly identified motifs matching known transcription factor binding motifs, present in about 5% of the tested target regions and at least by factor 10 less nearby unbound mSats, were those for *GATA2/6* (Figure 27E) and *FOXF2* downstream, and *RUNX1* and *MIXL1* upstream of the mSats. Interestingly, *GATA2* was downregulated in 8 cell lines upon EEts knockdown, which might indicate that in EwS EEts

promotes co-binding with GATA2 in the rare cases of motif colocalization. Even distant GGAA mSats might affect EEts binding to another mSat, as within a 100 kb window always EEts-bound GGAA mSats were closer to another mSat than not EEts-bound mSats (Figure 27F). Additionally, copy number gains of mSats seemingly favor EEts-binding detection at the respective site. Each of the 3,377 GGAA mSats at least once bound in EEts ChIP was inspected for EEts binding per cell line and for CNV. 32,480 times a mSat was not bound, 28,306 times bound. 11.9% of unbound mSats showed copy number loss, but only 8.3% of bound mSats. Vice versa, 16.5% of unbound mSats showed copy number gain, but 21.7% of bound mSats (P<0.0001).

Motif finding was also performed for EEts-binding sites observed in at least 50% of cell lines externally of mSats. As expected, *de novo* motif calling resulted as top hit with by far highest significance in a GGAA motif not flanked by another GGAA, hinting towards binding capability of EEts also to a single GGAA motif or small repeats (reverse complement in Figure 27G).

To see the effect of EEts binding to GGAA mSats on expression regulation of nearby genes, Figure 27A was rebuilt, subsetting the bound and unbound mSats by the effect of EEts knockdown on the closest gene (Figure 28A). Surprisingly, the overall rate of nearby downregulated genes was already 6.7% for four consecutive GGAA motifs (mean allele count) and did not show a constant increase over the number of consecutive repeats (Figure 28B). Even more, when comparing the rates of nearby downregulated genes for EEts bound versus unbound mSats, there was no significance for higher rates of regulated genes nearby bound mSats. Similiar results were achieved for upregulated genes upon EEts knockdown, besides a significance of slightly higher rates of upregulated genes (P=0.016) nearby bound mSats, which was not expected as insulator function of EEts at GGAA mSats has not been described, yet. Nevertheless, in previous ChIA-PET data (generated by Didier Surdez, Paris, France; not yet published) associations of promoters with GGAA mSats even 1 Mbp distant were shown. This means, that a gene might be regulated via EEts binding to a mSat far apart, what suggests that the rate of regulated genes might be much higher than displayed in Figure 28A. Indeed, when testing a 2 Mbp window around each GGAA mSat, 99% of all genes in the transcriptome data were overlapping.

When testing vice versa genes, that appeared as likely regulated upon EEts knockdown among all cell lines (defined as regulated in one direction in at least 33% of cell lines and never regulated in the opposite direction), the distance to the next GGAA mSat was significantly smaller for downregulated genes versus not regulated genes (P<0.0001), while the next GGAA mSat for upregulated genes was more upstream (Figure 28C). When again

looking more specifically for genes regulated per cell line and the nearest in the respective cell line EEts-bound GGAA mSat, the results were even more pronounced (Figure 28D). The closest EEts-bound GGAA mSat was little more often on the opposite strand of the gene body (50.1, 51.2 and 52.5% for not, up- and downregulated genes), more often downstream when on the opposite strand (54.7, 52.8 and 58.2% for not, up- and downregulated genes), and upstream when on the same strand (56.8, 61.3 and 61.9% for not, up- and downregulated genes). As the differences were most pronounced for genes downregulated upon EEts knockdown, even highly significant (P<0.0001) when compared to not regulated genes, strandness might be relevant for EEts-mediated gene regulation. The number of GGAA mSats per 2 Mbp-window around transcriptional start sites was higher for downregulated genes than for not regulated genes (on average 6.70 versus 5.96 GGAA mSats, 12.3% difference), and even more when focusing on EEts-bound GGAA mSats (on average 1.41 versus 0.98 EEts-bound GGAA mSats, 43.5% more, P<0.0001 for both differential rates). Those differences were less for genes upregulated upon EEts knockdown, but again there was a difference between binding localization relative to the TSS as 11.4% more GGAA binding upstream than downstream of the TSS was observed. In contrast to bound GGAA mSats, the distance to the next EEts binding externally of a GGAA repeat was rather downstream for genes upregulated upon EEts knockdown (Figure 28E). Of note, EW24 was excluded from this analysis because of the relatively very high number of identified binding sites. Thus, the actual distance to EEts-bound GGAA mSats versus single GGAA motifs differentially affect gene expression regulation.

Promoters of regulated genes were tested for transcription factor binding motif enrichment versus promotors of not regulated genes, subgrouping them into up- and downregulated upon fusion knockdown. These analyses did not reveal a motif exclusively or in most cases seen in the promoters of regulated genes, but not in the not regulated. Nevertheless, 12.8% of upregulated genes harboured a NFAT5 binding motif (versus 4.3% of not regulated genes). For the downregulated genes NFY binding motif was found in 27.7% of promoters and a motif highly similar to E2F4, 2, and 1 binding site in 11.1% (versus 15% and 4.1% in not regulated genes, respectively; Figure 28F). Noteworthy, *NFAT5* was upregulated in three cell lines, *NFY subunit C* was downregulated in 6 and *E2F4*, *2*, and *1* in 1, 15 and 12 cell lines upon fusion knockdown. Hence, likely *NFAT5* is suppressed and *NFYC*, *E2F1* and *E2F2* are promoted in EwS, decreasing expression of genes with NFAT5 binding site in their promoters and increasing expression of genes with NFY and E2F family binding sites.

Hence, high numbers of consecutive GGAA motifs adjacent to further GGAA repeats increase the probability of EEts binding, despite nearby transcription factor binding sites

(GATA2), further mSats, copy number gains, and still unknown factors. Enrichment of nearby EEts-bound GGAA mSats influence the effect of EEts on gene expression dependent on the actual distance and relative localization to the TSS, partly together with specific EEts-regulated transcription factor motifs in the promoter region.

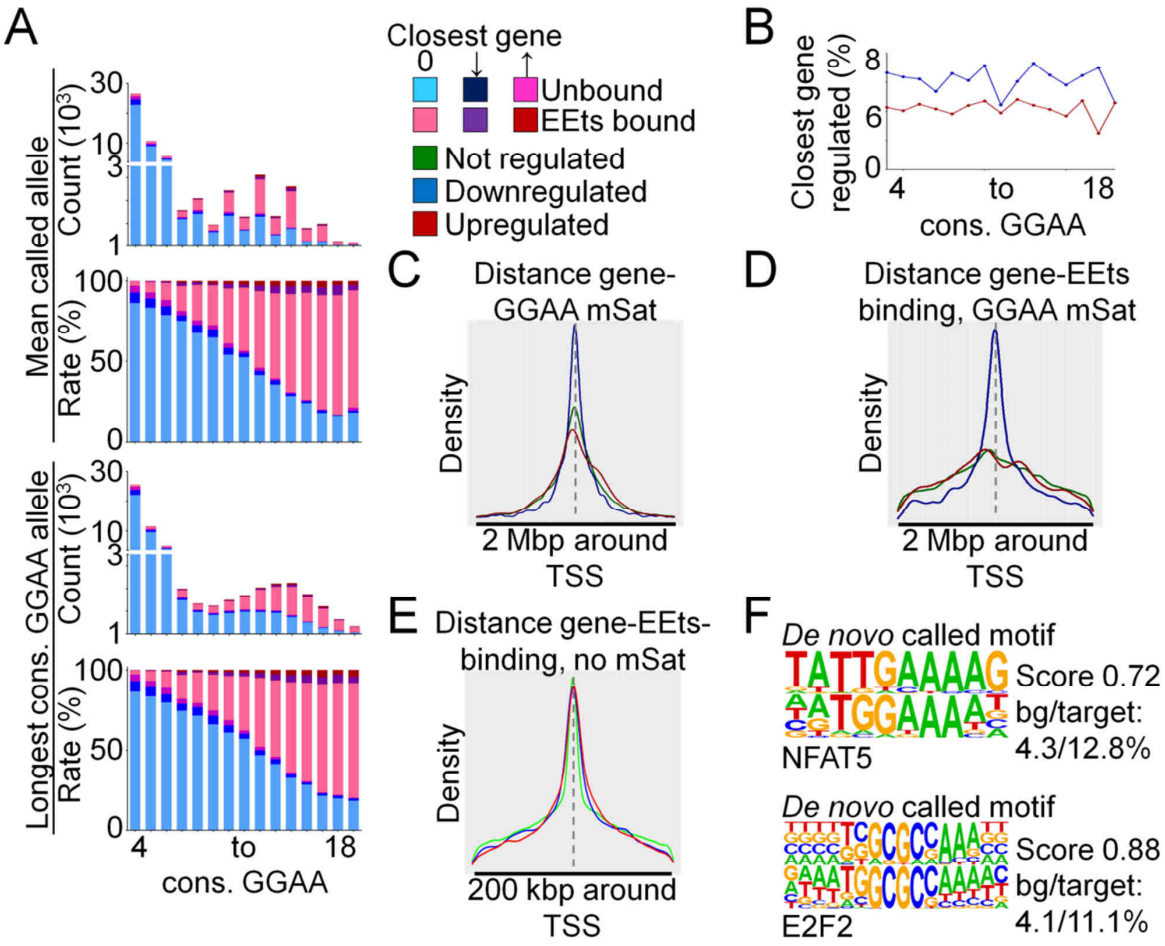

Figure 28: Genes up- and downregulated upon EEts knockdown differ in their distance to the next EEts-bound GGAA mSat and in transcription factor binding motifs in their promoters. A) Bar plots of alternately absolute count and relative count of GGAA mSats EEts-bound in ChIP versus unbound, further subsetted for the effect of EEts knockdown to the closest gene, stratified by the mean (upper panel) and maximum (lower panel) number of consecutive GGAAs of both alleles per locus; cons.: consecutive. B) Percentage of closest genes to a GGAA mSat that are regulated upon EEts knockdown, stratified by mean number of consecutive GGAA motifs for both alleles per locus; cons.: consecutive. C) Density plot indicating distance to the next GGAA mSat for genes not, up- or downregulated upon EEts knockdown. D) Density plot as in C, but for the distance to the next EEts-bound GGAA mSat in the respective tested cell line. E) Density plot as in D, but for the distance to the next EEts binding externally of a GGAA mSat. F) Exemplary stacked motifs of frequently observed transcription factor binding sites in promoters of genes upregulated (top) and downregulated (bottom) upon EEts knockdown; bg: background.

4.5. EWSR1-ETS heterogeneously regulates clinically relevant genes

One main motivation for creating the ESCLA with 18 cell lines was to overcome the present lack of studies on the overt inter-patient heterogeneity. Thus, in the next step the applicability of the ESCLA to model heterogeneity and the transferability of the results to clinically relevant genes were initially tested. First, genes with heterogeneous expression regulation were selected. Several cells per group (strong versus weak/no differential expression in transcriptome data upon EEts knockdown) were needed to reinforce statistical power, and between those groups a grey zone for the resolution of the here applied analyses might exist. Hence, heterogeneously regulated genes in the ESCLA were defined as being strongly regulated (top 33% of regulated genes of the respective cell lines) in at least 33% of cell lines, and being not or weakly regulated (lower 33% of regulated genes of the respective cell lines) in another 33% of cell lines. This approach found 256 genes in the ESCLA to be heterogeneously regulated upon EEts knockdown.

Second, to identify potentially clinically relevant and thus potentially clinical heterogeneity-mediating genes, the list of heterogeneously regulated genes was crossed with survival data. Of the heterogeneously regulated genes 209 were represented in the survival dataset. Applying best-percentile algorithm and Bonferroni correction, 22 genes were significantly associated with worse overall survival in the EEts-mediated expression state (high expression when downregulated upon EEts knockdown, hence likely driven by EEts in patients; low expression when upregulated upon EEts knockdown, hence likely suppressed by EEts in patients) (Table 20).

The inter-quartile range of gene expression values in the survival dataset was utilized as metric for heterogeneous expression. Actually, for 15 of the 22 identified heterogeneously regulated and survival associated genes (68%) the inter-quartile ranges were over the 80th percentile of the patients' cohort.

Many factors influencing EEts binding to GGAA mSats and gene expression, described above, are not suitable to explain the heterogeneity, as they do not differ between cell lines. The total number of GGAA mSats in a 1 Mbp window around the TSS of a gene, the distance to the next GGAA mSat, transcription factor binding sites nearby the GGAA mSat and in the promoter are unlikely to account for heterogeneity, except for cases of binding-motif perturbed or *de novo* generated SNVs and large InDels.

More likely, the binding rate of nearby GGAA mSats, their length and CNVs, and CNVs of the gene body might differ between the cell lines. However, when comparing all these factors for each gene in the group of cell lines with strong effect of fusion knockdown on the expression

of the respective gene and in the group without strong effect, either no difference or no uniform difference between groups was observed (Figure 29).

Table 20: Genes heterogeneously regulated in the ESCLA and survival associated. *P* value is Bonferonni corrected; Surv. corr: expression status correlating with worse survival; IQC perc.: percentile of the inter-quartile range in the expression data of the survival cohort.

