Université Paris Diderot VII |Université Sorbonne Paris Cité
École doctorale : « Hématologie, Oncogenèse et Biothérapies » - ED n°561
Institut Curie – INSERM U830
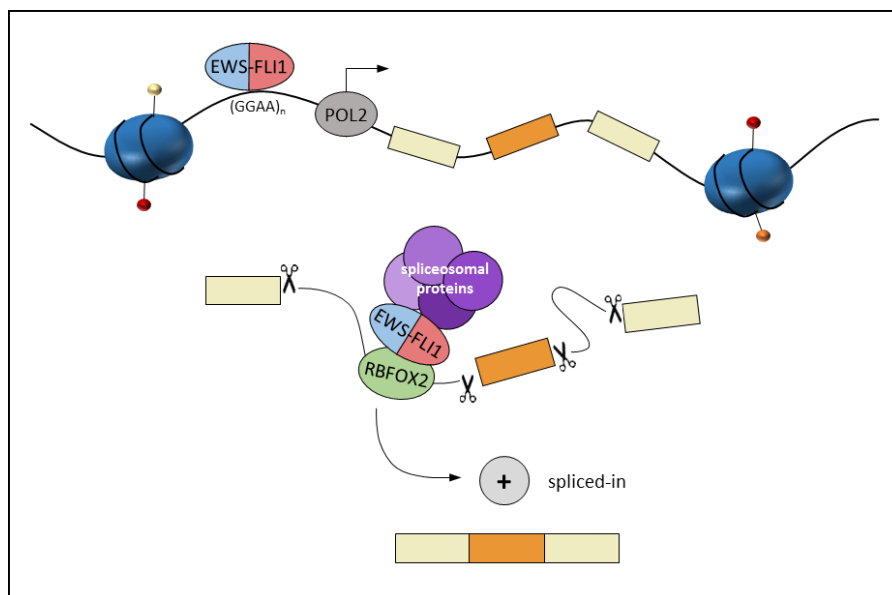
DOCTORAT

Discipline : Médecine
Spécialité : Oncogenèse

Thèse présentée et soutenue publiquement par

Olivier SAULNIER

Le 26 novembre 2018

Deciphering the splicing landscape of Ewing sarcoma

Thèse co-dirigée par le Dr. Olivier DELATTRE et le Dr. Martin DUTERTRE

JURY

Pr. Jean SOULIER - UMR 944/7212, Paris	Président
Dr. Reini FERNANDEZ de LUCO - UPR 1142, Montpellier	Rapporteur
Pr. Heinrich KOVAR - Children's Cancer Research Institute, Vienna	Rapporteur
Dr. Françoise REDINI - UMR 1238, Nantes	Examinateur
Dr. Eric LETOUZE - UMR 1162, Paris	Examinateur
Pr. Franck DEQUIEDT - Laboratory of Gene Expression in Cancer, Liège	Examinateur
Dr. Olivier DELATTRE - U830, Paris	Co-directeur de Thèse
Dr. Martin DUTERTRE - UMR3348, Orsay	Co-directeur de Thèse

Abstract

Cancer can be characterized by abnormal fusion transcription factors. These transcription factors may have gain of function, neomorphic DNA binding properties or aberrant transcriptional activity. This is the case for Ewing sarcoma, which is characterized by a chromosomal translocation EWSR1-ETS. Ewing sarcoma fusion oncoproteins have been mostly studied as aberrant transcription factors due to their ability to specifically bind GGAA repeat sequences and to activate *de novo* enhancers. In addition, its ability to recruit chromatin-remodeling proteins, to induce chromatin opening and to drive an aberrant transcriptional program is a neomorphic property of the EWSR1 moiety that depends on its low complexity domain. EWSR1-ETS fusions have recently been implicated in alternative splicing regulation but to date this function is mainly attributed to the EWSR1 part. However, ERG protein, a member of the ETS transcription factor family, has been lately shown to control post-transcriptional processes such as mRNA stability. Considering these observations, we decided to challenge this view by studying ERG as a *bona fide* splicing regulator.

This work highlights a new function of ERG subfamily proteins (ERG, FLI1 and FEV) in alternative splicing regulation. We have shown that ERG proteins interact with the master splicing regulator RBFOX2 to similarly regulate a common splicing program. We demonstrated that this new function is mediated *via* protein-protein interaction through the C-terminal domain of ERG. Because this domain remains in EWSR1-ETS fusions, we demonstrated that EWSR1-FLI1 protein is still able to bind RBFOX2 as expected. In addition, EWSR1-FLI1 induces massive changes of the splicing landscape of Ewing sarcoma and regulates an RBFOX2-dependent splicing program. However, in contrast to the collaborative effect observed for ERG, we found that EWSR1-FLI1 antagonizes RBFOX2-splicing function by repressing RBFOX2 binding to its pre-mRNAs targets. Importantly, we have found that mis-splicing of ADD3 by EWS-FLI1 leads to the repression of the mesenchymal phenotype of Ewing sarcoma cells. Our study provides direct evidence to understand how splicing dysregulation by an oncogenic transcription factor impacts on Ewing sarcoma biology.

Keywords: ERG – alternative splicing – RBFOX2 – EWS-FLI1 – Ewing sarcoma

Résumé

Certains cancers peuvent être caractérisés par un facteur de transcription aberrant. C'est le cas du sarcome d'Ewing qui est caractérisé par une translocation chromosomique générant une protéine de fusion appelée EWSR1-ETS. Ces protéines de fusion ont principalement été étudiées en tant que facteur de transcription car elles ont la capacité de se fixer sur des séquences de type répétition de GGAA dans le génome et d'activer la transcription de nombreux gènes. Les fusions EWSR1-ETS ont aussi la capacité de recruter les protéines du complexe du remodelage de la chromatine afin d'augmenter l'accessibilité aux régions riches en répétition de GGAA et donc promouvoir un programme transcriptionnel aberrant. Cette propriété dépend majoritairement de la partie EWSR1 et de son domaine de faible complexité qui lui permet, notamment, d'interagir avec de nombreuses protéines. Les fusions EWSR1-ETS ont aussi été impliquées dans la régulation de l'épissage alternatif ; mais à ce jour cette fonction reste peu décrite et est principalement attribuée à la partie EWSR1. Cependant, il a récemment été montré que la protéine ERG (qui est très homologue à FLI1) contrôle la stabilité des ARN messagers. Ces observations nous ont mené à tester le rôle potentiel de ERG (et par conséquent de FLI1) dans l'épissage alternatif afin de mieux décrire les mécanismes impliqués dans la régulation de l'épissage alternatif induite par les protéines de fusion EWSR1-ETS dans le sarcome d'Ewing.

Ce travail a permis d'identifier une nouvelle fonction des protéines de la sous-famille ERG (ERG, FLI1 et FEV) dans la régulation de l'épissage alternatif. Nous avons montré que les protéines ERG interagissent avec RBFOX2, un régulateur de l'épissage et que ERG et RBFOX2 induisent une régulation de l'épissage similaire suggérant, ainsi, un mécanisme de collaboration. Nos résultats démontrent que ERG interagit avec RBFOX2 par son extrémité C-terminale. De manière intéressante, ce domaine est retenu dans les fusions EWSR1-ETS. Nous avons donc confirmé que les fusions EWSR1-ETS étaient aussi capable d'interagir avec RBFOX2 et d'induire un programme d'épissage alternatif commun. Cependant, au contraire de la collaboration observée pour ERG et RBFOX2, nous avons montré que les fusions EWSR1-FLI1 ont un rôle opposé sur le programme d'épissage de RBFOX2. Nous avons également montré que EWS-FLI1 induisait l'épissage alternative du gène ADD3 ce qui a pour conséquence la répression du phénotype mésenchymateux des cellules du sarcome d'Ewing. Notre travail a permis d'identifier de nouveaux mécanismes afin de mieux comprendre comment la dérégulation de l'épissage alternatif par des facteurs de transcriptions oncogéniques influent sur la biologie du sarcome d'Ewing.

Mot clés : ERG – épissage alternatif – RBFOX2 – EWS-FLI1 – sarcome d'Ewing

Au sein de l'équipe du

Dr. Olivier DELATTRE

Génétique et Biologie des tumeurs pédiatriques

Unité INSERM U830

Institut Curie – Centre de Recherche

26 rue d'Ulm, 75005 PARIS

Acknowledgements

To my father and grandfather

The more I learn,
the more I realize
how much I don't know
                        - Albert Einstein

Table of contents

Index of Figures

Index of Tables

Preamble

Olivier Delattre's lab, called "Genetics and biology of childhood cancers" aims to better understand and describe molecular mechanisms at the origin of these tumors. A wide-range of biological questions are under investigation in the lab including chromatin conformation, tumor heterogeneity, mechanisms of dissemination and development of new therapeutic strategies. Ewing sarcoma is a pediatric tumor originating from bones and soft tissues and characterized by a chromosomal translocation EWSR1-ETS. In 85% of cases, the balanced translocation t(11;22)(q24;q12) leads to the chimeric fusion protein EWSR1-FLI1 that acts as an aberrant transcription factor. Neomorphic properties of EWSR1-FLI1 require the low complexity region of EWSR1 and the DNA-binding domain of FLI1 to hijack chromatin machinery, hence governing DNA accessibility and driving aberrant transcriptional programs. In addition to its transcriptional regulatory activty, EWSR1-FLI1 plays a role in alternative splicing regulation *via* interaction with core components of the spliceosome complex or by modulation of the RNA polymerase II elongation rate. However, the functional consequences and the underlying regulatory mechanisms are poorly understood.

The main project of my PhD was to identify, characterize and study the biological functions of alternative splicing events induced by EWSR1-FLI1 fusion protein. To do so, I have combined bioinformatics approaches as well as wet lab techniques to decipher the role of EWSR1-FLI1 on alternative splicing. I have used several RNA seq dataset from the lab, as well as previously published data from the literature, to identify EWSR1-FLI1-dependent splicing events. Curation of the current literature on splicing analysis tools was necessary to identify robust, reliable and convenient software for our purpose. Combined bioinformatics analysis, *in silico* prediction and *in vitro* validation allowed us to better characterize the splicing landscape of Ewing sarcoma. In addition, I have used siRNAs transfections and CRISPR technology to specifically decipher the functional role of a mis-splicing induced by EWSR1-FLI1 on Ewing sarcoma biology.

Moreover, in 2016, I had the opportunity to meet Dr. Franck Dequiedt (head of the "Protein Signaling and Interactions" lab, GIGA institute, Belgium) who is working on the role of ETS transcription factor in post-transcriptional processes. His lab has recently demonstrated that ERG transcription factor is associated to nascent RNA by interacting with RNA-binding proteins

(RBPs) to control mRNA stability. Considering these observations, we hypothesized that FLI1, which is highly homologous to ERG, might also play a role in splicing and could be important for EWSR1-FLI1 splicing function. Thus, we started a collaboration between both groups to understand mechanisms underlying EWSR1-ETS splicing function and the functional impact of alternative splicing regulation on Ewing sarcoma biology.

INTRODUCTION

I.   Ewing sarcoma

A. Clinical characteristics

1.   Historic context and localization

Ewing sarcoma was first described by James Ewing in 1921 as a new bone tumor entity called "diffuse endothelioma of bone" (Ewing, 1921). His first report was about seven young patients from fourteen to nineteen years old with tumors localized in bones and microscopically identical. Ewing sarcoma is the second most common malignant bone tumor in the pediatric population. This tumor mainly occurs in bones, such as pelvis, femur or tibia and in some rare cases, it can also arise from soft tissues **(Figure 1)**.
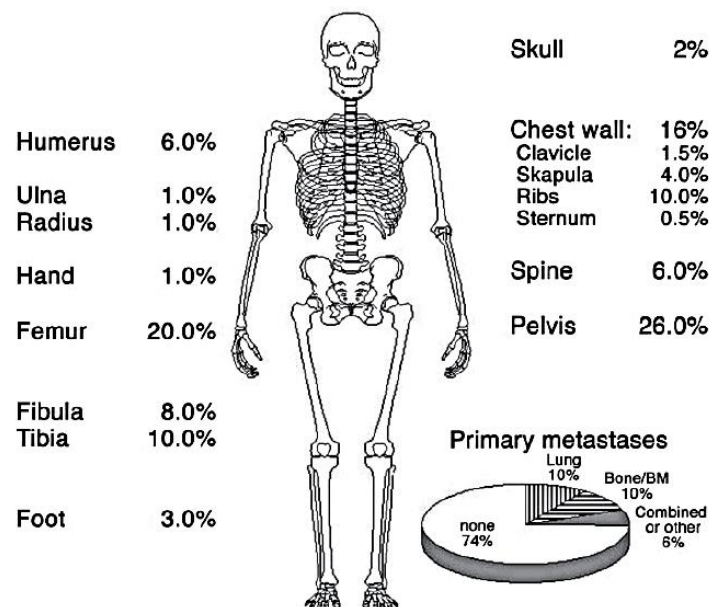
Figure 1: Most frequent primary tumor and metastatic sites in skeletal Ewing sarcoma. From Bernstein et al., 2006

2. Histology

The actual World Health Organization Classification (WHO) classification of soft tissue and bone tumors regroups Ewing sarcoma, primitive neuroectodermal tumor (pNET) and Askin tumors as one uniform tumor entity expressing FET-ETS fusion genes (de Alava et al., 2013; Doyle, 2014). Histological features include undifferentiated small round cells with scanty cytoplasm containing glycogen deposits. Tumor cells highly express the transmembrane glycoprotein CD99, which is used in routine for diagnosis **(Figure 2)** (Ambros et al., 1991). However, CD99 expression is not fully restricted to Ewing sarcoma and can also be present in other round cell sarcomas or in leukemia (Prakash et al., 2008). Ewing sarcoma are characterized by a chromosomal translocation between FET family of RNA-binding proteins and ETS transcription factor family, thus producing a chimeric transcription factor with neomorphic properties (Delattre et al., 1992). Tumors with uncommon features can be analyzed using fluorescence *in situ* hybridization (FISH) or reverse transcriptase polymerase chain reaction (RT-PCR) methods for the search of FET-ETS fusions. The search for an Ewing-specific biomarker in clinics is still relevant (Baldauf et al., 2018).
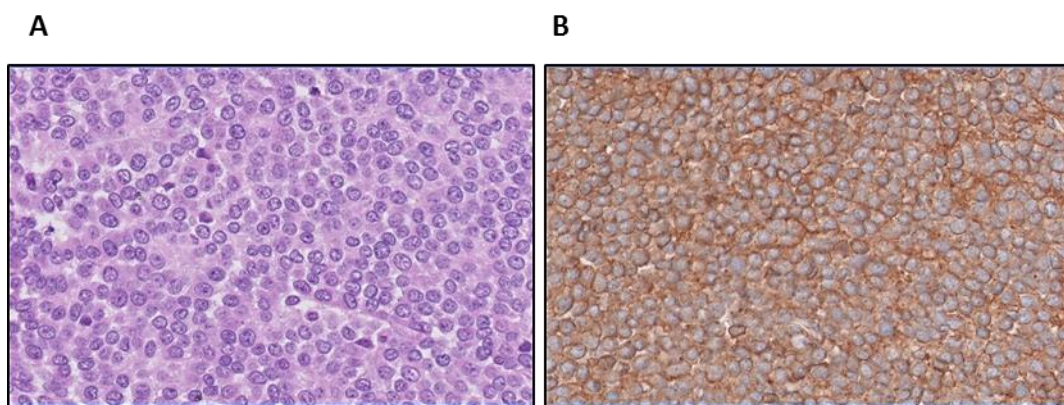
Figure 2: Histological and immunohistochemical features of Ewing sarcoma. **(A)** Classic Hematoxylin and Eosin staining of an Ewing sarcoma tumor showing small round blue cells with a minimal cytoplasm. **(B)** Tumor cells show a strong immune-reactivity for membranous CD99 protein. From Nadège Gruel, unpublished data.

3. Epidemiology

Ewing sarcoma has an incidence of 1.5 cases per million per year and this number has remained unchanged for decades (Jawad et al., 2009). It mainly affects children and young adults with a median age at diagnosis of 15 years. Men are slightly more affected than women with a sex ratio of 1.5:1 (Jawad et al., 2009). Although some rare familial cases exist (Randall et al., 2010), this disease is not associated with strong hereditary predisposition. To date, no environmental exposures have been linked to Ewing sarcoma oncogenesis. Nevertheless, Ewing sarcoma is much more common in Caucasians and is nearly absent in the African population (Jawad et al., 2009). Genome-wide association studies (GWAS) identified six loci (*EGR2*, *ADO*, *TARDBP*, *RRE1*, *KIZ* and *NKX2-2*) associated with Ewing sarcoma predisposition (Machiela et al., 2018; Postel-Vinay et al., 2012). Furthermore, risk haplotypes in these loci were less prevalent in the African population, which partially explain the geographic bias observed in this tumor type.

4. Prognosis factors and treatment

Ewing sarcoma is an aggressive tumor; around 25% of patients already show metastasis at diagnosis. This factor is the main prognosis factor and is associated with very low survival rate **(Figure 3)**. Therefore, only 20% of patients with metastasis at diagnosis survive after 5 years compared to 70% of overall survival for patients with localized tumors (Gaspar et al., 2015; Paulussen et al., 1998; Spraker et al., 2012).
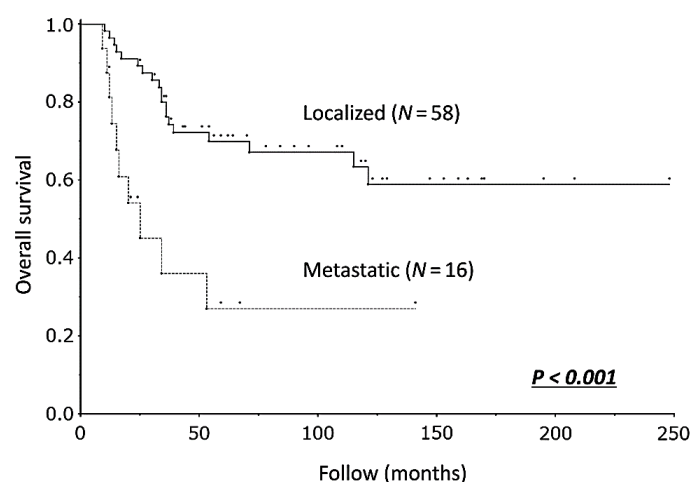
Figure 3: Overall survival of Ewing sarcoma tumors. From Takenaka et al., 2016.