Gene	Description	*P*	Surv. corr.	IQR perc.
AIF1	Actin binding protein regulating immune cells	0.001	high	88
CDC20	Cell cycle regulator	0.0004	high	91
CDC25A	Dual-specific phosphatase, DNA damage responder	0.0028	high	50
CDC25C	Cell cycle regulator, triggers mitosis	0.0194	high	88
CENPI	Centromere protein; E2F target	<0.001	high	79
DSEL	Dermatan sulfate epimerase-like	0.0002	low	90
DUSP26	Dual-specific phosphatase	0.0161	high	66
E2F1	Transcription factor, cell cycle regulator	0.0001	high	34
EXO1	Exonuclease	<0.001	high	94
GPN3	GTPase	0.0002	high	68
GTSE1	Cell cycle dependent protein	0.0003	high	80
KIF14	Microtubule regulator involved in mitosis	0.019	high	90
KIF15	Microtubule regulator involved in mitosis	0.0075	high	90
KIF2C	Microtubule regulator involved in mitosis	<0.001	high	88
MYBL2	Transcription factor, involved in cell cycle	0.0004	high	87
NEK2	Protein kinase involved in mitosis	<0.001	high	79
NUF2	Centromere associated protein	0.0011	high	97
SELENBP1	Selenium binding protein	0.0039	low	99
SPAG5	Associated with mitotic spindle apparatur	0.0093	high	89
TP53INP1	p53 interactor	<0.001	low	74
TPX2	Microtubule regulator involved in mitosis	0.0014	high	89
TRIP13	ATPase involved in cell cycle progression	0.0084	high	90

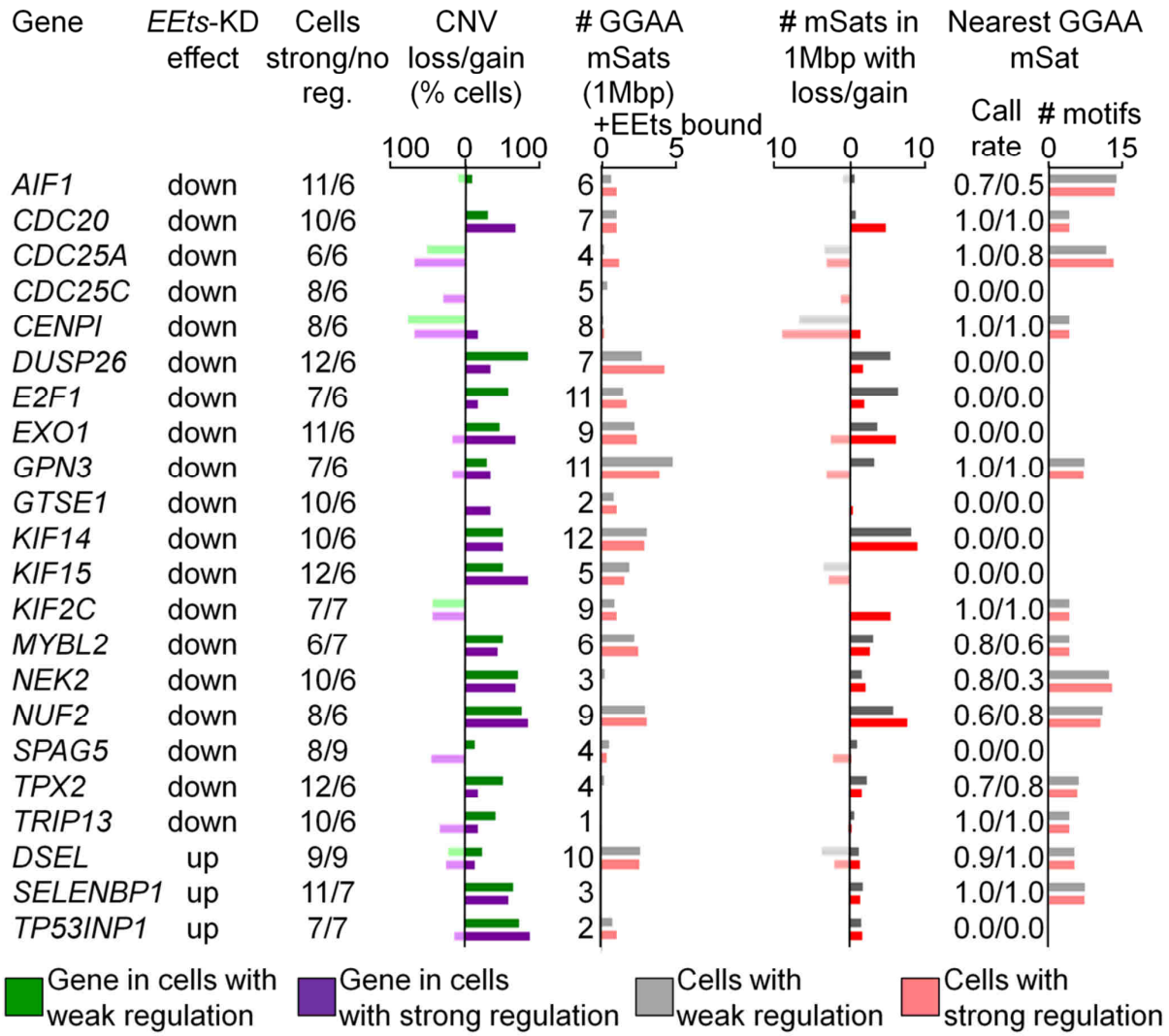

Figure 29: Heterogeneous regulation upon EEts knockdown for genes, which are heterogeneously expressed in patients and associated with overall survival, cannot be explained by CNV of the respective gene, number of GGAA mSat binding up to 1 Mbp distant from the TSS, CNV of mSats in 1 Mbp distance to the TSS, or the number of consecutive GGAA motifs in the nearest mSat, alone. All calculations were performed for the two groups of cell lines with strong effect of EEts knockdown on the gene expression and with no or weak effect. All values were normalized to the number of cells per group; reg.: regulation.

Thus, the ESCLA is capable to identify heterogeneously regulated genes, which are also heterogeneously expressed in patients. However, the many known and here partially described and evaluated factors influencing EEts-binding to GGAA mSats and mechanisms of EEts-dependent gene-expression regulation are not sufficient to identify, whether EEts binds to a GGAA mSat, and how this binding affects gene expression. Hence, further investigations are demanded to identify determinants of patients' heterogeneity. The ESCLA is likely a rich resource for this research journey.

5. Discussion

EwS is an aggressive bone-associated cancer mainly affecting children and young adults (Grünewald et al., 2018). Prognosis for localized disease is favorable, but it is poor for patients with non-localized disease (Bosma et al., 2018). The clinical heterogeneity is in contradiction to the few recurrent mutations in EwS (Tirode et al., 2014).

Actually, the *EEts* fusion appears as the major, if not single driver of EwS. It reprograms the genomic accessibility (Patel et al., 2012; Tomazou et al., 2015), binds to the genome (Guillon et al., 2009; Monument et al., 2014), and drives the expression of several genes, which were shown to promote cancer aggressiveness (Grunewald et al., 2012; Musa et al., 2019; Smith et al., 2006; Surdez et al., 2012). Moreover, expression of *EF1* in embryonal stem cells resulted in a highly similar phenotype to EwS (Gordon et al., 2016). Thus, a better understanding of the biology of the fusion oncogene, and parameters which affect the biology of the fusion oncogene and might explain inter-patient heterogneity, is highly important. This constituted the rationale to generate the ESCLA. The ESCLA not only comprises data of 18 EwS cell lines to model heterogeneity, but was also generated with focus on the *EEts* fusion biology. The WGS platform was purposely selected as Illumina 150 bp paired-end reads were described to enable high accurracy microsatellite calling (Willems et al., 2017), like GGAA mSats described as main binding sites of EEts (Gangwal et al., 2008; Guillon et al., 2009). The ChIP experiments were performed for distinct fusion types. Moreover, for comparative analyses transcriptome, proteome and methylome data were generated for all cell lines in fusion-high and -low state.

The cell lines were selected to represent the most common fusion types, *EF1* type 1, 2 and *EErg*. 93% of EwS are positive for one of these fusions (Berger et al., 2013). A fusion oncogene knockdown, but not knockout approach was chosen to evaluate effects of the fusion oncogene on gene expression. As EwS is dependent on the fusion, fusion knockout might have been lethal for EwS cell lines (Patel et al., 2012; Smith et al., 2006). For the same reason, an inducible shRNA construct was employed. CRISPRi (Qi et al., 2013) might be used as complementary strategy on the genome level to test in synopsis with the here generated data for fusion specifity of the observed effects. The achieved knockdown led to *EEts* rest-expression of 16 to 60% (mean 31%). Whereas knockdown of the *EF1* type1 and 2 fusion did not impair the expression of non-fused wildtype *EWSR1* and *FLI1*, the *EWSR1* expression was reduced upon shEErg expression. However, the knockdown effect on the fusion was stronger (46%) than on *EWSR1* and the *EWSR1* expression level twice as high

as for the fusion. Furthermore, an interaction of EF1 with EWSR1 has been described that interfers with the *EWSR1*-mediated transcriptional repression (Gorthi et al., 2018). Therefore, *EWSR1* is presumably not physiologically active in EwS and any effects of the shRNA-mediated *EWSR1* impairment on the transcriptome, which was the readout of the knockdown experiments, are unlikely. Strikingly, even in two *EF1* type 2 positive cells, in which the applied shRNA for fusion knockdown targeted *FLI1* in exon 9, a massive increase of non-fused *FLI1* was observed upon fusion knockdown. Actually, a delicately balanced expression regulation of *ETS* transcription factors in EwS might explain known differences in the proliferative versus metastatic capacities of EwS cell lines dependent on the fusion expression (Aynaud et al., 2020; Chaturvedi et al., 2012; Franzetti et al., 2017; Katschnig et al., 2017). Moreover, suppression of *ETS* transcription factors, which are known mediators of differentiation (Kar and Gutierrez-Hartmann, 2013), might stabilize the undifferentiated phenotype of EwS (Fletcher, 2014; Grünewald et al., 2018).

TMAs were generated from six xenografted cell lines. These TMAs enabled *in situ* validation of described EEts-regulated genes like *PAX7* and *SOX6* (Baldauf et al., 2018b; Charville et al., 2017; Marchetto et al., 2020). The TMAs add value to qRT-PCR and western blot analyses in 2D culture models. First, gene regulation can be evaluated in 3D-developed tumors grown in a mammal instead of in an artifical 2D culture. Second, the subcellular distribution of staining is an additional specificity indicator which is missing in whole cell lysate western blot analyses.

WGS, and consciously not whole exome sequencing as in a few former studies (Sand et al., 2015), was performed to cover regulatory regions with GGAA mSats. For library preparation, a PCR-free protocol was used to avoid any PCR-stutter artifacts. The achieved coverage with 150 bp paired-end reads was 37x. WGS data depticted mutation rates for *TP53*, *STAG2*, and rate of *CDKN2A* deletion of 78%, 50%, and 28%, respectively. Thus, at least for the cell lines in ESCLA, rates of genetic alterations are far higher than reported in patients (7%, 17% and 12%, respectively) (Tirode et al., 2014). Of note, the proposed mutual exclusivity of *STAG2* mutation and *CDKN2A* deletion (Tirode et al., 2014) was not observed for TC32. Additionally, frequent copy number alterations like chr1q (Mackintosh et al., 2012), chr8 and chr12 gain and chr16q loss were displayed by the data (Tirode et al., 2014). From this qualitative WGS data, HipSTR (Willems et al., 2017) inferred genotypes of 3,831 GGAA mSats from 8,311 potential sites. The potential sites were retrieved from the reference genome (hg19) demanding at least four consecutive GGAA motifs. Thus, sites potentially expanded to five repeats in EwS, described as minimal number for EEts-mediated enhancer activity (Gangwal et al., 2008), were also called. Actually, 132 of 1,593 GGAA mSats with

four motifs in the reference genome had at least five repeats in some EwS cell lines. Up to 24 consecutive motifs were genotyped by HipSTR. However, this tool is limited to call maximum mSat length close to the read length. 1,895 (23%) of potential GGAA mSats were at least 150 bp long in the reference genome. This warrants long-read sequencing to increase knowledge about GGAA-mSat genotypes in EwS. Unfortunately, although these methods have been improved, they have still relatively high error rates compared to sequencing on an Illumina HiSeq Xten platform (De Roeck et al., 2019; Fu et al., 2019). Such error rates might result in falsely called mSat-interrupting bases, which can heavily corrupt association studies of mSat length and effects on genes, especially as the consecutive GGAA number was the most important factor for EEts binding in the here generated results.

Described EF1 binding and additionally EErg binding to GGAA mSat (Gangwal et al., 2008; Guillon et al., 2009; Monument et al., 2014) was confirmed in the ChIP-Seq data. Actually, most consensus sites from two cell lines (Riggi et al., 2014) were present in most, if not all cell lines. Having 18 cell lines analysed with various numbers of called peaks, probably several consensus sites were missed out due to false negatives, so that further studies on EEts-genome interaction should not be limited to consensus sites, but also include common sites. Still, 280 consensus sites were identified for the 18 EwS cell lines, including 50 sites which were not described in the previous consensus set. As a high overlap of the 280 sites with the consensus set of Riggi *et al.* was observed, and as TC32 was an outlier with only 1,253 peaks for the fusion, these 280 consensus site likely represent a conserved set of EEts-genome interactions in EwS and not random hits. Furthermore, reported *NR0B1* and *MYBL2* associated mSats were bound in most cell lines (Johnson et al., 2017a; Musa et al., 2019). Interestingly, even after application of a cross-cell line normalization method, based on the assumption of equal peaks for genes expressed at the same level in all cells, peak heights showed strong inter-cell line heterogeneity. For all cell lines, at least 100 bp-read sequencing was performed, as peaks in mSats were expected and standard 50 bp reads might have not been sufficient for qualititative mapping (Zhang et al., 2016). For EW24, the only cell line whose ChIP products were sequenced with paired-end reads, far more peaks were called than for any other cell line. Further experiments in the same and other cell lines with paired-end sequencing have to clarify whether the multitude of peaks were the result of, for instance, a contamination, or if several true peaks were missed out by single-end sequencing. 44% of EEts binding sites were located in intergenic regions, 47% in introns. Actually, EEts-mediated enhancer activity by binding to an intron, instead of a distant mSat, has been already published (Dallmayer et al., 2019; Marchetto et al., 2020). EEts binding was colocalized with the H3K27ac enhancer mark. This observation is in line with the described enhancer activity of and H3K27ac at EEts-bound GGAA mSats (Monument et al., 2014; Patel et al., 2012; Riggi et al., 2014; Tomazou et al., 2015). Super-enhancers (SEs)

were called with the ROSE algorithm (Lovén et al., 2013; Whyte et al., 2013). SEs often overlapped with known targets of the EwS fusion oncogene, like *PRKCB*, *SOX6* and *GLG1* indicated for TC71 in Figure 16C (Baldauf et al., 2018a; Marchetto et al., 2020; Surdez et al., 2012). For those 58 super-enhancers which were found in all cell lines, the overlapping and nearby genes overrepresented gene ontology terms in context of development, transcription and metabolism, including eight of 98 genes involved in embryonic skeletal system morphogenesis. Super-enhancers are described as determinating cell identity (Whyte et al., 2013). Therefore, the overrepresented genes might be an indicator of the cell of origin. Indeed, the identified gene ontologies are in line with the current hypothesis that EwS origins from mesenchymal stem cells (Sheffield et al., 2017; Tirode et al., 2007). Nevertheless, the EwS cell origin is still a matter of debate. As *EEts* drives the expression of a multitude of genes, fusion knockdown might result in a transcriptome profile more similar to the cell of origin. Thus, a t-SNE plot with transcriptome data of EwS cell lines without and with fusion knockdown and 71 published normal tissue types (Baldauf et al., 2018b) was generated. However, the fusion knockdown was not sufficient to change the transcriptome profile of the cell lines strong enough to reach higher similarity of *EEts*-low cells with any normal tissue than with *EEts*-high cells.

Described targets of *EEts* mediated gene expression regulation were reidentified or confirmed for *EErg*-positive cell lines. However, *EEts*-knockdown in EwS cell lines was heterogeneous, hence comparative analyses were performed (semi-)rank based. The proteome data were strongly and significantly correlated with the transcriptome data, thus mutually validating each other. The strong correlation of transcriptome and proteome suggests that transcriptional regulation is of higher relevance in EwS than posttanscriptional modifications. Indeed, publications on posttranscriptional modification in EwS are rare (Baumuratova et al., 2010). Surprisingly, the overlap of genes regulated in all cell lines was very limited. Actually, only four genes were identified, namely *FEZF1*, *TRPM4*, *PAX7* and *PPP1R1A*. The transcription factor *PAX7* is known to be highly expressed in EwS (Baldauf et al., 2018b), but its function in EwS is still unclear. Unfortunately, the *PAX7* expression is not fully specific for EwS, thus not suitable as diagnostic marker (Baldauf et al., 2018a). *PPP1R1A* was very recently described to regulate cell cycle progression in EwS (Luo et al., 2020). *FEZF1* represses transcription, is involved in neuronal development, and has been already implicated in the context of cancer (Lan et al., 2018). *TRPM4* is a cation channel expressed in various normal tissues and cancers (Wong and Hussain, 2020). Nevertheless, integration of all transcriptome data by building average ranks for gene regulation across cell lines and subsequent GSEA led to enrichment of gene sets annotated with cell cycle and proliferation. These gene sets were in line with known *EF1* mediated proliferation (Franzetti et al., 2017) and the previous notion that one third of *EF1* regulated genes are involved in

cell cycle (Patel et al., 2012). The overrepresented gene ontology terms in context of DNA replication and cell cycle matched well several reports on the proliferative activity of EwS with still quiet genomic landscaping (Franzetti et al., 2017; Patel et al., 2012; Tirode et al., 2014), demanding a precise cell cycle regulation. In line, EwS is highly susceptible to inhibition of DNA damage repair via CDK12/13 (Iniguez et al., 2018). This implies a common biology of EwS cell lines, that is seemingly mediated by different genes in the different cell lines. Hence, targeting a single gene might be less efficient in EwS than targeting pathways, what further emphasize the importance of understanding the biology of EEts.

An EwS specific, but highly inter-tumor heterogeneous methylation profile was reported (Sheffield et al., 2017). When comparing the methylation profile of the cell lines with data on primary tumors (Capper et al., 2018), the metric of all six cell lines depicting an EwS profile decreased upon fusion knockdown, indicating an effect of the fusion on the EwS-specific methylome. However, not any uniform differentially methylated region was identified upon fusion knockdown. Furthermore, the cell lines clearly clustert apart from primary EwS in t-SNE analysis of the methylome, although an overlap of methylation profiles of cell lines and primary EwS was shown in a previous study (Sheffield et al., 2017). The same was observed for the transcriptome data. At least, EwS cell lines were neighboring to primary EwS in both datasets. Possibly the observed methylation profile in cell lines is one extreme of the reported methylation spectrum of EwS which is favored for growth *in vitro* and rather not depicted *in vivo*. The biological meaning of the different methylation profiles identified by Sheffied *et al*. is not elucidated. However, in synopsis of the ChIP-Seq data, histone modifications appear to be the more relevant epigenetic marks for EwS. Hence, more ChIP-Seq experiments in *EEts*-low status should be performed to further outline the epigenetic effects of *EEts*. Successful identification of epigenetic EwS characteristics by this approach has been demonstrated already (Tomazou et al., 2015).

The clear clustering of EwS cell lines in t-SNE plots for both transcriptome and methylome challenge the validity of the cell lines as model for EwS. Additionally, current publications question the value of cell lines due to poor reproducibility of cell-culture based results (Hirsch and Schildknecht, 2019; Niepel et al., 2019). However, alternative models are scarce. Furthermore, EwS is genomically rather stable, thus the cell biology of cell lines might stay closer to the primary tumors in EwS than in many other tumor entities. Eventually, other culture conditions, e.g. 3D growth, might already improve similarity of cell-line biology to the tumor (Kapałczyńska et al., 2016). Alternatively, patient-derived xenografts (PDX) are on the rise (Siolas and Hannon, 2013), and first EwS PDX were established (Marchetto et al., 2020;

Rokita et al., 2019). Future studies have to elucidate their value as tumor model compared to cell lines.

For the first time, cell lines with distinct fusion types were systematically compared. Previously, several studies discussed differences of fusion types in regard of the phenotype prognosis, first assuming (de Alava et al., 1998) and later reject the hypothesis of any fusion-type specific effect (Le Deley et al., 2010). Indeed, neither for transcriptome and proteome nor for methylome clustering of one fusion type from others was observed in t-SNE plots. Despite the similarities in t-SNE analyses, high overlaps of top results in GSEA and of shared regulated genes and SEs were detected. Still, *EF1* type 2 positive cell lines displayed higher overlaps with *EF1* type 2 postive cell lines, as did *EErg* positive cells with others *EErg* positive cells. The reason for those seemingly fusion-type privately regulated genes might be simply a stochastic phenomenon, as far less *EF1* type 2 and *EErg* positive cell lines were tested than for *EF1* type 1 (4 and 3 cell lines versus 11). Hence, the risk for any outlier reducing the overlap rate was far less for these fusion types. Nevertheless, there might be differences between EwS with distinct fusion types as a consequence of the developmental history of the fusion oncogene. The *EErg* fusion cannot be explained by simple reciprocal translocation, as the fusion partners are localized on opposite strands. Thus, likely the *EErg* fusion arises from chromoplexy (Anderson et al., 2018). The developmental process of the EwS cell lines' fusions were assessed with three methods, PCR of the inverse fusion transcript, read counts from BreakDancer (Chen et al., 2009) on the plus and minus strand of *EWSR1*, and rebuilding rearrangement loops with ChainFinder (Baca et al., 2013). For all *EF1* type 2 positive cells a reciprocal fusion was detected. In contrast, six of the eleven (55%) *EF1* type 1 positive cells did not yield any evidence for reciprocal translocation, but for chromoplexy. Hence, in nine of 18 cell lines (50%) chromoplexy occured. The rate was slightly higher than in a published cohort of primary EwS (42%) (Anderson et al., 2018). Noteworthy, chromoplexy always affects at least one other genomic locus. For EW24 even several loci were identified as involved in rearrangement loops. Studies on a large series of primary EwS might elucidate whether chromoplexy often affects the same genes in EwS or rather random loci. This could imply distinct biology of EwS dependent on the developmental history of the fusion, and in consequence the necessity of EwS-treatment stratification by chromoplexy versus reciprocal translocation. The here generated ChainFinder results were based on the cell line data solely, while usually such tools are used after extensive data quality adjustment with available normal controls (Anderson et al., 2018; Baca et al., 2013). On the one site the successful identification of rearrangement loops in the cell lines corroborate the data quality of the ESCLA, on the other site some rearrangements might not

have been fully called. Additionally, chromothripsis has been described to cause gene fusions (Anderson et al., 2018; Stephens et al., 2011). As chromothripsis affects rather more genes on individual chromosomes than chromoplexy, there might be differences between EwS with chromothripsis versus chromoplexy as genomic rearrangement event. Thus, also this mechanism of fusion development should be further investigated for EwS.

In sum, the EwS cell lines with distinct fusions depicted many similarities and not any clear difference. Except for proteasome subunits in *EF1* type 1 positive cell lines, not any gene was recurrently represented in fusion-private gene sets (GSEA), but not regulated by other fusions. Thus, all cell lines were jointly analysed. The further analyses were conducted combining readouts of the several datasets included in the ESCLA to investigate *EEts*-mediated gene regulation and potential factors of heterogeneity.

When it comes to heterogeneity, very little is known for EwS. As EwS is genomically silent, but the fusion oncogene interacts with highly polymorphic GGAA mSats (Gangwal et al., 2008; Mirkin, 2007), likely these mSats cause heterogeneity. Indeed, of 3,831 mSats genotyped across at least 6 cell lines with HipSTR, 1,467 (38%) corresponded to the reference, but nearly as many (1,297; 34%) mSats had even more than two alternative alleles. Actually, examples for the clinical impact of GGAA-mSat length were published. A SNP, whoose A allele is more frequent in Caucasions than Africans, connects two adjacent GGAA mSats creating one long consecutive repeat stretch. This GGAA mSat has enhancer activity for *EGR2*, a susceptibility gene. In line, the EwS incidence in Caucasians is 10x higher than for Africans (Grünewald et al., 2015). In 2019, the relevance of a GGAA-mSat length for prognosis was shown in a publication coauthored by the author of this thesis (Musa et al., 2019). *MYBL2* was higher expressed in patients with longer GGAA mSats and was associated with survival (Musa et al., 2019). Actually, *MYBL2* was reidentified in the ESCLA as heterogeneously regulated and survival associated gene.

In line with previous publications, higher numbers of consecutive GGAA repeats displayed higher rates for EEts binding (Gangwal et al., 2008; Guillon et al., 2009; Monument et al., 2014). Yet, a sweet spot for EEts binding was not observed, although in luciferase reporter assays a strong enhancer activity increase was seen at 18-26 repeats before a dip and later increase at about 50 repeats (Johnson et al., 2017a; Monument et al., 2014). However, only once 24 repeats were genotyped. Thus, the read length was too small to achieve many genotypes with the length of the "sweet spot" or even beyond. Nevertheless, the here generated data are based on a diploid human genome and not on an artifical mono-allelic model. Therefore, this data, as far as sufficient calls are availabe, should be more reliable. Besides higher binding rates with increasing number of GGAA repeats, also the absolute

number of genotyped mSats accelarated again beyond seven to nine GGAA-motif repeats. Rerunning the analysis with other tumor entities and/or normal tissues is necessary to evaluate if the mentioned increase in mSat calls with more than seven to nine repeats is specific for EwS and might indicate higher EwS susceptibility for children with longer mSats. Even more, this analysis could clarify whether GGAA mSats expand in EwS or if the observed mSat diversity is actually germline variants. Current data clearly favor the hypothesis of germline variants. For instance, the frequency of the risk allele at rs79965208 is higher in Caucasians than in Africans. Thus EwS susceptibility due to high *EGR2* expression is likely the consequence of a germline variant (Grünewald et al., 2015). Furthmermore, mSat instability was not observed for EwS (Monument et al., 2014). Moreover, the *MYBL2* associated mSat was tested in 76 matched pairs of EwS patients' normal and tumor samples, and no differences in the pairs were detected (Musa et al., 2019).