Ewing sarcoma treatment includes neoadjuvant chemotherapy, surgery and/or radiotherapy. Chemotherapy is given before surgery to reduce the global tumor size and to target potential micrometastases. In Europe, standard chemotherapy includes six cycles of combination of four agents: vincristine, ifosfamide, doxorubicine, etoposide (VIDE protocol – Euro-EWING 99) (Juergens et al., 2006; Ladenstein et al., 2010). Surgery consists in the resection of the remaining tumor mass without affecting normal adjacent tissues as far as possible. In some cases, there is no tissue reconstruction necessary or feasible (*e.g.* tumors in the radius or fibula). Radiotherapy can be applied before or after surgery. Patients treated with both surgery and radiotherapy have a lower risk to develop local recurrence compared to patients treated with surgery only. However, there is no statistical difference observed in overall survival between these two subgroups (Foulon et al., 2016).

Despite development of pharmacological inhibitors for kinase fusion genes such as *EML4-ALK* in lung cancer (Kwak et al., 2010), targeting FET-ETS fusions in Ewing sarcoma remains complex, in particular due to the lack of enzymatic activity. Nevertheless, several groups focused on downstream FET-ETS targets to induce death of tumor cells. To date, there are few targeted therapies that have been pushed in clinical trials but the success of these studies is moderate. For example, the insulin-like growth factor 1 receptor (*IGF1R*) has been considered as a potential therapeutic target in Ewing sarcoma since decades. Several studies have shown that IGF1R blockade reduces tumor growth, hence suggesting an important role in Ewing sarcoma initiation (Manara et al., 2007; Scotlandi et al., 1998; Toretsky et al., 1997; Yee et al., 1990). However, clinical studies have demonstrated that only a small subset of patients (from 8% to 15%) benefit from this therapy (O'Neill et al., 2013). It is essential to decipher molecular mechanisms of action of this therapy to predict patients that will most likely respond to this therapy.

In addition, poly(ADP-ribose) polymerase (PARP) inhibitors are known to be efficient for tumors with BRCA1/2 deficiency. A recent report of homologous recombination defects in Ewing sarcoma (Gorthi et al., 2018) gives new insights on the sensitivity of Ewing sarcoma cells to PARP inhibitors and could be a promising therapeutic target to develop (Brenner et al., 2012; Garnett et al., 2012).

In the era of immunotherapy, Ewing sarcoma is considered as "cold tumor" or "immune desert" due to a very small fraction of tumors that exhibit immune infiltration (Grünewald et al., 2018). Additionally, Ewing sarcoma tumors do not express the immunosuppressive molecule programmed cell death 1 ligand 1 (PDL1) but can upregulate it under inflammatory stimulation (Machado et al., 2018; Spurny et al., 2018). Although still under evaluation, it is crucial to decipher the mechanisms of immune evasion in Ewing sarcoma to develop effective immune-based therapeutic strategies. Immunotherapy has the potential to improve current Ewing sarcoma management and to decrease toxicity and long-term treatment effects.

B. *"The origin of cancer: once upon a cell"* (ADELIH conference 2018)

Even though extensive studies on the cellular origin of Ewing sarcoma in the past decades has been done, this question is still under debate and remains unclear, in particular due to the neomorphic function of EWSR1-FLI1. Ewing sarcoma primarily arises from bone; hence, progenitors might come from either neural-crest-derived stem cells or bone marrow-derived mesenchymal cells.

Interestingly, Chan and colleagues reported, very recently, the first isolation of human skeletal stem cells that can differentiate into progenitors of bone, cartilage and stroma but not fat, muscle, fibroblasts or hematopoietic (Chan et al., 2018). This study might reveal new perspectives on the identification of the cell of origin of Ewing sarcoma tumors.

1. Neural-crest stem cells

It has been shown that Ewing sarcoma cells express neural-crest stem cells (NCSCs) markers including *CD57*, *ENO2* and genes of the Notch pathway (Baliko et al., 2007; Franchi et al., 2001; Wahl et al., 2010). In addition, ectopic expression of *EWSR1-FLI1* induces a neural crest-like phenotype and abrogates the existing cell differentiation program (Hu-Lieskovan et al., 2005; Teitell et al., 1999).

These experiments were performed in multiple cellular contexts suggesting that the observed phenotype upon EWSR1-FLI1 expression might be attributed to its transcriptional reprogramming function rather than intrinsic properties of the supposed cell of origin.

2. Mesenchymal stem cells

In 1970, Friendestein and colleagues identified an adherent cell population that morphologically looks like fibroblasts and is able to form colony units (Friedenstein et al., 1970). Caplan and colleagues introduced the term mesenchymal stem cells (MSCs) in 1991 to a population from the bone marrow that exhibited ability to differentiate into bone and cartilage (Caplan, 1991). MSCs are multipotent cells that have self-renewal capacity and the ability to differentiate into several cell types including osteocytes, chondrocytes, adipocytes, and myocytes (Gang et al., 2004; Pittenger et al., 1999).

Expression of *EWSR1-FLI1* in MSCs blocked their differentiation, thus revealing the impact of the fusion protein on the undifferentiated status of Ewing cells (Torchia et al., 2003). In addition, ectopic expression of *EWSR1-FLI1* in murine MSCs led to cell transformation and tumor growth with Ewing-like properties such as *CD99* expression (Castillero-Trejo et al., 2005; Riggi et al., 2005). Furthermore, comparison of transcriptomic profiles of Ewing sarcoma cells depleted for EWSR1-FLI1 converge towards MSCs and these cells are able to differentiate along the adipogenic and osteogenic lineages (Tirode et al., 2007). Overall, these studies underline the complex role of EWS-FLI1 in a putative mesenchymal cell of origin.

C. Genetics of Ewing sarcoma

1. The hallmark of Ewing sarcoma: FET-ETS fusions

The genetic hallmark of Ewing sarcoma is a balanced chromosomal translocation between a member of the RNA-binding protein family FET and a member of the ETS transcription factor family. The first translocation, t(11;22)(q24;q12), has been identified in 1983 (Aurias, 1983; Turc-Carel et al., 1983). In 85% of Ewing sarcoma, translocation results in the formation of a fusion gene between *EWSR1* and *FLI1*, encoding an RNA-binding protein and a transcription factor respectively **(Figure 4)** (Delattre et al., 1992). This primary oncogenic event leads to a fusion gene containing the 5' low complexity region of *EWSR1* and the 3' DNA binding domain of the *FLI1* gene. *EWSR1* gene and *EWSR1-FLI1* fusion gene will be referred as *EWS* and *EWS-FLI1*, respectively, for the rest of the manuscript. *EWS-FLI1* fusion gene is ubiquitously expressed because of the EWS reporter activity. However, *FLI1-EWS* fusion has rarely been found expressed in Ewing sarcoma because the FLI1 promoter is quiescent. In addition, a recent study suggested that a proportion of Ewing sarcoma translocations are generated from a complex mechanism, called "chromoplexy" (Anderson et al., 2018). This mechanism causes a sudden burst of chromosomal rearrangements in the cell resulting in disruption of the reciprocal fusion *FLI1-EWS*, which is not observed in chromoplexy-induced translocations.
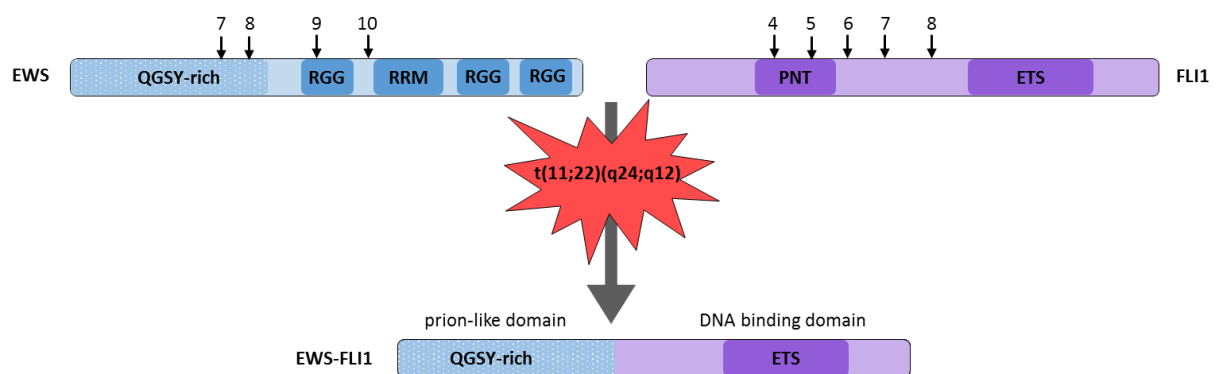
Figure 4: Schematic representation of EWSR1, FLI1 and EWS-FLI1 (type I fusion) domains. Black arrows indicate genomic breakpoints. Abbreviations: RNA recognition motif (RRM); Arg-Gly-Gly repeats (RGG); Pointed domain (PNT); ETS DNA-binding domain (ETS).