Despite the longest consecutive GGAA stretch, also additional nearby motifs with few interspersed bases correlated with higher EEts-binding rates. This association was most apparent when simply calculating the percentage of GGAA bases of all bases called by HipSTR besides the longest consecutive stretch. Another still unresolved issue is connected to these results: What should be considered as GGAA mSat in EwS? Several authors refer to consecutive GGAA repeats as a GGAA mSat (Gangwal et al., 2008; Guillon et al., 2009). Others allow also interruptions up to 20 bases between each motif (Johnson et al., 2017b), aggravating the exact definition of the start and end position of a mSat. Here, the mSat regions were defined with the parameters given to Tandem Repeats Finder. At least four consecutive GGAA repeats were demanded, and whenever the motif was interrupted by an alternate base, three motif bases compensated the mismatch, and one full GGAA motif an InDel. Other thresholds will result in other start and end positions of the mSats and the here described associations of the mSat architecture with EEts binding might differ. Hence, the number of consecutive GGAA motifs enables most precise and reproducible description of a GGAA mSats, and it was also strongest correlated with fusion binding (Pearson's r=0.99). Thus, in further analyses only the longest consecutive GGAA mSat was included.

Despite the number of GGAA motifs, higher rates of copy number gains in EEts-bound GGAA mSats were observed. This might be due to the applied method, as the more copies are present the more likely a binding might be precipitated and sequenced. Contrarily, it might be a hint that the recurrent CNVs in EwS, whose relevance was not investigated except for prognosis association (Mackintosh et al., 2012), supports EwS tumor initiation and/or progression. For instance, *NFAT5*, whose binding motif was found in 12.8% of genes upregulated upon EEts knockdown, hence likely downregulated in EwS, is located on chr16q

which is frequently lost in EwS (Tirode et al., 2014). Additionally, in few cases specific transcription factor binding sites nearby the mSat and the distance to the next GGAA mSat might have effect on the EEts-binding probability. However, only factors influencing the EEts binding probability were affirmed here, as the number of consecutive GGAA motifs, or newly introduced, but not any determinant was identified. Consequently, it remains to be elucidated whether EEts binding to GGAA mSats is a probabilistic event favored by some factors, or if there are more, still unknown influencing factors.

EEts binding externally of GGAA mSats to a single GGAA motif, as observed in the ESCLA, has been described previously (Guillon et al., 2009; Riggi et al., 2014). In context of the hypothesis, that fusion oncogene versus *ETS* transcription factor expression is tightly regulated, the occupation of the *ETS* transcription factors' binding sites might be another mechanism to reduce their effects. As *ETS* transcription factors in the *EEts* fusion lack one of normally two p300 binding domains, their transcriptonal activator potential is less than for wildtype *ETS* transcripton factors (Riggi et al., 2014).

The investigation of the above-mentioned factors influencing EEts-binding probability was motivated by the association of EEts binding to GGAA mSats with enhancer activity on nearby genes (Gangwal et al., 2008; Musa et al., 2019; Patel et al., 2012; Riggi et al., 2014). Indeed, the described GGAA-mSat enhancer activity seems to be dependent on *EEts* and thus exclusive for EwS: In a study on mSats associated with nearby gene regulation, the GGAA motif was not enriched in the regulative mSats compared to other mSats (OR 0.79, P=0.018). Moreover, of 1,380 fine-mapped mSats with associated gene regulation, only eight consisted out of the GGAA motif (Fotsing et al., 2019).

Surprisingly, when focusing on the closest gene for each GGAA mSat, the rate of regulated genes was not higher for bound than for unbound GGAA mSats. At a first glance, this appears contradictionary to previous studies, which successfully searched fusion-specific ChIP-Seq peaks nearby the candidate gene, and actually found a correlation of the mSat length with enhancer activity on the candidate gene (Gangwal et al., 2008; Marchetto et al., 2020; Musa et al., 2019). However, wide range enhancer activity of GGAA mSats was described (Guillon et al., 2009). Unpublished ChIA-PET data found enhancer to promoter looping with even 1 Mbp distance. Actually, virtually each gene is close enough to a GGAA mSat in the human genome to be regulated by such a mSat. Again in line with reports on the enhancer activity of EEts-bound GGAA mSats (Gangwal et al., 2008; Wei et al., 2010), the distance to the next GGAA mSat, especially an EEts-bound GGAA mSat, was smaller for genes downregulated upon EEts knockdown versus not regulated genes. Additionally, the number of bound GGAA mSats up to 1 Mbp distant from TSS was 43% higher nearby downregulated genes. The enhancer activity of the mSats might affect more genes

downstream than upstream, as the nearest bound GGAA mSat was more often upstream on the same, and downstram on the opposite strand. Although the closest bound GGAA mSats were slightly more often on the opposite strand and downstream, more GGAA mSats were found upstream in 1 Mbp distance. Besides these observations on the relative localization of GGAA mSats to regulated genes, transcription factor binding sites for NFAT5, NFY, and E2F family members were identified in the promoters of 13% of genes upregulated, and 28% and 11% of genes downregulated upon fusion knockdown, respectively. Interestingly, genes coding for these transcription factors appeared regulated by the fusion into the same direction as their targets. A potential synergism of EEts and transcription factors has been already described for E2F3 as well as an E2F enrichment in *EF1* target genes (Schwentner et al., 2015; Tomazou et al., 2015).

Thus, the here presented data support an enhancer activity of EEts-bound GGAA mSats, but indicate a multifactorial and probabilistic enhancer effect on target genes. For EEts binding, already several factors were identified, like the longest consecutive GGAA stretch, additional motifs, nearby mSats, and CNV. For the enhancer effect of EEts binding to GGAA mSats, additional factors were identified, like distance to TSS, relative position to TSS dependent on the strand, enrichment of nearby EEts-bound mSats, and transcription factor binding motifs in promoters. In light of all these parameters, some published experiments might need critical reevaluation. For instance, published luciferase reporter assays on GGAA mSats, cloned together with their flanking bases, were interpreted as indicator for the enhancer activity of the respective GGAA mSat (Dallmayer et al., 2019; Marchetto et al., 2020; Musa et al., 2019). Another more general interpretation, in line with the here presented data, might be that actually any GGAA mSat, independent on the flanking region, has EEts-mediated enhancer activity. This could be further investigated in reporter assays with synthetic GGAA mSats without flanking regions. As another example, CRISPRi for specific GGAA mSats was shown to impair target gene expression (Johnson et al., 2017a; Musa et al., 2019). However, as the results here did not support the concept of one GGAA mSat regulating one gene, probably also other genes might be affected by such CRISPRi experiments. Identification of those genes might further elucidate the interactions among GGAA mSats and with genes. The most relevant experiments to understand those interactions are likely chromosome conformation capture experiments. Unfortunately, these experiments currently lack precise positional resolution (Pal et al., 2019).

The mechanism of *EEts* mediated gene suppression was not further clarified, as the associations seen for genes downregulated upon *EEts* knockdown were usually far stronger than for the upregulated genes. Nevertheless, identification of the suppressive *EEts* effect on *NFAT5*, located on the recurrently lossed chr16q arm, might already explain the upregulation

of 13% of upregulated genes. The observed EEts binding to single GGAA motifs might also reason suppression of *ETS* transcription factor family targets. However, *NFAT5* has not been described to be regulated by *ETS* transcription factors. Additionally, an alternate localization of bound GGAA mSats for genes upregulated upon fusion knockdown was observed. The next bound GGAA mSats was more often upstream than downstream located, and more distant to the TSS. How such relative positions in the genome might translate into gene suppression, remains to be elucidated. Again, chromatin conformation capture might shed light on potential interactions of GGAA mSats with genes upregulated upon *EEts* knockdown, hence, likely in EwS suppressed genes.

Finally, 22 genes were identified as being heterogeneously regulated by *EEts* and associated with overall survival. Notably, 14 of these 22 genes were involved in cell cycle regulation. This further supports the notion that EwS requires strictly regulated cell cycle progression to stay genomically silent (Gorthi et al., 2018; Iniguez et al., 2018). Furthermore, the heterogeneous regulation and expression of these genes hint towards the employment of different cell-cycle regulating processes in distinct EwS tumors. Moreover, two transcription factors were among the 22 heterogeneously regulated genes, namely *E2F1* and *MYBL2*. The regulation of *E2F1* is in line with the enriched E2F family binding site in 11% of promoters of genes downregulated upon *EEts* knockdown. Furthermore, in each cell line with strongly affected *E2F1* expression upon fusion knockdown, the known *E2F* target *CENPI* (Thangavelu et al., 2017), another cell-cycle associated gene among the 22 genes, was at least moderately affected. As above-mentioned, *MYBL2* was reidentified and its clinical relevance for EwS has been already published (Musa et al., 2019). However, factors described in this thesis to impact EEts-mediated gene regulation were not sufficient to explain inter-cell line heterogeneity in the expression regulation of the 22 survival associated genes.

In summary, the ESCLA is a high-quality dataset and likely a rich resource for further investigations, which already reidentified and supported previous findings on EEts-fusion binding and genes regulated by *EEts*. Additionally, the ESCLA provided new insights into a rather probabilistic model of *EEts*-mediated gene regulation, new impulses for further investigations, emphasized the importance of chromatin conformation capture to better understand *EEts* interactions, and highlights the relevance of a well understood *EEts* biology to promote personalized medicine.

6. Conclusions and limitations

6.1. Conclusions

- Numerous EwS cell line models with inducible EWSR1-ETS fusion knockdown were generated as a resource for further investigations of fusion-mediated effects.

- A tissue microarray with *in vivo* expanded EwS cell lines with and without fusion knockdown (*EF1*) was constructed and enables *in situ* evaluation of fusion regulated proteins.

- The Ewing Sarcoma Cell Line Atlas - ESCLA - was generated with all constructed cell line models. The ESCLA is multidimensional and comprehensive with high quality data, enabling assessment of gene regulation, fusion binding, epigenetics and genetics of the cell line models.

- Analyses of the ESCLA corroborated previously described mutations, targets of EWSR1-ETS, EWSR1-ETS binding sites and colocalization with histone acetylation and super-enhancers. Furthermore, GGAA mSats were genotyped, additional (consistently) EWSR1-ETS regulated genes and consensus EWSR1-ETS binding sites were identified.

- First time systematic evaluation of commonalities between EwS cell lines with three distinct fusion types, *EF1* type 1, 2 and *EErg*, showed highly similar super-enhancer landscape, methylation profile and expression regulation, but different rates of chromoplexy.

- The number of consecutive GGAA motifs as main determinant of EWSR1-ETS binding to a GGAA mSat was validated by the ESCLA data. Additionally, copy number gains and transcription factor binding sites of e.g. GATA2 were identified to favor EWSR1-ETS binding. Previously reported differences in the distance of EWSR1-ETS regulated genes to EWSR1-ETS binding sites were refined. Moreover, the position of the GGAA mSat relative to the stand and transcription start site was shown to differ for genes regulated by EWSR1-ETS. Combination of transcriptome, ChIP and WGS data revealed that EWSR1-ETS regulated genes are enriched for binding sites of transcription factors, which were as well regulated by EWSR1-ETS, namely *NFAT5*, *NFYC* and *E2F* family members.

- Expression regulation heterogeneity was demonstrable in the ESCLA, in line with observations in primary EwS associated with survival.

- Hence, the ESCLA constitutes a rich resource for the Ewing sarcoma research community, promotes biomarker identification, models heterogeneity and, thus, might give inroads to personalized medicine, ultimately improving patient care.

6.2. Limitations

- This thesis bases on cell line models, which were shown here to be only similar, but not equal to primary EwS tumors regarding transcriptome and methylome. Hence, the results might be not fully representative for EwS and likely several true interactions in EwS are not ascertainable by the ESCLA. Further efforts should be made to find alternative EwS models and test them for their representativity for EwS. Additionally, culture conditions other than 2D (e.g. sphere formation, *in vivo* expansion) should be tested for their potential to adapt cell line behavior closer to primary EwS tumors.

- The fusion knockdown efficiency highly differed among the cell lines and had inconsistent effects on cell viability. Hence, some regulation was possibly missed out and artefacts might have occurred due to selective pressure. To increase the rate of accurately identified fusion regulated genes, similar knockdown levels are desirable, and might be achieved by further generation and testing of single cell clones and/or retransfection.

- The ESCLA was built with sequencers, analysers and processing pipelines in accordance with state of the art. Nevertheless, all analyses have limitations that will be, at least partly be overcome with technical improvements. The sequencing was performed with 150 bp reads; GGAA mSats near to this length or even longer can only be predicted with a certain rate of specificity. Long-read sequencing will enable the identification of far longer mSats and a clear phasing, but have currently still high error rates. ChIP is notoriously difficult to normalise. The results should be validated with independent replicates and sequencing with a uniform protocol. Additionally, the optimal ChIP-Seq protocol for read mapping to GGAA mSats has not been systematically evaluated yet. The microarray data were already filtered for

(presumably) translated genes, but several gene models are already deprecated and other predicted, here not considered genes have been experimentally validated. Thus, reanalyses of microarray data with updated gene lists as well as updated chip description files promise discovery of further fusion driven genes.

- All analyses on the ESCLA were performed with computational and statistical tools, most of which are based on several assumptions. Especially when running several analyses in a row, thus making many assumptions, false interpretations can easily occur. Hence, although analyses were done with appropriate tools and parameters, to the best of my knowledge, all results are and have to be understood as correlations and predictions which demand extensive experimental validation before translation into clinics.

- The analyses based on the ESCLA were focused on gene expression, especially for genes with clinical relevance, and further investigation of previously described and presumed expression regulating factors. Hence, many other potential regulatory aspects in EwS were not considered, but can partly be assessed in the ESCLA. For instance, ClariomD microarray also quantifies various long non-coding RNAs whose relevance for EwS was not investigated here. Network analysis might identify most relevant hubs in regulatory circuits. Additionally, the ESCLA might be extended, as the cell line models were stably generated. For instance, drug response assays in EEts-high versus -low cells and metabolomic analyses warrant further insights into the biology of EwS.