In the remaining 15% of cases that do not harbor the EWS-FLI1 translocation, numerous other fusion genes have been found, mainly implicating *EWS* with other *ETS* family genes **(Table 1)** (Jeon et al., 1995; Kaneko et al., 1996; Ng et al., 2007; Peter et al., 1997; Shing et al., 2003; Zucman et al., 1993).

Family	Translocation	Fusion gene	Frequency	References
EWS-ETS	t(11;22)(q24;q12)	EWS-FLI1	85%	Delattre et al., 1992
	t(21;22)(q22;q12)	EWS-ERG	10%	Zucman et al., 1993
	t(2;22)(q33;q12)	EWS-FEV	<1%	Peter et al., 1997
	t(7;22)(p22;q12)	EWS-ETV1	<1%	Jeon et al., 1995
	t(17;22)(q12;q12)	EWS-ETV4	<1%	Kaneko et al., 1996
FUS-ETS	t(16;21)(p11;q22)	FUS-ERG	<1%	Shing et al., 2003
	t(2;16)(q36;p11)	FUS-FEV	<1%	Ng et al., 2007

Table 1: Chromosomal translocations observed in Ewing sarcoma.

In addition, more than ten distinct types of the EWS-FLI1 transcripts exist. The two most frequent fusion transcripts are fusions of EWS exon 7 / exon 6 FLI1 (called fusion type I) and EWS exon 7 / exon 5 FLI1 (fusion type II) (Zucman et al., 1993).

2. The FET RNA-binding protein family

The FET protein family is composed of three RNA-binding proteins (RBPs) called fused in sarcoma (FUS), Ewing sarcoma breakpoint region 1 (EWSR1) and TATA-binding protein-associated factor 15 (TAF15). Structurally, FET proteins share a high sequence similarity and share an RNA-binding domain, multiple Arg-Gly-Gly boxes and a low complexity region (Tan and Manley, 2009). The FET members are ubiquitously expressed in almost all tissues and are mainly located in the nucleus. They are implicated in various processes, including transcription, post-transcriptional regulation and DNA repair (Hallier et al., 1998; Hoell et al., 2011; Wang et al., 2013; Yang et al., 1998; Zhang et al., 1998). They serve as transcriptional coregulators interacting with several transcription factors, the RNA polymerase II and the TFIID complex (Tan and Manley, 2009). Through their RNA binding domain, they have regulatory roles in multiple post-transcriptional processes, from splicing to export to translation notably by interacting with RNA-processing proteins (Meissner et al., 2003; Yang et al., 1998). They have been involved in the formation of stress granules, which are cytoplasmic RNA-protein aggregates induced by stresses (Ramaswami et al., 2013).

One important specificity of these proteins is the prion-like domain (PrLD), which is a low-complexity region rich polar, uncharged amino acids such as glutamine (Q), glycine (G), serine (S) and tyrosine (Y). This domain is similar in composition to prion and is conserved through evolution. Prions have been discovered in the 60's in mammals and are known to cause neurodegenerative disorders in humans such as Kuru or Creutzfeldt-Jacob.

Proteins displaying such domain (mainly RBPs) can undergo physiological and reversible phase transition between soluble, hydrogel and fibrous states **(Figure 5)** (Han et al., 2012; Kato et al., 2012). These amyloid-like fibrils are structurally formed by β-sheet assembly, which are insoluble and resistant to degradation (Rambaran and Serpell, 2008). Moreover, they have the ability to form membrane-free compartments, thus sequestrating substrates such as RNA into these transient organelles. For instance, fused in sarcoma (*FUS*) gene encodes for an RNA-binding protein composed of a RNA recognition motif (RRM) and a prion-like domain. FUS protein is normaly localized in the nucleus and regulates several processes mainly associated with RNA processing. However, FUS mutant proteins have been shown to mislocalize to the cytoplasm and to form pathological protein aggregates in motor neurons (Kwiatkowski et al., 2009). Mutations in *FUS* have been linked to several diseases such as Amyotrophic Lateral Sclerosis (ALS) and Frontotemporal dementia (FTD) (Alberti et al., 2009; Couthouis et al., 2011; Da Cruz and Cleveland, 2011; Harrison and Shorter, 2017; Mackenzie et al., 2010; Patel et al., 2015).

In addition, Maharana and colleagues have suggested a mechanism by which RNA concentration in the nucleus governs the ability of RBPs to mislocalize to the cytoplasm and to form pathological solid aggregates, hence sequestering RNA molecules and proteins in the cytoplasm (Maharana et al., 2018).
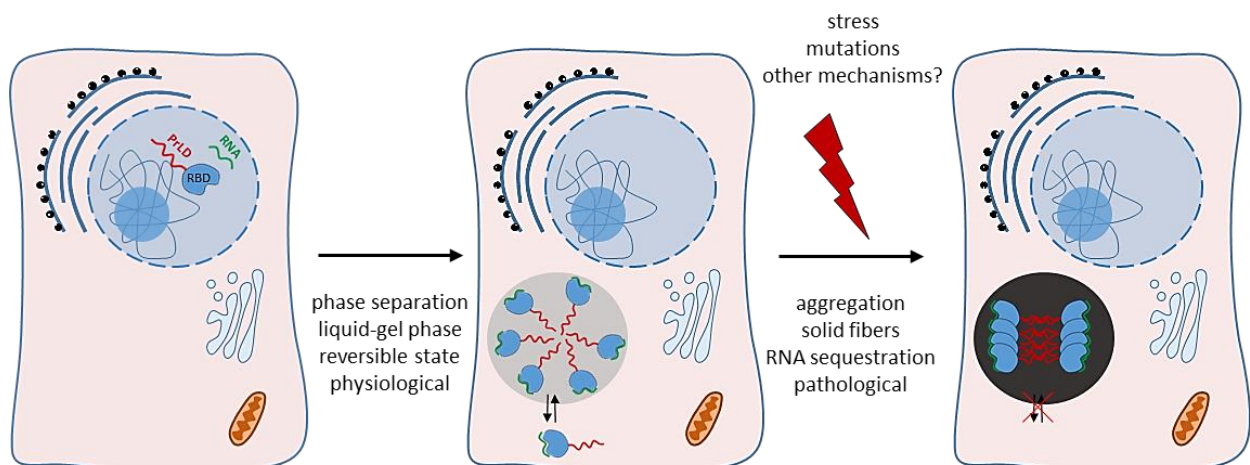
Figure 5: Schematic representation of how RBPs with prion-like domain undergo phase separation. RNA-binding proteins that arbor prion-like domain are soluble in the nucleus and go through phase separation in the cytoplasm to form membrane-less compartments. Under specific stress, mutation or other mechanisms, RBPs harbor mis-folding and undergo pathological protein aggregates leading to an important cell dysfunction. Abbreviations: prion-like domain (PrLD); RNA-binding domain (RBD). PrLD is represented in red and RNA in green.

3. The ETS transcription factor family

The E-twenty-six transformation-specific (ETS) family is one of the largest transcription factor family that includes 27 proteins subdivided in 12 subgroups. All ETS proteins share a conserved ETS DNA-binding domain. ETS transcription factors are implicated in a wide-range of biological processes and influence gene regulation in particular during embryonic development and differentiation (Oikawa and Yamada, 2003; Schober et al., 2005; Sharrocks, 2001). All ETS members bind a core purine-rich motif (GGA[A/T]), however flanking sequences or specific binding partners can affect transcriptional activity (Karim et al., 1990; Nye et al., 1992). The ERG subfamily is composed of three proteins: ETS-related gene (ERG), friend leukemia integration 1 transcription factor (FLI1) and fifth Ewing variant (FEV). This family has been largely implicated in many chromosomal translocations in cancer, such as EWS-FLI1 in Ewing sarcoma (Delattre et al., 1992) or TMPRSS2-ERG in prostate cancer (Tomlins et al., 2005). ETS transcription factors have been widely studied as oncogenic transcription activators (Sizemore et al., 2017), however only a few suggested a role of ETS proteins in post-transcriptional regulation (Guillouf et al., 2006; Rambout et al., 2016), as I will discuss below **(See Section II.D)**.

Several reports have demonstrated that ETS proteins are implicated in post-transcriptional processes such as splicing or mRNA degradation. For instance, Spi-1-/PU.1 is able to bind RNA through its DNA-binding domain and induce alternative splicing (Guillouf et al., 2006; Hallier et al., 1996, 1998). Recently, ERG has been shown to control mRNA degradation through interaction with RBPs (Rambout et al., 2016). Altogether, these observations suggest that ETS proteins are implicated in many cellular processes and should not be considered only as transcription factors.