- Although previous reports on factors influencing EWSR1-ETS binding to GGAA mSats and EWSR1-ETS mediated gene expression regulation were confirmed and refined, and additional factors were identified, still binding and expression regulation were not fully predictable. Hence, some influencing factors were not detected and are still unknown. The genome-wide analyses counterspeaked the assumption of direct enhancer activity of an EWSR1-ETS bound GGAA mSat on the nearest gene. Instead, the here presented results hint towards a complex regulatory network over a relatively wide genomic range. Thus, to identify interacting sites in this regulatory network the ESCLA should be augmented with chromosome conformation capture analyses. Additionally, genome editing of GGAA mSats and subsequent genome wide ChIP-Seq and transcriptome analyses might elucidate the relevance of single mSats in EWSR1-ETS mediated gene expression regulation.

- The survival dataset was relatively large for a rare cancer, but small compared to studies on for example mamma carcinoma. Additionally, the overlap of genes described in the survival dataset and investigated by ClariomD microarray was 50.8%. A larger, possibly multi-centric survival cohort on the same microarray platform as the ESCLA or RNAseq analysis will likely increase the number of correctly identified differential fusion driven clinically relevant genes.

7. References

1000 Genomes Project Consortium, Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E., and McVean, G.A. (2010). A map of human genome variation from population-scale sequencing. Nature *467*, 1061–1073.

Abyzov, A., Urban, A.E., Snyder, M., and Gerstein, M. (2011). CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. Genome Res. *21*, 974–984.

Agra, N., Cidre, F., García-García, L., de la Parra, J., and Alonso, J. (2013). Lysyl Oxidase Is Downregulated by the EWS/FLI1 Oncoprotein and Its Propeptide Domain Displays Tumor Supressor Activities in Ewing Sarcoma Cells. PLoS ONE *8*, e66281.

de Alava, E., Kawai, A., Healey, J.H., Fligman, I., Meyers, P.A., Huvos, A.G., Gerald, W.L., Jhanwar, S.C., Argani, P., Antonescu, C.R., et al. (1998). EWS-FLI1 fusion transcript structure is an independent determinant of prognosis in Ewing's sarcoma. J. Clin. Oncol. *16*, 1248–1255.

Ambros, I.M., Ambros, P.F., Strehl, S., Kovar, H., Gadner, H., and Salzer-Kuntschik, M. (1991). MIC2 is a specific marker for Ewing's sarcoma and peripheral primitive neuroectodermal tumors. Evidence for a common histogenesis of Ewing's sarcoma and peripheral primitive neuroectodermal tumors from MIC2 expression and specific chromosome aberration. Cancer *67*, 1886–1893.

Anderson, N.D., de Borja, R., Young, M.D., Fuligni, F., Rosic, A., Roberts, N.D., Hajjar, S., Layeghifard, M., Novokmet, A., Kowalski, P.E., et al. (2018). Rearrangement bursts generate canonical gene fusions in bone and soft tissue tumors. Science *361*.

Andrews, S. (2019). Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data.

Antonescu, C. (2014). Round cell sarcomas beyond Ewing: emerging entities. Histopathology *64*, 26–37.

Applebaum, M.A., Worch, J., Matthay, K.K., Goldsby, R., Neuhaus, J., West, D.C., and DuBois, S.G. (2011). Clinical features and outcomes in patients with extraskeletal ewing sarcoma. Cancer *117*, 3027–3032.

Aryee, M.J., Jaffe, A.E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A.P., Hansen, K.D., and Irizarry, R.A. (2014). Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics *30*, 1363–1369.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. *25*, 25–29.

Augusto Corrêa dos Santos, R., Goldman, G.H., and Riaño-Pachón, D.M. (2017). ploidyNGS: visually exploring ploidy with Next Generation Sequencing data. Bioinformatics *33*, 2575–2576.

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.

Baca, S.C., Prandi, D., Lawrence, M.S., Mosquera, J.M., Romanel, A., Drier, Y., Park, K., Kitabayashi, N., MacDonald, T.Y., Ghandi, M., et al. (2013). Punctuated Evolution of Prostate Cancer Genomes. Cell *153*, 666–677.

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

Baldauf, M.C., Gerke, J.S., Kirschner, A., Blaeschke, F., Effenberger, M., Schober, K., Rubio, R.A., Kanaseki, T., Kiran, M.M., Dallmayer, M., et al. (2018b). Systematic identification of cancer-specific MHC-binding peptides with RAVEN. Oncoimmunology *7*, e1481558.

Bandrés, E., Malumbres, R., Escalada, A., Cubedo, E., Gonz??lez, I., Honorato, B., Zarate, R., Garc??a-Foncillas, J., and de Alava, E. (2005). Gene Expression Profile of Ewing Sarcoma Cell Lines Differing in Their EWS-FLI1 Fusion Type: J. Pediatr. Hematol. Oncol. *27*, 537–542.

Baumuratova, T., Surdez, D., Delyon, B., Stoll, G., Delattre, O., Radulescu, O., and Siegel, A. (2010). Localizing potentially active post-transcriptional regulations in the Ewing's sarcoma gene regulatory network. BMC Syst. Biol. *4*, 146.

Beck, R., Monument, M.J., Watkins, W.S., Smith, R., Boucher, K.M., Schiffman, J.D., Jorde, L.B., Randall, R.L., and Lessnick, S.L. (2012). EWS/FLI-responsive GGAA microsatellites exhibit polymorphic differences between European and African populations. Cancer Genet. *205*, 304–312.

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. *27*, 573–580.

Berger, M., Dirksen, U., Braeuninger, A., Koehler, G., Juergens, H., Krumbholz, M., and Metzler, M. (2013). Genomic EWS-FLI1 Fusion Sequences in Ewing Sarcoma Resemble Breakpoint Characteristics of Immature Lymphoid Malignancies. PLoS ONE *8*, e56408.

Bosma, S.E., Ayu, O., Fiocco, M., Gelderblom, H., and Dijkstra, P.D.S. (2018). Prognostic factors for survival in Ewing sarcoma: A systematic review. Surg. Oncol. *27*, 603–610.

Bosma, S.E., Lancia, C., Rueten-Budde, A.J., Ranft, A., Gelderblom, H., Fiocco, M., van de Sande, M.A.J., Dijkstra, P.D.S., and Dirksen, U. (2019). Easy-to-use clinical tool for survival estimation in Ewing sarcoma at diagnosis and after surgery. Sci. Rep. *9*, 11000.

Boulay, G., Sandoval, G.J., Riggi, N., Iyer, S., Buisson, R., Naigles, B., Awad, M.E., Rengarajan, S., Volorio, A., McBride, M.J., et al. (2017). Cancer-Specific Retargeting of BAF Complexes by a Prion-like Domain. Cell *171*, 163-178.e19.

Brenner, J.C., Feng, F.Y., Han, S., Patel, S., Goyal, S.V., Bou-Maroun, L.M., Liu, M., Lonigro, R., Prensner, J.R., Tomlins, S.A., et al. (2012). PARP-1 Inhibition as a Targeted Strategy to Treat Ewing's Sarcoma. Cancer Res. *72*, 1608–1613.

Brohl, A.S., Solomon, D.A., Chang, W., Wang, J., Song, Y., Sindiri, S., Patidar, R., Hurd, L., Chen, L., Shern, J.F., et al. (2014). The Genomic Landscape of the Ewing Sarcoma Family of Tumors Reveals Recurrent STAG2 Mutation. PLoS Genet. *10*, e1004475.

Cameron, D.L., Schröder, J., Penington, J.S., Do, H., Molania, R., Dobrovic, A., Speed, T.P., and Papenfuss, A.T. (2017). GRIDSS: sensitive and specific genomic rearrangement detection using positional de Bruijn graph assembly. Genome Res. *27*, 2050–2060.

Cameron, D.L., Baber, J., Shale, C., Valle-Inclan, J.E., Besselink, N., Cuppen, E., Priestley, P., and Papenfuss, A.T. (2020). GRIDSS2: harnessing the power of phasing and single breakends in somatic structural variant detection (Bioinformatics).

Capper, D., Jones, D.T.W., Sill, M., Hovestadt, V., Schrimpf, D., Sturm, D., Koelsche, C., Sahm, F., Chavez, L., Reuss, D.E., et al. (2018). DNA methylation-based classification of central nervous system tumours. Nature *555*, 469–474.

Carrillo, J., García-Aragoncillo, E., Azorín, D., Agra, N., Sastre, A., González-Mediero, I., García-Miguel, P., Pestaña, A., Gallego, S., Segura, D., et al. (2007). Cholecystokinin down-regulation by RNA interference impairs Ewing tumor growth. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. *13*, 2429–2440.

Carter, C.S., and Patel, R.M. (2019). Important Recently Characterized Non-Ewing Small Round Cell Tumors. Surg. Pathol. Clin. *12*, 191–215.

Çelik, H., Sciandra, M., Flashner, B., Gelmez, E., Kayraklıoğlu, N., Allegakoen, D.V., Petro, J.R., Conn, E.J., Hour, S., Han, J., et al. (2018). Clofarabine inhibits Ewing sarcoma growth through a novel molecular mechanism involving direct binding to CD99. Oncogene *37*, 2181–2196.

Chansky, H.A., Barahmand-pour, F., Mei, Q., Kahn-Farooqi, W., Zielinska-Kwiatkowska, A., Blackburn, M., Chansky, K., Conrad, E.U., Bruckner, J.D., Greenlee, T.K., et al. (2004). Targeting of EWS/FLI-1 by RNA interference attenuates the tumor phenotype of Ewing's sarcoma cells in vitro. J. Orthop. Res. *22*, 910–917.

Charville, G.W., Wang, W.-L., Ingram, D.R., Roy, A., Thomas, D., Patel, R.M., Hornick, J.L., van de Rijn, M., and Lazar, A.J. (2017). EWSR1 fusion proteins mediate PAX7 expression in Ewing sarcoma. Mod. Pathol. Off. J. U. S. Can. Acad. Pathol. Inc *30*, 1312–1320.

Chaturvedi, A., Hoffman, L.M., Welm, A.L., Lessnick, S.L., and Beckerle, M.C. (2012). The EWS/FLI Oncogene Drives Changes in Cellular Morphology, Adhesion, and Migration in Ewing Sarcoma. Genes Cancer *3*, 102–116.

Chen, K., Wallis, J.W., McLellan, M.D., Larson, D.E., Kalicki, J.M., Pohl, C.S., McGrath, S.D., Wendl, M.C., Zhang, Q., Locke, D.P., et al. (2009). BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. Nat. Methods *6*, 677–681.

Chen, S., Deniz, K., Sung, Y.-S., Zhang, L., Dry, S., and Antonescu, C.R. (2016). Ewing sarcoma with ERG gene rearrangements: A molecular study focusing on the prevalence of FUS-ERG and common pitfalls in detecting EWSR1-ERG fusions by FISH. Genes. Chromosomes Cancer *55*, 340–349.

Christensen, L., Joo, J., Lee, S., Wai, D., Triche, T.J., and May, W.A. (2013). FOXM1 Is an Oncogenic Mediator in Ewing Sarcoma. PLoS ONE *8*, e54556.

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

Cingolani, P., Patel, V.M., Coon, M., Nguyen, T., Land, S.J., Ruden, D.M., and Lu, X. (2012). Using Drosophila melanogaster as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. Front. Genet. *3*, 35.

Crompton, B.D., Stewart, C., Taylor-Weiner, A., Alexe, G., Kurek, K.C., Calicchio, M.L., Kiezun, A., Carter, S.L., Shukla, S.A., Mehta, S.S., et al. (2014). The genomic landscape of pediatric Ewing sarcoma. Cancer Discov. *4*, 1326–1341.

Daemen, A., Peterson, D., Sahu, N., McCord, R., Du, X., Liu, B., Kowanetz, K., Hong, R., Moffat, J., Gao, M., et al. (2015). Metabolite profiling stratifies pancreatic ductal adenocarcinomas into subtypes with distinct sensitivities to metabolic inhibitors. Proc. Natl. Acad. Sci. *112*, E4410–E4417.

Dallmayer, M., Li, J., Ohmura, S., Alba Rubio, R., Baldauf, M.C., Hölting, T.L.B., Musa, J., Knott, M.M.L., Stein, S., Cidre-Aranaz, F., et al. (2019). Targeting the CALCB/RAMP1 axis inhibits growth of Ewing sarcoma. Cell Death Dis. *10*, 116.

De Roeck, A., De Coster, W., Bossaerts, L., Cacace, R., De Pooter, T., Van Dongen, J., D'Hert, S., De Rijk, P., Strazisar, M., Van Broeckhoven, C., et al. (2019). NanoSatellite: accurate characterization of expanded tandem repeat length and sequence through whole genome long-read sequencing on PromethION. Genome Biol. *20*, 239.

Delattre, O., Zucman, J., Plougastel, B., Desmaze, C., Melot, T., Peter, M., Kovar, H., Joubert, I., de Jong, P., and Rouleau, G. (1992). Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours. Nature *359*, 162–165.

Desmaze, C., Brizard, F., Turc-Carel, C., Melot, T., Delattre, O., Thomas, G., and Aurias, A. (1997). Multiple chromosomal mechanisms generate an EWS/FLI1 or an EWS/ERG fusion gene in Ewing tumors. Cancer Genet. Cytogenet. *97*, 12–19.