4. Other genetic alterations

Ewing sarcoma, as most pediatric cancers, has a silent genomic profile and presents one of the lowest mutation rate (0.3/Mb) across all cancer types **(Figure 6)** (Lawrence et al., 2013; Mugneret et al., 1988)
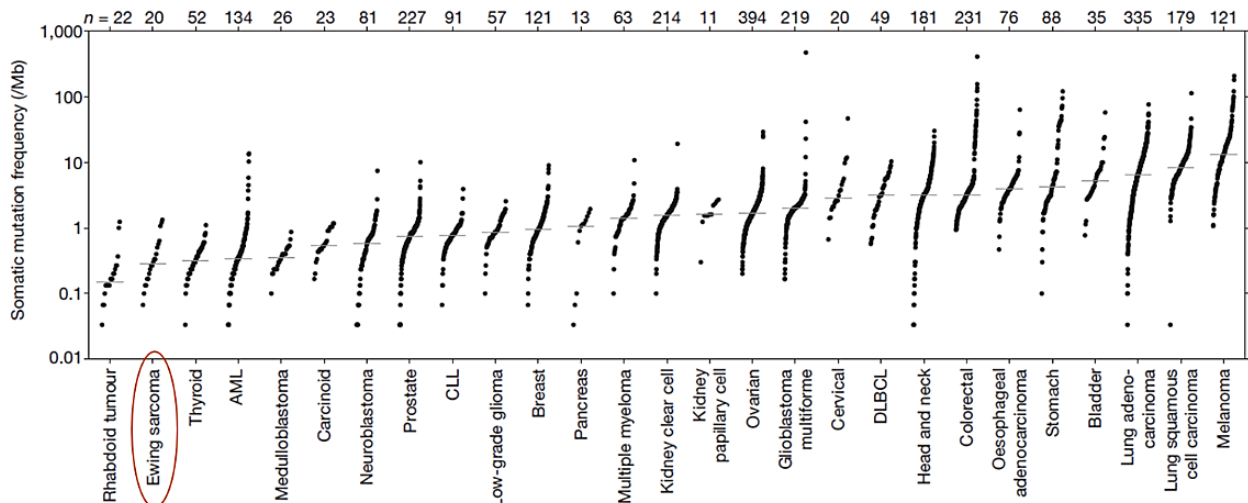
Figure 6: Somatic mutation frequency across human cancers. The lowest mutation rates (left) are found in pediatric cancers, whereas the highest frequencies (right) are found in tumors induced by carcinogens, such as ultraviolet light and tobacco. Adapted from Lawrence et al., 2014.

Recently, several groups have identified the mutational landscape of Ewing sarcoma using whole-genome sequencing (Brohl et al., 2014; Crompton et al., 2014; Tirode et al., 2014). Recurrent mutations were found in the Cohesin subunit SA-2 (*STAG2*) (20%), Cyclin Dependent Kinase Inhibitor 2A (*CDKN2A*) (13%) and Tumor Protein P53 (*TP53*) (6%) genes. Interestingly, these studies showed that prognosis of patients harboring both *STAG2* and *TP53* mutations is particularly unfavorable, suggesting a potential cooperation between these two mutations to increase tumorigenesis. Altogether, these findings underline the genetic background of Ewing sarcoma and suggest STAG2 as a potential target for therapeutic development. The development of therapies targeting chromatin remodelers and epigenetic regulators could be a promising avenue such as lysine-specific histone demethylase 1A (LSD1) inhibitors (Bennani-Baiti et al., 2012; Sankar et al., 2014).

D. Oncogenic neomorphic properties of EWS-FLI1 fusion gene

EWS-FLI1 fusion is a chimeric transcription factor with neomorphic functions that aberrantly modulates expression of thousands of genes. In particular because of its ability to bind GGAA microsatellites, thus activating *de novo* enhancers (Boulay et al., 2017). EWS-FLI1 protein have oncogenic properties and is critical for cell transformation and proliferation (May et al., 1993). However, very few cells tolerate stable expression of EWS-FLI1 and the cellular context seems to be essential to recapitulate Ewing sarcoma biology.

1. Transcription factor properties

EWS-FLI1 binds DNA through the DNA-binding domain of FLI1, which remains in the fusion. Several studies have identified, using chromatin immunoprecipitation (ChIP) data, the DNA binding motif of EWS-FLI1 **(Figure 7)** (Boeva et al., 2010; Gangwal et al., 2008; Guillon et al., 2009). Both, FLI1 and EWS-FLI1, bind the same canonical ETS binding motif composed of a GGAA site.

However, EWS-FLI1 can bind GGAA microsatellites likewise with one rule: "the more GGAA there is, the more expressed it is", with an upper limit around 20 GGAA repeats (Gangwal et al., 2008; Guillon et al., 2009; Johnson et al., 2017). This binding motif is specific of the fusion protein and is not observed for the wild type FLI1 transcription factor (or any ETS family members) highlighting the neomorphic function of EWS-FLI1.
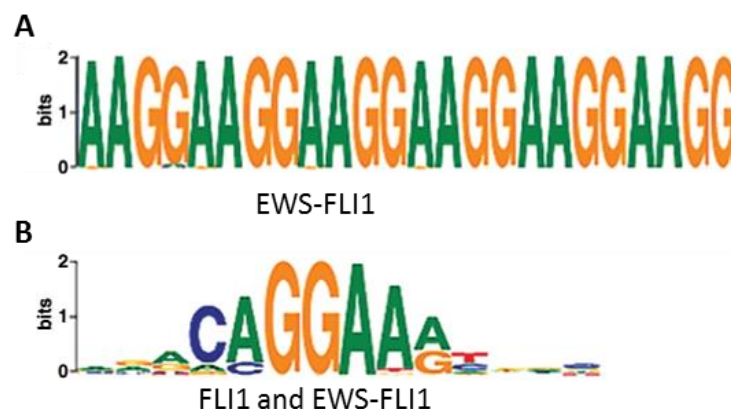
Figure 7: Enriched motifs oserved in EWS-FLI1 ChIP-seq peaks by the MICSA algorithm. **(A)** Most enriched motif found representing GGAA microsatellites and **(B)** canonical ETS binding motif. Adapted from Boeva et al., 2010.

EWS-FLI1 oncoprotein modulates a wide-range of biological processes such as cell-cycle regulation, telomerase activity, cell migration and chromatin conformation (Cidre-Aranaz and Alonso, 2015). Epigenomic studies have demonstrated that depending on the binding site, which can either be (GGAA)$_n$ or ETS motif, EWS-FLI1 fusion can have distinct transcriptional regulatory activities. Activated genes encodes for proteins mostly involved in cell-cycle process and are associated with GGAA microsatellite binding site. Otherwise, repressed genes are enriched in extracellular matrix (ECM) pathway and are associated with canonical ETS binding site (Riggi et al., 2014; Tomazou et al., 2015).

2. Inherited germline variant

Furthermore, several transcriptional targets of EWS-FLI1 have been extensively described. For instance, Early Growth Response 2 (*EGR2*) is important for Ewing sarcoma tumorigenicity and is highly expressed in Ewing tumors compared to other tumor types and normal tissues (Grünewald et al., 2015). This gene is located nearby the locus identified by GWAS as genetic variant associated with Ewing sarcoma susceptibility (Machiela et al., 2018; Postel-Vinay et al., 2012). The Single Nucleotide Polymorphism (SNP) rs79965208 is located within a long sequence of 16 GGAA repeats that is bound by EWS-FLI1 **(Figure 8)**. For this SNP, the reference allele is T, whereas A is the alternative allele. Having the T reference allele splits the sequence of GGAA repeats into two sequences of 11 and 4 repeats. The homozygous A/A genotype is associated with an overexpression of EGR2 gene compared to A/T or T/T genotypes. Additionally, the T reference allele is predominant in the African population compared to Caucasians suggesting a role of this SNP on Ewing sarcoma susceptibility and oncogenesis. Altogether, the study deciphers how inherited germline variants cooperate with EWS-FLI1 fusion protein to promote tumorigenesis.
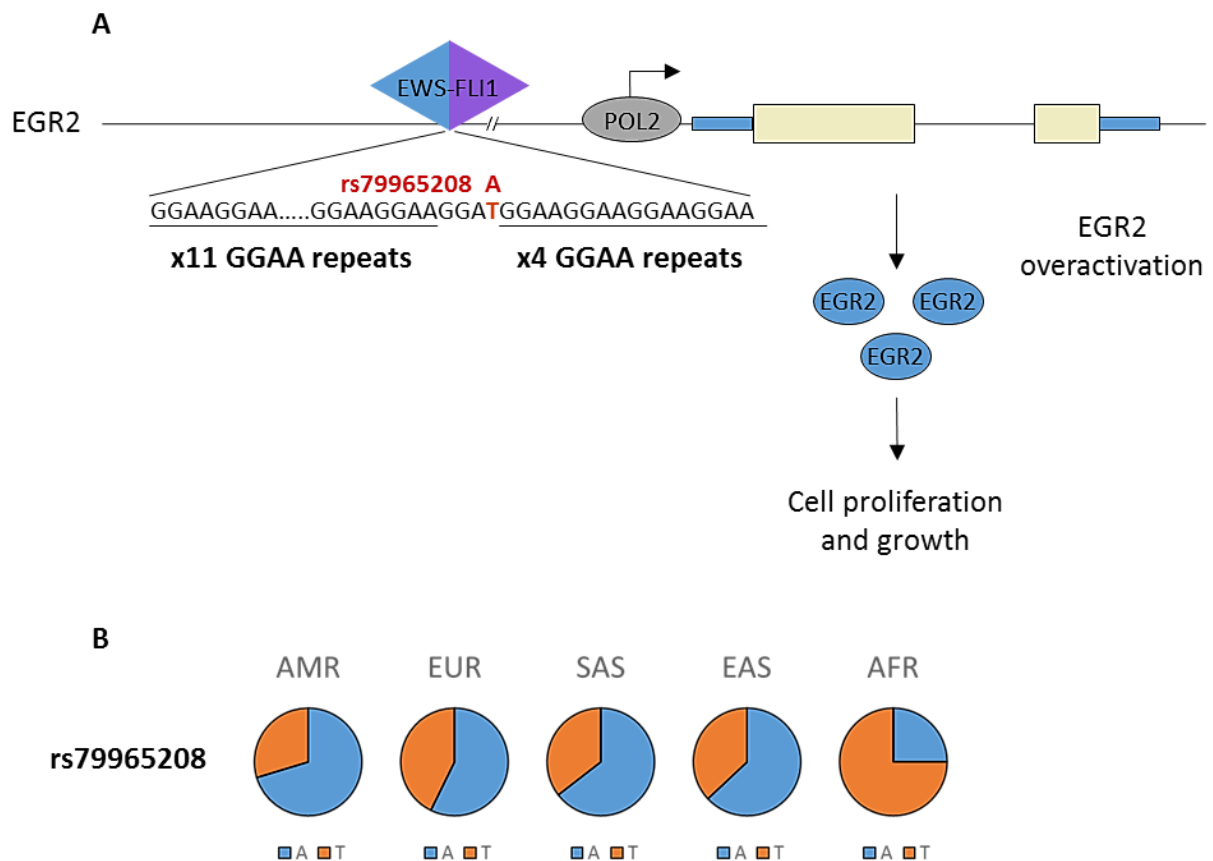
Figure 8: Schematic representation of how a genetic susceptibility factor, which is frequent in European population, interacts with EWS-FLI1 oncoprotein to drive abberant transcriptional program. **(A)** EWS-FLI1 binds a GGAA microsatellite nearby *EGR2* locus. Interestingly, a SNP converts a GGAT into a GGAA motif, therefore connecting two GGAA repeats into one long repetition of sixteen GGAA, thus enhancing *EGR2* expression. **(B)** Allele frequency distribution for rs79965208 in dbSNP (https://www.ncbi.nlm.nih.gov/projects/SNP). Abbreviations: The Americas (AMR), Europe (EUR), South Asia (SAS), East Asia (EAS), Africa (AFR).