Dirks, W.G., MacLeod, R.A.F., Nakamura, Y., Kohara, A., Reid, Y., Milch, H., Drexler, H.G., and Mizusawa, H. (2010). Cell line cross-contamination initiative: an interactive reference database of STR profiles covering common cancer cell lines. Int. J. Cancer *126*, 303–304.

Elzi, D.J., Song, M., Houghton, P.J., Chen, Y., and Shiio, Y. (2015). The role of FLI-1-EWS, a fusion gene reciprocal to EWS-FLI-1, in Ewing sarcoma. Genes Cancer *6*, 452–461.

Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics *32*, 3047–3048.

Ewing, J. (2006). THE CLASSIC: Diffuse Endothelioma of Bone. Clin. Orthop. *450*, 25–27.

Fletcher, C.D.M. (2014). The evolving classification of soft tissue tumours - an update based on the new 2013 WHO classification. Histopathology *64*, 2–11.

Fotsing, S.F., Margoliash, J., Wang, C., Saini, S., Yanicky, R., Shleizer-Burko, S., Goren, A., and Gymrek, M. (2019). The impact of short tandem repeat variation on gene expression. Nat. Genet. *51*, 1652–1659.

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

Fu, S., Wang, A., and Au, K.F. (2019). A comparative evaluation of hybrid error correction methods for error-prone long reads. Genome Biol. *20*, 26.

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., et al. (2008). Microsatellites as EWS/FLI response elements in Ewing's sarcoma. Proc. Natl. Acad. Sci. U. S. A. *105*, 10149–10154.

Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. ArXiv12073907 Q-Bio.

Gaspar, N., Hawkins, D.S., Dirksen, U., Lewis, I.J., Ferrari, S., Le Deley, M.-C., Kovar, H., Grimer, R., Whelan, J., Claude, L., et al. (2015). Ewing Sarcoma: Current Management and Future Approaches Through Collaboration. J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol. *33*, 3036–3046.

Ginsberg, J.P., Goodman, P., Leisenring, W., Ness, K.K., Meyers, P.A., Wolden, S.L., Smith, S.M., Stovall, M., Hammond, S., Robison, L.L., et al. (2010). Long-term Survivors of Childhood Ewing Sarcoma: Report From the Childhood Cancer Survivor Study. JNCI J. Natl. Cancer Inst. *102*, 1272–1283.

Giovannini, M., Biegel, J.A., Serra, M., Wang, J.Y., Wei, Y.H., Nycum, L., Emanuel, B.S., and Evans, G.A. (1994). EWS-erg and EWS-Fli1 fusion transcripts in Ewing's sarcoma and primitive neuroectodermal tumors with variant translocations. J. Clin. Invest. *94*, 489–496.

Gomez, N.C., Hepperla, A.J., Dumitru, R., Simon, J.M., Fang, F., and Davis, I.J. (2016). Widespread Chromatin Accessibility at Repetitive Elements Links Stem Cells with Human Cancer. Cell Rep. *17*, 1607–1620.

Gordon, D.J., Motwani, M., and Pellman, D. (2016). Modeling the initiation of Ewing sarcoma tumorigenesis in differentiating human embryonic stem cells. Oncogene *35*, 3092–3102.

Gorthi, A., Romero, J.C., Loranc, E., Cao, L., Lawrence, L.A., Goodale, E., Iniguez, A.B., Bernard, X., Masamsetti, V.P., Roston, S., et al. (2018). EWS-FLI1 increases transcription to cause R-loops and block BRCA1 repair in Ewing sarcoma. Nature *555*, 387–391.

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

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.

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. Primer *4*, 5.

Gu, Z., Gu, L., Eils, R., Schlesner, M., and Brors, B. (2014). circlize implements and enhances circular visualization in R. Bioinformatics *30*, 2811–2812.

Gu, Z., Eils, R., Schlesner, M., and Ishaque, N. (2018). EnrichedHeatmap: an R/Bioconductor package for comprehensive visualization of genomic signal associations. BMC Genomics *19*, 234.

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.

Gymrek, M., Willems, T., Guilmatre, A., Zeng, H., Markus, B., Georgiev, S., Daly, M.J., Price, A.L., Pritchard, J.K., Sharp, A.J., et al. (2016). Abundant contribution of short tandem repeats to gene expression variation in humans. Nat. Genet. *48*, 22–29.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Hirsch, C., and Schildknecht, S. (2019). In Vitro Research Reproducibility: Keeping Up High Standards. Front. Pharmacol. *10*, 1484.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. *37*, 1–13.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57.

Hulsen, T., de Vlieg, J., and Alkema, W. (2008). BioVenn – a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. BMC Genomics *9*, 488.

Hung, Y.P., Fletcher, C.D.M., and Hornick, J.L. (2016). Evaluation of NKX2-2 expression in round cell sarcomas and other tumors with EWSR1 rearrangement: imperfect specificity for Ewing sarcoma. Mod. Pathol. *29*, 370–380.

Hung, Y.P., Lee, J.P., Bellizzi, A.M., and Hornick, J.L. (2017). PHOX2B reliably distinguishes neuroblastoma among small round blue cell tumours. Histopathology *71*, 786–794.

Iniguez, A.B., Stolte, B., Wang, E.J., Conway, A.S., Alexe, G., Dharia, N.V., Kwiatkowski, N., Zhang, T., Abraham, B.J., Mora, J., et al. (2018). EWS/FLI Confers Tumor Cell Synthetic Lethality to CDK12 Inhibition in Ewing Sarcoma. Cancer Cell *33*, 202-216.e6.

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.

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. (2017a). Role for the EWS domain of EWS/FLI in binding GGAA-microsatellites required for Ewing sarcoma anchorage independent growth. Proc. Natl. Acad. Sci. U. S. A. *114*, 9870–9875.

Johnson, K.M., Taslim, C., Saund, R.S., and Lessnick, S.L. (2017b). Identification of two types of GGAA-microsatellites and their roles in EWS/FLI binding and gene regulation in Ewing sarcoma. PloS One *12*, e0186275.

Johnson, W.E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostat. Oxf. Engl. *8*, 118–127.

Kaatsch, P., Strothotte, J., Becker, C., Bielack, S., Dirksen, U., and Blettner, M. (2016). Pediatric bone tumors in Germany from 1987 to 2011: incidence rates, time trends and survival. Acta Oncol. Stockh. Swed. *55*, 1145–1151.

Kapałczyńska, M., Kolenda, T., Przybyła, W., Zajączkowska, M., Teresiak, A., Filas, V., Ibbs, M., Bliźniak, R., Łuczewski, Ł., and Lamperska, K. (2016). 2D and 3D cell cultures – a comparison of different types of cancer cell cultures. Arch. Med. Sci.

Kar, A., and Gutierrez-Hartmann, A. (2013). Molecular mechanisms of ETS transcription factor-mediated tumorigenesis. Crit. Rev. Biochem. Mol. Biol. *48*, 522–543.

Katschnig, A.M., Kauer, M.O., Schwentner, R., Tomazou, E.M., Mutz, C.N., Linder, M., Sibilia, M., Alonso, J., Aryee, D.N.T., and Kovar, H. (2017). EWS-FLI1 perturbs MRTFB/YAP-1/TEAD target gene regulation inhibiting cytoskeletal autoregulatory feedback in Ewing sarcoma. Oncogene *36*, 5995–6005.

Kedage, V., Selvaraj, N., Nicholas, T.R., Budka, J.A., Plotnik, J.P., Jerde, T.J., and Hollenhorst, P.C. (2016). An Interaction with Ewing's Sarcoma Breakpoint Protein EWS Defines a Specific Oncogenic Mechanism of ETS Factors Rearranged in Prostate Cancer. Cell Rep. *17*, 1289–1301.

Kinsey, M., Smith, R., and Lessnick, S.L. (2006). NR0B1 Is Required for the Oncogenic Phenotype Mediated by EWS/FLI in Ewing's Sarcoma. Mol. Cancer Res. *4*, 851–859.

Kridis, W.B., Toumi, N., Chaari, H., Khanfir, A., Ayadi, K., Keskes, H., Boudawara, T., Daoud, J., and Frikha, M. (2017). A Review of Ewing Sarcoma Treatment: Is it Still a Subject of Debate? Rev. Recent Clin. Trials *12*, 19–23.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: An information aesthetic for comparative genomics. Genome Res. *19*, 1639–1645.

Lan, Y., Xiao, X., Luo, Y., He, Z., and Song, X. (2018). FEZF1 is an Independent Predictive Factor for Recurrence and Promotes Cell Proliferation and Migration in Cervical Cancer. J. Cancer *9*, 3929–3938.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Layer, R.M., Chiang, C., Quinlan, A.R., and Hall, I.M. (2014). LUMPY: a probabilistic framework for structural variant discovery. Genome Biol. *15*, R84.

Le Deley, M.-C., Delattre, O., Schaefer, K.-L., Burchill, S.A., Koehler, G., Hogendoorn, P.C.W., Lion, T., Poremba, C., Marandet, J., Ballet, S., et al. (2010). Impact of *EWS-ETS* Fusion Type on Disease Progression in Ewing's Sarcoma/Peripheral Primitive Neuroectodermal Tumor: Prospective Results From the Cooperative Euro-E.W.I.N.G. 99 Trial. J. Clin. Oncol. *28*, 1982–1988.

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.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinforma. Oxf. Engl. *25*, 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinforma. Oxf. Engl. *25*, 2078–2079.

Liu, F., Zhang, Y., Zhang, L., Li, Z., Fang, Q., Gao, R., and Zhang, Z. (2019). Systematic comparative analysis of single-nucleotide variant detection methods from single-cell RNA sequencing data. Genome Biol. *20*, 242.

Longhi, A., Ferrari, S., Tamburini, A., Luksch, R., Fagioli, F., Bacci, G., and Ferrari, C. (2012). Late effects of chemotherapy and radiotherapy in osteosarcoma and Ewing sarcoma patients: The Italian Sarcoma Group Experience (1983-2006). Cancer *118*, 5050–5059.

Lovén, J., Hoke, H.A., Lin, C.Y., Lau, A., Orlando, D.A., Vakoc, C.R., Bradner, J.E., Lee, T.I., and Young, R.A. (2013). Selective inhibition of tumor oncogenes by disruption of super-enhancers. Cell *153*, 320–334.

Luo, W., Xu, C., Phillips, S., Gardenswartz, A., Rosenblum, J.M., Ayello, J., Lessnick, S.L., Hao, H.-X., and Cairo, M.S. (2020). Protein phosphatase 1 regulatory subunit 1A regulates cell cycle progression in Ewing sarcoma. Oncotarget *11*, 1691–1704.

Machado, I., Noguera, R., Pellin, A., Lopez-Guerrero, J.A., Piqueras, M., Navarro, S., and Llombart-Bosch, A. (2009). Molecular diagnosis of Ewing sarcoma family of tumors: a comparative analysis of 560 cases with FISH and RT-PCR. Diagn. Mol. Pathol. Am. J. Surg. Pathol. Part B *18*, 189–199.

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.

Mackintosh, C., Ordóñez, J.L., García-Domínguez, D.J., Sevillano, V., Llombart-Bosch, A., Szuhai, K., Scotlandi, K., Alberghini, M., Sciot, R., Sinnaeve, F., et al. (2012). 1q gain and CDT2 overexpression underlie an aggressive and highly proliferative form of Ewing sarcoma. Oncogene *31*, 1287–1298.

Marchetto, A., Ohmura, S., Orth, M.F., Knott, M.M.L., Colombo, M.V., Arrigoni, C., Bardinet, V., Saucier, D., Wehweck, F.S., Li, J., et al. (2020). Oncogenic hijacking of a developmental transcription factor evokes vulnerability toward oxidative stress in Ewing sarcoma. Nat. Commun. *11*.

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. *90*, 5752–5756.

May, W.A., Grigoryan, R.S., Keshelava, N., Cabral, D.J., Christensen, L.L., Jenabi, J., Ji, L., Triche, T.J., Lawlor, E.R., and Reynolds, C.P. (2013). Characterization and Drug Resistance Patterns of Ewing's Sarcoma Family Tumor Cell Lines. PLoS ONE *8*, e80060.

Mi, H., Muruganujan, A., Ebert, D., Huang, X., and Thomas, P.D. (2019). PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Res. *47*, D419–D426.

Miller, H.E., Gorthi, A., Bassani, N., Lawrence, L.A., Iskra, B.S., and Bishop, A.J.R. (2020). Reconstruction of Ewing Sarcoma Developmental Context from Mass-Scale Transcriptomics Reveals Characteristics of EWSR1-FLI1 Permissibility. Cancers *12*, 948.

Mirkin, S.M. (2007). Expandable DNA repeats and human disease. Nature *447*, 932–940.

Mitra, R., Chen, X., Greenawalt, E.J., Maulik, U., Jiang, W., Zhao, Z., and Eischen, C.M. (2017). Decoding critical long non-coding RNA in ovarian cancer epithelial-to-mesenchymal transition. Nat. Commun. *8*, 1604.

Monument, M.J., Johnson, K.M., McIlvaine, E., Abegglen, L., Watkins, W.S., Jorde, L.B., Womer, R.B., Beeler, N., Monovich, L., Lawlor, E.R., et al. (2014). Clinical and biochemical function of polymorphic NR0B1 GGAA-microsatellites in Ewing sarcoma: a report from the Children's Oncology Group. PloS One *9*, e104378.

Mousavi, N., Margoliash, J., Pusarla, N., Saini, S., Yanicky, R., and Gymrek, M. (2020). TRTools: a toolkit for genome-wide analysis of tandem repeats. Bioinformatics btaa736.

Musa, J., Cidre-Aranaz, F., Aynaud, M.-M., Orth, M.F., Knott, M.M.L., Mirabeau, O., Mazor, G., Varon, M., Hölting, T.L.B., Grossetête, S., et al. (2019). Cooperation of cancer drivers with regulatory germline variants shapes clinical outcomes. Nat. Commun. *10*, 4128.

Niepel, M., Hafner, M., Mills, C.E., Subramanian, K., Williams, E.H., Chung, M., Gaudio, B., Barrette, A.M., Stern, A.D., Hu, B., et al. (2019). A Multi-center Study on the Reproducibility of Drug-Response Assays in Mammalian Cell Lines. Cell Syst. *9*, 35-48.e5.

Novina, C.D., and Sharp, P.A. (2004). The RNAi revolution. Nature *430*, 161–164.

Orth, M.F., Hölting, T.L.B., Dallmayer, M., Wehweck, F.S., Paul, T., Musa, J., Baldauf, M.C., Surdez, D., Delattre, O., Knott, M.M.L., et al. (2020). High Specificity of BCL11B and GLG1 for EWSR1-FLI1 and EWSR1-ERG Positive Ewing Sarcoma. Cancers *12*, 644.

Ottaviano, L., Schaefer, K.-L., Gajewski, M., Huckenbeck, W., Baldus, S., Rogel, U., Mackintosh, C., de Alava, E., Myklebost, O., Kresse, S.H., et al. (2010). Molecular characterization of commonly used cell lines for bone tumor research: a trans-European EuroBoNet effort. Genes. Chromosomes Cancer *49*, 40–51.

Pal, K., Forcato, M., and Ferrari, F. (2019). Hi-C analysis: from data generation to integration. Biophys. Rev. *11*, 67–78.

Papp, G., Mihály, D., and Sápi, Z. (2017). Unusual Signal Patterns of Break-apart FISH Probes Used in the Diagnosis of Soft Tissue Sarcomas. Pathol. Oncol. Res. *23*, 863–871.

Patel, M., Simon, J.M., Iglesia, M.D., Wu, S.B., McFadden, A.W., Lieb, J.D., and Davis, I.J. (2012). Tumor-specific retargeting of an oncogenic transcription factor chimera results in dysregulation of chromatin and transcription. Genome Res. *22*, 259–270.

Pattenden, S.G., Simon, J.M., Wali, A., Jayakody, C.N., Troutman, J., McFadden, A.W., Wooten, J., Wood, C.C., Frye, S.V., Janzen, W.P., et al. (2016). High-throughput small molecule screen identifies inhibitors of aberrant chromatin accessibility. Proc. Natl. Acad. Sci. *113*, 3018–3023.

PDQ Pediatric Treatment Editorial Board (2002). Ewing Sarcoma Treatment (PDQ®): Health Professional Version. In PDQ Cancer Information Summaries, (Bethesda (MD): National Cancer Institute (US)), p.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. Cell *152*, 1173–1183.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Remmele, W., and Stegner, H.E. (1987). [Recommendation for uniform definition of an immunoreactive score (IRS) for immunohistochemical estrogen receptor detection (ER-ICA) in breast cancer tissue]. Pathol. *8*, 138–140.

Richard, G.-F., Kerrest, A., and Dujon, B. (2008). Comparative Genomics and Molecular Dynamics of DNA Repeats in Eukaryotes. Microbiol. Mol. Biol. Rev. *72*, 686–727.

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.

Rokita, J.L., Rathi, K.S., Cardenas, M.F., Upton, K.A., Jayaseelan, J., Cross, K.L., Pfeil, J., Egolf, L.E., Way, G.P., Farrel, A., et al. (2019). Genomic Profiling of Childhood Tumor Patient-Derived Xenograft Models to Enable Rational Clinical Trial Design. Cell Rep. *29*, 1675-1689.e9.

Rossi, S., Orvieto, E., Furlanetto, A., Laurino, L., Ninfo, V., and Tos, A.P.D. (2004). Utility of the immunohistochemical detection of FLI-1 expression in round cell and vascular neoplasm using a monoclonal antibody. Mod. Pathol. *17*, 547–552.

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.

Sand, L., Szuhai, K., and Hogendoorn, P. (2015). Sequencing Overview of Ewing Sarcoma: A Journey across Genomic, Epigenomic and Transcriptomic Landscapes. Int. J. Mol. Sci. *16*, 16176–16215.

Sannino, G., Marchetto, A., Ranft, A., Jabar, S., Zacherl, C., Alba-Rubio, R., Stein, S., Wehweck, F.S., Kiran, M.M., Hölting, T.L.B., et al. (2019). Gene expression and immunohistochemical analyses identify SOX2 as major risk factor for overall survival and relapse in Ewing sarcoma patients. EBioMedicine *47*, 156–162.

Sbaraglia, M., Righi, A., Gambarotti, M., and Dei Tos, A.P. (2020). Ewing sarcoma and Ewing-like tumors. Virchows Arch. Int. J. Pathol. *476*, 109–119.

Schwentner, R., Papamarkou, T., Kauer, M.O., Stathopoulos, V., Yang, F., Bilke, S., Meltzer, P.S., Girolami, M., and Kovar, H. (2015). EWS-FLI1 employs an E2F switch to drive target gene expression. Nucleic Acids Res. *43*, 2780–2789.

Sheffield, N.C., Pierron, G., Klughammer, J., Datlinger, P., Schönegger, A., Schuster, M., Hadler, J., Surdez, D., Guillemot, D., Lapouble, E., et al. (2017). DNA methylation heterogeneity defines a disease spectrum in Ewing sarcoma. Nat. Med. *23*, 386–395.

Sherry, S.T. (2001). dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. *29*, 308–311.

Shigemizu, D., Fujimoto, A., Akiyama, S., Abe, T., Nakano, K., Boroevich, K.A., Yamamoto, Y., Furuta, M., Kubo, M., Nakagawa, H., et al. (2013). A practical method to detect SNVs and indels from whole genome and exome sequencing data. Sci. Rep. *3*, 2161.

Siolas, D., and Hannon, G.J. (2013). Patient-Derived Tumor Xenografts: Transforming Clinical Samples into Mouse Models. Cancer Res. *73*, 5315–5319.

Smith, R., Owen, L.A., Trem, D.J., Wong, J.S., Whangbo, J.S., Golub, T.R., and Lessnick, S.L. (2006). Expression profiling of EWS/FLI identifies NKX2.2 as a critical target gene in Ewing's sarcoma. Cancer Cell *9*, 405–416.

Sorensen, P.H., Lessnick, S.L., Lopez-Terrada, D., Liu, X.F., Triche, T.J., and Denny, C.T. (1994). A second Ewing's sarcoma translocation, t(21;22), fuses the EWS gene to another ETS-family transcription factor, ERG. Nat. Genet. *6*, 146–151.

Stephens, P.J., Greenman, C.D., Fu, B., Yang, F., Bignell, G.R., Mudie, L.J., Pleasance, E.D., Lau, K.W., Beare, D., Stebbings, L.A., et al. (2011). Massive genomic rearrangement acquired in a single catastrophic event during cancer development. Cell *144*, 27–40.

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., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U. S. A. *102*, 15545–15550.

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

Talevich, E., Shain, A.H., Botton, T., and Bastian, B.C. (2016). CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. PLOS Comput. Biol. *12*, e1004873.

Tange, Ole (2020). GNU Parallel 20201022 ('Samuel Paty') (Zenodo).

Tate, J.G., Bamford, S., Jubb, H.C., Sondka, Z., Beare, D.M., Bindal, N., Boutselakis, H., Cole, C.G., Creatore, C., Dawson, E., et al. (2019). COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res. *47*, D941–D947.

Thangavelu, P.U., Lin, C.-Y., Vaidyanathan, S., Nguyen, T.H.M., Dray, E., and Duijf, P.H.G. (2017). Overexpression of the E2F target gene *CENPI* promotes chromosome instability and predicts poor prognosis in estrogen receptor-positive breast cancer. Oncotarget *8*, 62167–62182.

Theisen, E.R., Selich-Anderson, J., Miller, K.R., Tanner, J.M., Taslim, C., Pishas, K.I., Sharma, S., and Lessnick, S.L. (2020). Chromatin profiling reveals relocalization of lysine-specific demethylase 1 by an oncogenic fusion protein. Epigenetics 1–20.

Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform. *14*, 178–192.

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.

Tirode, F., Surdez, D., Ma, X., Parker, M., Le Deley, M.C., Bahrami, A., Zhang, Z., Lapouble, E., Grossetête-Lalami, S., Rusch, M., et al. (2014). Genomic landscape of Ewing sarcoma defines an aggressive subtype with co-association of STAG2 and TP53 mutations. Cancer Discov. *4*, 1342–1353.

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.

Tsherniak, A., Vazquez, F., Montgomery, P.G., Weir, B.A., Kryukov, G., Cowley, G.S., Gill, S., Harrington, W.F., Pantel, S., Krill-Burger, J.M., et al. (2017). Defining a Cancer Dependency Map. Cell *170*, 564-576.e16.

Tsuda, Y., Dickson, B.C., Swanson, D., Sung, Y.-S., Zhang, L., Meyers, P., Healey, J.H., and Antonescu, C.R. (2020). Ewing sarcoma with FEV gene rearrangements is a rare subset with predilection for extraskeletal locations and aggressive behavior. Genes. Chromosomes Cancer *59*, 286–294.

Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr. Protoc. Bioinforma. *43*, 11.10.1-11.10.33.

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. *38*, e164.

Wei, G.-H., Badis, G., Berger, M.F., Kivioja, T., Palin, K., Enge, M., Bonke, M., Jolma, A., Varjosalo, M., Gehrke, A.R., et al. (2010). Genome-wide analysis of ETS-family DNA-binding in vitro and in vivo. EMBO J. *29*, 2147–2160.

Welford, S.M., Hebert, S.P., Deneen, B., Arvand, A., and Denny, C.T. (2001). DNA Binding Domain-independent Pathways Are Involved in EWS/FLI1-mediated Oncogenesis. J. Biol. Chem. *276*, 41977–41984.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307–319.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (New York: Springer-Verlag).

Wiederschain, D., Wee, S., Chen, L., Loo, A., Yang, G., Huang, A., Chen, Y., Caponigro, G., Yao, Y.-M., Lengauer, C., et al. (2009). Single-vector inducible lentiviral RNAi system for oncology target validation. Cell Cycle Georget. Tex *8*, 498–504.

Willems, T., Zielinski, D., Yuan, J., Gordon, A., Gymrek, M., and Erlich, Y. (2017). Genome-wide profiling of heritable and *de novo* STR variations. Nat. Methods *14*, 590–592.

Wong, K.K., and Hussain, F.A. (2020). TRPM4 is overexpressed in breast cancer associated with estrogen response and epithelial-mesenchymal transition gene sets. PLOS ONE *15*, e0233884.

Wright, M.N., Gola, D., and Ziegler, A. (2017). Preprocessing and Quality Control for Whole-Genome Sequences from the Illumina HiSeq X Platform. Methods Mol. Biol. Clifton NJ *1666*, 629–647.

Zaccarini, D.J., Deng, X., Tull, J., Maciak, C., Valente, A.L., and Zhang, S. (2018). Expression of TLE-1 and CD99 in Carcinoma: Pitfalls in Diagnosis of Synovial Sarcoma. Appl. Immunohistochem. Mol. Morphol. AIMM *26*, 368–373.

Zhang, Q., Zeng, X., Younkin, S., Kawli, T., Snyder, M.P., and Keleş, S. (2016). Systematic evaluation of the impact of ChIP-seq read designs on genome coverage, peak identification, and allele-specific binding detection. BMC Bioinformatics *17*, 96.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

Zucman, J., Melot, T., Desmaze, C., Ghysdael, J., Plougastel, B., Peter, M., Zucker, J.M., Triche, T.J., Sheer, D., and Turc-Carel, C. (1993). Combinatorial generation of variable fusion proteins in the Ewing family of tumours. EMBO J. *12*, 4481–4487.

8. Appendix

8.1. List of figures

8.2. List of tables

8.3. Abbreviations

ChIP-Seq	Chromatin immunoprecipitation and high-throughput sequencing
CNV	Copy number variation
DIN	DNA integrity number
DMR	Differentially methylated region
Dox	Doxycycline
EErg	EWSR1-ERG
EEts	EWSR1 fused to ETS transcription factor

EF1	EWSR1-FLI1
ESCLA	Ewing sarcoma cell line atlas
EwS	Ewing sarcoma
FC	Fold change
FCS	Fetal calf serum
GO	Gene Ontology
GSEA	Gene set enrichment analysis
HRP	Horseradish peroxidase
InDel	Insertion or deletion
IHC	Immunohistochemistry
INV	Inversion
LLD	Lowest lethal dose
miRNA	micro RNA
mSat	Microsatellite
NES	Normalized enrichment score
PAGE	Polyacrylamid gel electrophoresis
PCR	Polymerase chain reaction
qRT-PCR	quantitative real-time PCR
RNAi	RNA interference
RPKM	Reads per kilo base per million mapped reads
SE	Super-enhancer
SEM	Standard error of the mean
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
shRNA	Short hairpin RNA
siRNA	Short interfering RNA
SNV	Single nucleotide variant

SV	Structural variant
TAC	Transcriptome Analysis Console
TAE	Tris-acetate-EDTA
TMA	Tissue microarray
t-SNE	t-Distributed Stochastic Neighbor Embedding
TSS	Transcription start site
TTS	Transcription termination site
WGS	Whole genome sequencing

9. Acknowledgements

First, I want to thank the MMRS and the hosting Faculty of Medicine of the LMU Munich for offering the Ph.D. medical research program, for the admission of my application and guidance during my studies.