3. EWS-FLI1 as a mediator of chromatin remodeling

It has been demonstrated that EWS-FLI1 interacts with the BRG1/BRM (BAF) chromatin-remodeling complex *via* the prion-like domain of EWS (Boulay et al., 2017). BAF complex plays a crucial role on chromatin accessibility at enhancers and promoters regions (Tolstorukov et al., 2013). EWS-FLI1 recruits BAF complex on GGAA microsatellites to promote chromatin opening and *de novo* enhancer formation (Boulay et al., 2017) **(Figure 9)**. This observation is in agreement with the fact that Ewing sarcoma cell lines are the only known cells that have open chromatin at GGAA repeat regions (Riggi et al., 2014). This study highlighted a fascinating example of how a chromatin remodeling machinery is hijacked by neomorphic properties of a gene fusion to drive aberrant transcriptional programs in cancer.
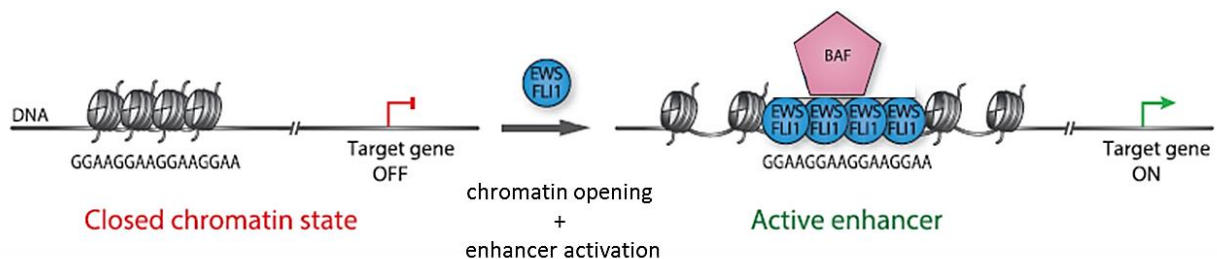
Figure 9: Mechanistic model of EWS-FLI1 binding at GGAA microsatellites and *de novo* enhancer activation in Ewing sarcoma. GGAA repeats are often associated to a closed chromatin state. In Ewing sarcoma, EWS-FLI1 recruits BAF chromatin complex to activate chromatin opening at these sites, thereby enhancing gene expression. Adapted Boulay et al., 2017.

Recently, the same group demonstrated how targeting GGAA repeats might be a promising therapeutic strategy for Ewing sarcoma (Boulay et al., 2018). They used CRISPR-dCas9-KRAB technology (Gilbert et al., 2014; Thakore et al., 2015) to induce deposition of the repressive histone mark H3K9me3 on *SOX2* GGAA repeats enhancer. Epigenetic silencing of this locus is sufficient to inhibit EWS-FLI1 binding, to abolish *SOX2* expression and to impair tumor growth *in vivo* (Boulay et al., 2018). This study showed a direct link between a unique GGAA microsatellite and EWS-FLI1 oncogenic transcriptional activity. In addition to their impact on transcription and chromatin, a few studies have suggested a role of FET-ETS fusion proteins in splicing regulation as I will discuss below **(See Section II.D)**.

4. Ewing sarcoma phenocopy BRCA1-deficient tumors

Recently, it has been shown that aberrant transcriptional regulation induced by EWS-FLI1 fusion leads to R-loops accumulation (Gorthi et al., 2018). R-Loops are three-stranded nucleic acid structures including an RNA-DNA hybrid and a single-stranded DNA. In the human genome, R-loops accounts for 5 to 8% of the genome and influence many cellular processes (Chen et al., 2015; Sanz et al., 2016). Gorthi and colleagues have demonstrated that breast cancer type 1 susceptibility (*BRCA1*) gene, which is crucial for DNA damage response, seems to be partially inactivated in Ewing sarcoma (Gorthi et al., 2018). Indeed, BRCA1 is sequestrated at R-loops sites with transcription complexes leading to its partial inactivation. This study underlines the mechanism of Ewing sarcoma sensitivity to poly(ADP-ribose) polymerase (PARP) inhibitors (Brenner et al., 2012; Garnett et al., 2012; Gorthi et al., 2018). Overall, the study highlights an aberrant transcriptional activity in Ewing sarcoma leading to R-loops formation and BRCA1 sequestration resulting in impaired homologous recombination and promising therapeutic development.

5. EMT-like, plasticity and heterogeneity

Several research groups have shown that Ewing sarcoma cells can have a certain plasticity between epithelial to a more mesenchymal phenotype (Chaturvedi et al., 2012, 2014; Franzetti et al., 2017; Katschnig et al., 2017; Wiles et al., 2013). Ewing sarcoma is quite complex to classify between mesenchymal-like tumors, because of its origin, (such as osteosarcoma or chondrosarcoma) and epithelial-like tumors, because of its morphological aspect (such as synovial or epithelioid sarcomas). For instance, tight junction-related proteins such as claudin-1 and ZO-1 are expressed in Ewing sarcoma, whereas epithelial E-cadherin marker (CDH1) is not (Schuetz et al., 2005).