Next, I want to thank my thesis advisory committee, namely PD Dr. Dr. Thomas G. P. Grünewald, Prof. Dr. rer. nat. Andreas Jung and Prof. Dr. med. Thomas Kirchner for their supervision and support. I thank Prof. Dr. Thomas Kirchner for the allowance to conduct my thesis at his institute. Special thanks go to PD Dr. Dr. Grünewald for assignment of the Ph.D. project and the position in frame of his Max-Eder grant, and his constant input and supervision. Furthermore, I want to thank all members of AG Grünewald who provided a friendly, professional and synergistic work environment and supported me. I want to express my graditute for the many helpful people at the Institute of Pathology of the LMU Munich, who supported me in several aspects, especially Sabine Sagebiel-Kohler for her constant willingness to help, Anja Heier and Andrea Sendelhofert for histological stainings, and Mario Gipp and Heike Prelle for TMA construction.

For help in conducting ChIP experiments, hosting me for two weeks for an internship, sequencing service, processing and providing data I thank Dr. Olivier Delattre, Dr. Didier Surdez, Sakina Zaidi and Dr. Sandrine Grossetête-Lalami from the Insitut Curie Research Center, Paris, France.

For her guidance into bioinformatics and a three weeks internship I want to thank Prof. Melissa Gymrek, UCSD, San Diego, California, USA.

I also want to thank those who cared externally for expert sequencing and chip analyses of my samples, the team of IMGM Laboratories (Martinsried, Germany), High Throughput Sequencing Unit of the Core Facilities Genomics and Proteomics at the German Cancer Research Center (Heidelberg, Germany), the Protein Science research unit and the Genotyping Platform of the Genome Analysis Center of the German Research Center for Environmental Health (Munich and Neuherberg, Germany), especially Dr. Stefanie Hauck, Dr. Fabian Metzger, Dr. Eva Reischl and Dr. Jennifer Kriebel. For support in interpreting methylome data and providing a t-SNE plot my thanks go to Dr. Martin Sill (DKFZ, Heidelberg, Germany).

Furthermore, I thank Dr. Aruna Marchetto (Molecular Genetics, Pränatal-Medizin, Munich, Germany) and Gabriele Meyer (DKFZ, Heidelberg, Germany) for proofreading my thesis and

Michael Hain (DKFZ, Heidelberg, Germany) for providing server infrastructure for bioinformatic analyses.

For financing my position, I want to thank the Deutsche Krebshilfe and the Kind-Philipp-Stiftung for granting a travel grant.

Special thanks go also to my family and friends who supported me over the years.

This thesis would not have been possible without the help of all these people. Again, thank you all for this great experience!

10. Scientific record

10.1. Publications

As from 02.12.2020

- Authorships on original research articles (thereof first-authorships): 21 (4)
- Authorships on review articles (thereof first-authorships) 3 (2)
- Citations: 779
- H-Index: 12

Galai, G., Ben-David, H., Levin, L., **Orth, M.F.**, Grünewald, T.G.P., Pilosof, S., Bershtein, S., and Rotblat, B. (2020). Pan-Cancer Analysis of Mitochondria Chaperone-Client Co-Expression Reveals Chaperone Functional Partitioning. **Cancers** 12, 825.

Gerke, J.S., **Orth, M.F.**, Tolkach, Y., Romero-Pérez, L., Wehweck, F.S., Stein, S., Musa, J., Knott, M.M.L., Hölting, T.L.B., Li, J., et al. (2020). Integrative clinical transcriptome analysis reveals TMPRSS2-ERG dependency of prognostic biomarkers in prostate adenocarcinoma. **Int. J. Cancer** 146, 2036–2046.

Herrmann, A.B., Müller, M.-L., **Orth, M.F.**, Müller, J.P., Zernecke, A., Hochhaus, A., Ernst, T., Butt, E., and Frietsch, J.J. (2020). Knockout of LASP1 in CXCR4 expressing CML cells promotes cell persistence, proliferation and TKI resistance. **J. Cell. Mol. Med.**

Marchetto, A., Ohmura, S., **Orth, M.F.**, Knott, M.M.L., Colombo, M.V., Arrigoni, C., Bardinet, V., Saucier, D., Wehweck, F.S., Li, J., et al. (2020). Oncogenic hijacking of a developmental transcription factor evokes vulnerability toward oxidative stress in Ewing sarcoma. **Nat. Commun.** 11.

Orth, M.F., Buecklein, V.L., Kampmann, E., Subklewe, M., Noessner, E., Cidre-Aranaz, F., Romero-Pérez, L., Wehweck, F.S., Lindner, L., Issels, R., et al. (2020). A comparative view on the expression patterns of PD-L1 and PD-1 in soft tissue sarcomas. Cancer Immunol. Immunother. **CII** 69, 1353–1362.

Orth, M.F., Hölting, T.L.B., Dallmayer, M., Wehweck, F.S., Paul, T., Musa, J., Baldauf, M.C., Surdez, D., Delattre, O., Knott, M.M.L., et al. (2020). High Specificity of BCL11B and GLG1 for EWSR1-FLI1 and EWSR1-ERG Positive Ewing Sarcoma. **Cancers** 12, 644.

Steinestel, K., Trautmann, M., Jansen, E.-P., Dirksen, U., Rehkämper, J., Mikesch, J.-H., Gerke, J.S., **Orth, M.F.**, Sannino, G., Arteaga, M.-F., et al. (2020). Focal adhesion kinase confers pro-migratory and antiapoptotic properties and is a potential therapeutic target in Ewing sarcoma. **Mol. Oncol.** 14, 248–260.

Triki, M., Rinaldi, G., Planque, M., Broekaert, D., Winkelkotte, A.M., Maier, C.R., Janaki Raman, S., Vandekeere, A., Van Elsen, J., **Orth, M.F.**, et al. (2020). mTOR Signaling and SREBP Activity Increase FADS2 Expression and Can Activate Sapienate Biosynthesis. **Cell Rep.** 31, 107806.

Dallmayer, M., Li, J., Ohmura, S., Alba Rubio, R., Baldauf, M.C., Hölting, T.L.B., Musa, J., Knott, M.M.L., Stein, S., Cidre-Aranaz, F., Wehweck F.S., Romero-Pérez, L., Gerke, J.S., **Orth, M.F.**, et al. (2019). Targeting the CALCB/RAMP1 axis inhibits growth of Ewing sarcoma. **Cell Death Dis.** 10, 116.

Musa, J., Cidre-Aranaz, F., Aynaud, M.-M., **Orth, M.F.**, Knott, M.M.L., Mirabeau, O., Mazor, G., Varon, M., Hölting, T.L.B., Grossetête, S., et al. (2019). Cooperation of cancer drivers with regulatory germline variants shapes clinical outcomes. **Nat. Commun.** 10, 4128.

Orth, M.F., Gerke, J.S., Knösel, T., Altendorf-Hofmann, A., Musa, J., Alba-Rubio, R., Stein, S., Hölting, T.L.B., Cidre-Aranaz, F., Romero-Pérez, L., et al. (2019). Functional genomics identifies AMPD2 as a new prognostic marker for undifferentiated pleomorphic sarcoma. **Int. J. Cancer** 144, 859–867.

Sannino, G., Marchetto, A., Ranft, A., Jabar, S., Zacherl, C., Alba-Rubio, R., Stein, S., Wehweck, F.S., Kiran, M.M., Hölting, T.L.B., Musa, J. Romero-Pérez, L., Cidre-Aranaz, F., Knott, M.M.L., Li, J., Jürgens, H., Satre, A., Alonso, J., Silveira, W., Hardiman G., Gerke, J.S., **Orth, M.F.**, et al. (2019). Gene expression and immunohistochemical analyses identify SOX2 as major risk factor for overall survival and relapse in Ewing sarcoma patients. **EBioMedicine** 47, 156–162.

Vriens, K., Christen, S., Parik, S., Broekaert, D., Yoshinaga, K., Talebi, A., Dehairs, J., Escalona-Noguero, C., Schmieder, R., Cornfield, T., Carlton, C. Romero-Pérez, L., Rossi, M., Rinaldi, G., **Orth, M.F.**, et al. (2019). Evidence for an alternative fatty acid desaturation pathway increasing cancer plasticity. **Nature** 566, 403–406.

Baldauf, M.C., Gerke, J.S., Kirschner, A., Blaeschke, F., Effenberger, M., Schober, K., Rubio, R.A., Kanaseki, T., Kiran, M.M., Dallmayer, M., et al. (2018). Systematic identification of cancer-specific MHC-binding peptides with RAVEN. **Oncoimmunology** 7, e1481558.

Baldauf, M.C., Gerke, J.S., **Orth, M.F.**, Dallmayer, M., Baumhoer, D., de Alava, E., Hartmann, W., Kirchner, T., and Grünewald, T.G.P. (2018). Are EWSR1-NFATc2-positive sarcomas really Ewing sarcomas? **Mod. Pathol.** Off. J. U. S. Can. Acad. Pathol. Inc 31, 997–999.

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

Boeva, V., Louis-Brennetot, C., Peltier, A., Durand, S., Pierre-Eugène, C., Raynal, V., Etchevers, H.C., Thomas, S., Lermine, A., Daudigeos-Dubus, E., Geoerger, B., **Orth, M.F.**, et al. (2017). Heterogeneity of neuroblastoma cell identity defined by transcriptional circuitries. **Nat. Genet.** 49, 1408–1413.

Elia, I., Broekaert, D., Christen, S., Boon, R., Radaelli, E., **Orth, M.F.**, Verfaillie, C., Grünewald, T.G.P., and Fendt, S.-M. (2017). Proline metabolism supports metastasis formation and could be inhibited to selectively target metastasizing cancer cells. **Nat. Commun.** 8, 15267.

Sannino, G.*, **Orth, M.F.***, and Grünewald, T.G.* (2017; EDITORIAL). Next steps in Ewing sarcoma (epi-)genomics. **Future Oncol.** Lond. Engl. 13, 1207–1211.

Christen, S., Lorendeau, D., Schmieder, R., Broekaert, D., Metzger, K., Veys, K., Elia, I., Buescher, J.M., **Orth, M.F.**, Davidson, S.M., et al. (2016). Breast Cancer-Derived Lung

Metastases Show Increased Pyruvate Carboxylase-Dependent Anaplerosis. **Cell Rep.** 17, 837–848.

Endres, M., Kneitz, S., **Orth, M.F.**, Perera, R.K., Zernecke, A., and Butt, E. (2016). Regulation of matrix metalloproteinases (MMPs) expression and secretion in MDA-MB-231 breast cancer cells by LIM and SH3 protein 1 (LASP1). **Oncotarget** 7, 64244–64259.

Musa, J., **Orth, M.F.**, Dallmayer, M., Baldauf, M., Pardo, C., Rotblat, B., Kirchner, T., Leprivier, G., and Grünewald, T.G.P. (2016; REVIEW). Eukaryotic initiation factor 4E-binding protein 1 (4E-BP1): a master regulator of mRNA translation involved in tumorigenesis. **Oncogene** 35, 4675–4688.

Orth, M.F., Cazes, A., Butt, E., and Grunewald, T.G.P. (2015; REVIEW). An update on the LIM and SH3 domain protein 1 (LASP1): a versatile structural, signaling, and biomarker protein. **Oncotarget** 6, 26–42.

Hailer, A., Grunewald, T.G.P., **Orth, M.**, Reiss, C., Kneitz, B., Spahn, M., and Butt, E. (2014). Loss of tumor suppressor mir-203 mediates overexpression of LIM and SH3 Protein 1 (LASP1) in high-risk prostate cancer thereby increasing cell proliferation and migration. **Oncotarget** 5, 4144–4153.

* *equal contribution*

10.2. Conferences

- DGP annual meeting 2016, 19.-21.5.2016, Berlin - Germany: Poster presentation
- Molecular Biology of Cancer Retreat – Civita 2016, 05.-07.10.2016, Civita – Italy: Short talk
- DKTK Retreat 2016, 11.-12.10.2016, Heidelberg – Germany: Poster presentation
- Mildred Scheel Cancer Conference 2017, 14.-16.06.2017, Bonn – Germany: Poster presentation
- Crete Conference “Genes, Cell Biology, Cancer and beyond”, 11.-13.10.2017, Heraklion – Greece: Talk and short talk
- Childhood Cancer Research Initiatives conference 2018, 19.-20.10.2018, Vienna – Austria: Poster presentation

10.3. Scientific experience

- Science academy at the Life Science Lab, German Cancer Research Center (DKFZ), Heidelberg, Germany
- Internships during collateral degree course Experimental Medicine

- 10/2015-04/2016 master thesis in Experimental Medicine, University of Würzburg and LMU Munich, Germany; Master of Science degree *with distinction, 1.1*
- 09/2011 – 04/2018 experimental doctoral thesis at the Institute of Clinical Biochemistry and Pathobiochemistry, University Hospital Würzburg, Germany. Medical doctor degree (Dr. med.) with *magna cum laude, 1.0*
- Since 07/2016 scientific work in the Max-Eder Research Group for Pediatric Sarcoma Biology, Institute of Pathology, LMU Munich, Germany, with international collaborations and internships