To decipher this complexity, it has been recently suggested that Ewing sarcoma tumors reside in an intermediate state called "metastable" phenotype which allows tumor cells to acquire either epithelial or mesenchymal features (Sannino et al., 2017). This phenotype is associated with aggressiveness because cells exhibit features from both states and can adapt to tumor environment (Jolly et al., 2015). This plasticity observed in Ewing sarcoma cells is mostly mediated by fluctuations in *EWS-FLI1* expression (Franzetti et al., 2017). These fluctuations in the master fusion oncoprotein allow cells to transiently acquire epithelial or mesenchymal features. Drastic transcriptomic and epigenetic reprogramming occur during this process, to allow cells to switch from a highly proliferative undifferentiated cell towards spindle-shaped cell with invasive properties. These observations suggest that there is a cooperation between (i) cells displaying high expression of *EWS-FLI1*, which are proliferative and contribute to tumor development and (ii) cells that have a low *EWS-FLI1* expression, which migrate and potentiate metastatic seeding *in vivo* (Franzetti et al., 2017) **(Figure 10)**. Nonetheless, EMT is not a binary process and it has been shown that intermediate states and transition fluctuations within this process are common features (Pastushenko et al., 2018). We hypothetized that Ewing sarcoma cells can have a certain plasticity according to the expression of *EWS-FLI1*. However, stochastic, cell-autonomous or non-cell autonomous mechanisms underlying this plasticity needs to be elucidated. In addition, the tumor microenvironment could play a crucial role in this cell-plasticity and needs to be investigated in this context.
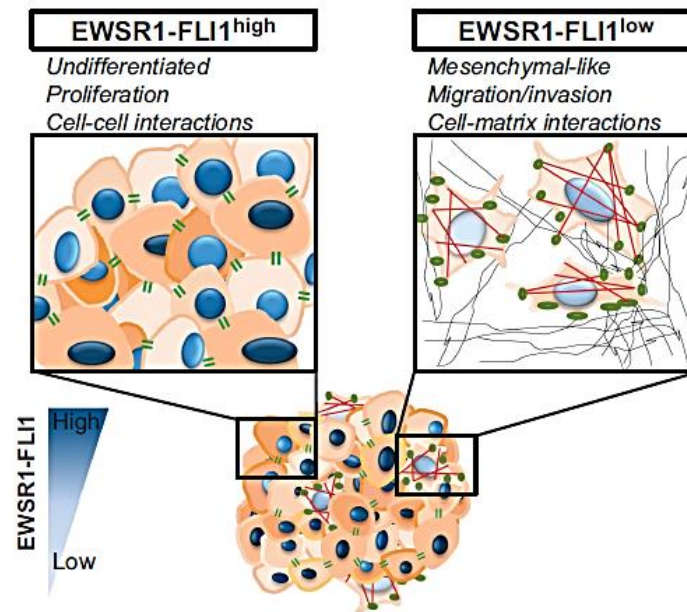
Figure 10: Schematic mechanism of Ewing sarcoma dissemination based on EWS-FLI1 fluctuation. EWS-FLI1^{High} cells have an undifferentiated phenotype, high proliferation rate and strong cell-cell interaction, whereas EWS-FLI1^{Low} cells have a mesenchymal-like phenotype with increase migratory/invasive capacities and important cell-matrix interactions. From Franzetti et al., 2017.

In addition, Katschnig and colleagues have identified a molecular mechanism implicated in the downregulation of cytoskeleton genes. Genes of the Rho pathway tightly regulate cell cytoarchitecture. They have found that EWS-FLI1 inhibits activation of the Rho pathway genes by interfering with transcription activators MRTFB/TEAD (Katschnig et al., 2017). Authors suggested that this regulation is mediated by the AP-1 transcription factor family, which interacts with both TEAD and EWS-FLI1 (Kim et al., 2006; Liu et al., 2016; Zanconato et al., 2015). This study establishes a novel mechanism of cell morphology modulation by EWS-FLI1. In addition, analysis of these protein complexes could be central to understand the repressive transcriptional activity of EWS-FLI1 and to develop better strategies for metastasis treatment.

Moreover, Pedersen and colleagues have identified the Wnt/beta-catenin axis as an important pathway for Ewing sarcoma tumorigenicity (Pedersen et al., 2016). Despite low frequency of Wnt/beta-catenin mutations in Ewing tumors, they have found that Wnt pathway seems to be activated at diverse level and Wnt-activated tumors are more clinically aggressive. This heterogeneity could partially be mediated by *LRG5* expression, which is restricted to a subset of Ewing tumors. Furthermore, Wnt activation leads to overexpression of metastasis-associated and pro-migratory genes, which are normally repressed by EWS-FLI1 (Pedersen et al., 2016).

According to this non-genetic heterogeneity, Sheffield and colleagues have recently shown that Ewing sarcoma DNA methylation profiles also exhibit heterogeneity. Indeed, clustering DNA methylation profiles of 140 Ewing tumors highlighted an Ewing-specific hypomethylation signature, which separated Ewing tumors from other cancer types (Sheffield et al., 2017). This is to put in relation with EWS-FLI1's role on epigenomic reprogramming as we have seen previously. Moreover, Ewing tumors exhibited a high interindividual heterogeneity with a continuous spectrum of DNA methylation profiles. It would be interesting to correlate *EWS-FLI1* expression to DNA methylation map to decipher epigenetic heterogeneity in Ewing sarcoma. Despite a silent genetic background, Ewing sarcoma shows a high intertumor epigenetic heterogeneity and is defined as a continuous disease spectrum.

II.   Alternative splicing

A. The basics of splicing and alternative splicing

1.   Historic context

In the 70's, Phillip Allen Sharp and Richard John Roberts identified a messenger RNA (mRNA), produced by an adenovirus, originates from four non-contiguous DNA segments (Berget et al., 1977; Chow et al., 1977). One year later, Walter Gilbert suggested to call these expressed sequences "exons", which are separated by intragenic regions called "introns" (Gilbert, 1978). For this major breakthrough, the Nobel Prize in Physiology or Medicine 1993 was attributed to P. Sharp and R.J. Roberts. In the early 80's, alternative splicing was revealed with the finding of distinct alternative transcripts of the immunoglobulin heavy locus (*IGH*) gene emerging from the same DNA region (Early et al., 1980). As detailed below, transcript splicing is a reaction consisting in the removal of introns and the ligation of exons. We now know that most mammalian coding genes, as well as many long non-coding genes have introns, and that alternative splicing occurs in about 95% of genes (Wang et al., 2008).

2.   Biological function

Alternative Splicing (AS) is a biological process by which one single DNA sequence can generate structurally and functionally different mRNAs molecules. This process is highly prevalent in higher eukaryotes because it contributes to transcriptome (and proteome) diversity and complexity (Barbosa-Morais et al., 2012; Blencowe, 2006; Chen et al., 2014). Alternative splicing is widely regulated in a tissue-, cell type- and developmental-stage specific manner and plays an important role in various cell differentiation programs (Daguenet et al., 2015; Kalsotra and Cooper, 2011). In addition, alternative splicing tunes important cellular processes such as apoptosis (Schwerk and Schulze-Osthoff, 2005; Wu et al., 2003) and has been involved in all the cellular hallmarks of cancer (David and Manley, 2010; Oltean and Bates, 2014), as I will discuss below. At the molecular level, alternative splicing not only affects the protein encoded itself but can also reshape protein interaction networks **(Figure 11)** (Buljan et al., 2012).
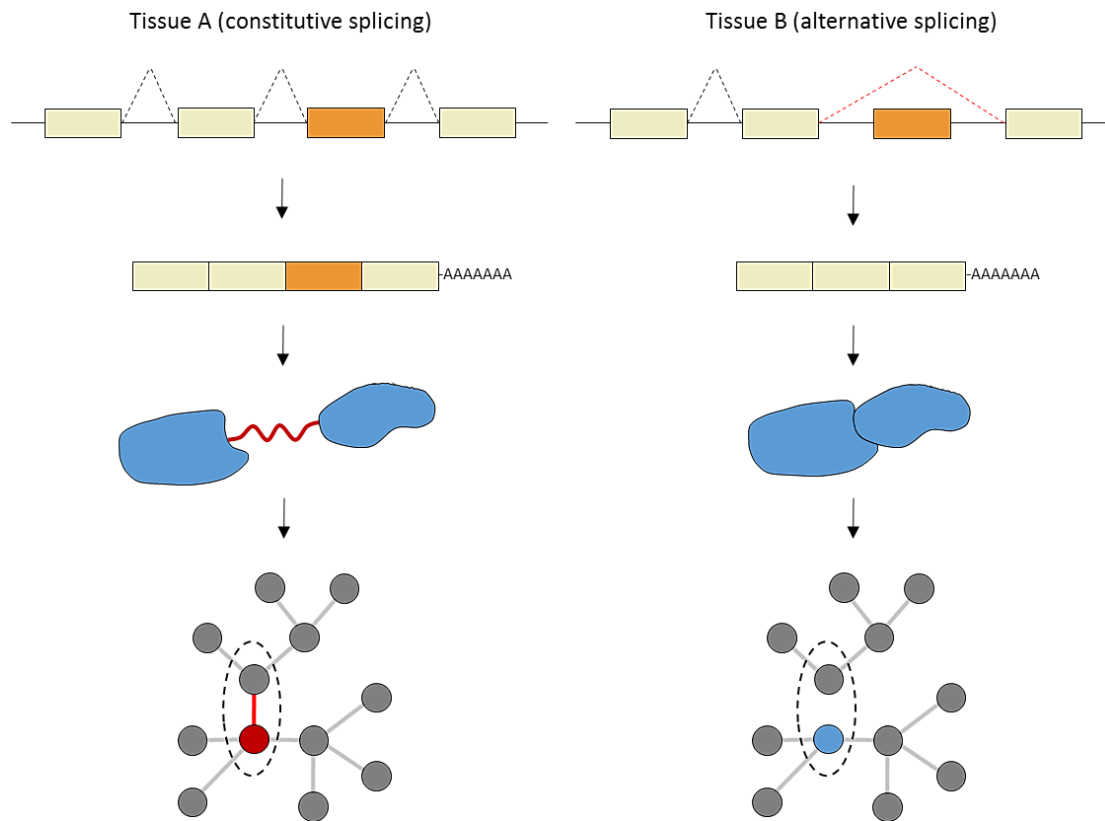
Figure 11: Alternative splicing modify protein interaction networks. Alternative splicing can give rise to different isoforms composition, such as the skipping of an interacting protein segment, hence modiying its binding ability. Adapted from Buljan et al 2012.

3. Splicing in numbers

In the human genome, there are 9 exons per gene on average. Mean exon length is about 170bp compared to 5400bp for introns (more than 5% of introns are greater than 200 kbp!) (Sakharkar et al., 2004). These numbers highlight the challenge to coordinate the spliceosome assembly and to operate the splicing reaction. The number of distinct transcripts produced by human coding genes has been estimated to 200,000 (Harrow et al., 2012; Pertea, 2012). This represent more than six transcripts per gene even if the majority of protein coding genes have only one dominant transcript per gene in a given cell type and biological condition (Ezkurdia et al., 2015; Gonzàlez-Porta et al., 2013).

Over the past decade, the number of transcripts has kept increasing whereas the number of protein coding genes has decreased to around 22,000. Advances in next generation sequencing (NGS) have largely contributed to this phenomenon. In particular, third generation sequencing technologies, such as Pacbio and Nanopore, which allow sequencing of full-length mRNA molecules, overrode most algorithmic constraints on transcriptome reconstruction. In addition to the coding repertoire, many long non-coding RNAs (lncRNAs) are also alternatively spliced and are themselves implicated in the splicing regulation of coding proteins genes (Mattick and Makunin, 2006; Rinn and Chang, 2012).

4. Different types of alternative splicing events

Due to its complexity, alternative splicing can give rise to a large diversity of mRNAs **(figure 12)**. There are at least six types of alternative splicing events: (i) skipped exon (SE) is the most common (and most commonly studied) in mammals: one or multiple exons are spliced-in or spliced-out, (ii) mutually exclusive exons (MXE): both exons cannot belong to the same transcript. If exon A is included, exon B is skipped and conversely, (iii) alternative 5' splice site (A5SS), (iv) alternative 3' splice site (A3SS), (v) intron retention (IR), which corresponds to the retention of a whole intron in the mature transcript. Most of the time, intron retention leads to a premature stop codon and degradation through nonsense-mediated decay (NMD) and (vi) alternative last exon (ALE), also called intronic polyadenylaton or splicing-dependent alternative polyadenylation (APA). It should be noted that in addition to alternative splicing, a large part of transcriptome diversity in Human corresponds to splicing-independent APA (APA within the last exon of genes) and to alternative first exons (AFE), the latter of which is due to alternative promoters and is thus related to transcriptional events (Pal et al., 2011; Reyes and Huber, 2018).
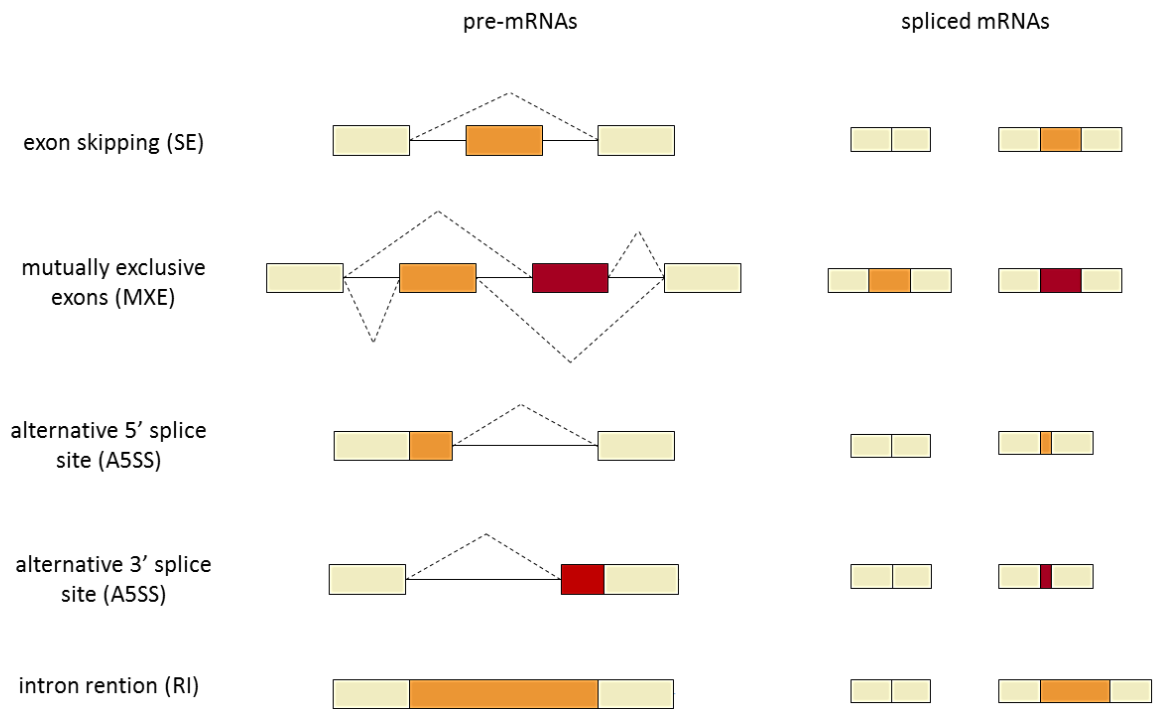
Figure 12: Five major events of alternative splicing. ALEs, which are less studied, are not represented..

B. Multiple layers of regulation

1. The core splicing signals

Intronic regions are extremely variable among Eukaryotes, but sequences located at the extremity of introns are highly conserved and are necessary for the spliceosome recognition. There are four sequences that are essential for the splicing reaction **(Figure 13)** (Breathnach and Chambon, 1981; Breathnach et al., 1978; Reed and Maniatis, 1985). The 5' donor site, located at the exon|intron boundary, is composed of the consensus sequence CAG|**GT**RAGT. The most conserved sequence is the dinucleotide GT (Moore and Sharp, 1993). The 3' acceptor site, located at the boundary intron|exon, is constituted of the sequence Y**AG**|G. The branchpoint is more degenerated than other splice signals. The sequence YNYTR**A**Y is located between 18 to 40 bases upstream the 3' acceptor site. In addition to these sequences, intronic regions harbor a pyrimidine-rich sequence called the polypyrimidine tract (PPT). This region is located between the branchpoint and the 3' acceptor site and is generally 10-20nt long.
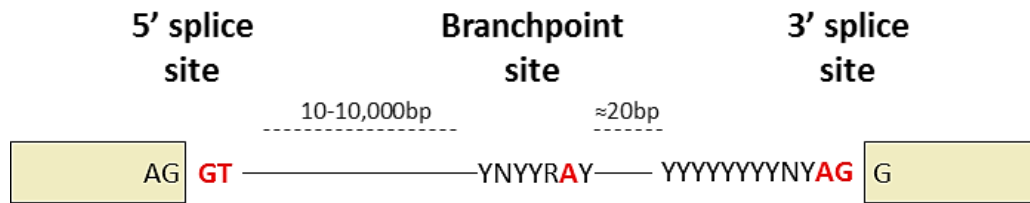
Figure 13: Schematic overview of splicing sites composition.

2. Spliceosomal proteins

"The spliceosome: the most complex macromolecular machine in the cell" (Nielsen et al., 2003)

Splicing is orchestrated by two spliceosome machineries: (i) the U2-dependent spliceosome that is necessary for the removal of U2-type introns, which account for 99% of splicing reactions and (ii) the U12-dependent spliceosome, which plays a role on the U12-type introns and is specific to a subset of eukaryotes (Patel and Steitz, 2003).

The spliceosome is a huge and highly dynamic machinery and is composed of RNA-protein complexes called small nuclear ribonucleoproteins (snRNPs). There are five types of snRNPs composed of five small nuclear RNAs (snRNAs): U1, U2, U4/U6 and U5; a common set of seven Sm proteins: B/B', D3, D2, D1, E, F and G; and a number of other specialized proteins. Another non-snRNP complex is the Prp19 complex. In total, the spliceosome comprises more than 150 proteins such as RNA-Binding Proteins (RBPs), RNA helicases or phosphatases (Wahl et al., 2009) making it the most complex macromolecular machine in the cell (Jurica and Moore, 2003; Nilsen, 2003). In order to orchestrate such an enormous complex, spliceosome assembly has to be ordered and well regulated in a stepwise fashion (Will and Lührmann, 2011).

3. Spliceosome assembly

The first step of the spliceosome assembly is the recognition of the 5' donor site by the snRNP U1. Non-snRNP factors U2AF65 and SF1 are recruited to the polypyrimidine tract and the branchpoint respectively (E complex) **(Figure 14)**. Then, the A complex is formed by the binding of the snRNP U2 on the branchpoint site. The next step consists in the pre-assembly of three snRNP U4/U6 and U5 (B complex). Once this complex is associated to pre-mRNA, the subunits U1 and U4 are released and a conformational change occurs giving rise to an active spliceosome complex (B' complex). Next, the 5' splice site is cleaved by catalytic reaction generating the C complex. The second enzymatic reaction is processed, the 3' splice site is cleaved, the intronic region (called "lariat") is released and exon boundaries are ligated. Spliceosome is then disassembled and the Exon Junction Complex (EJC) is deposited upstream of the exon-exon junction. In addition, it has been shown that core spliceosome components affect splicing outcomes partially due to functional interactions with RBPs (Papasaikas et al., 2015).
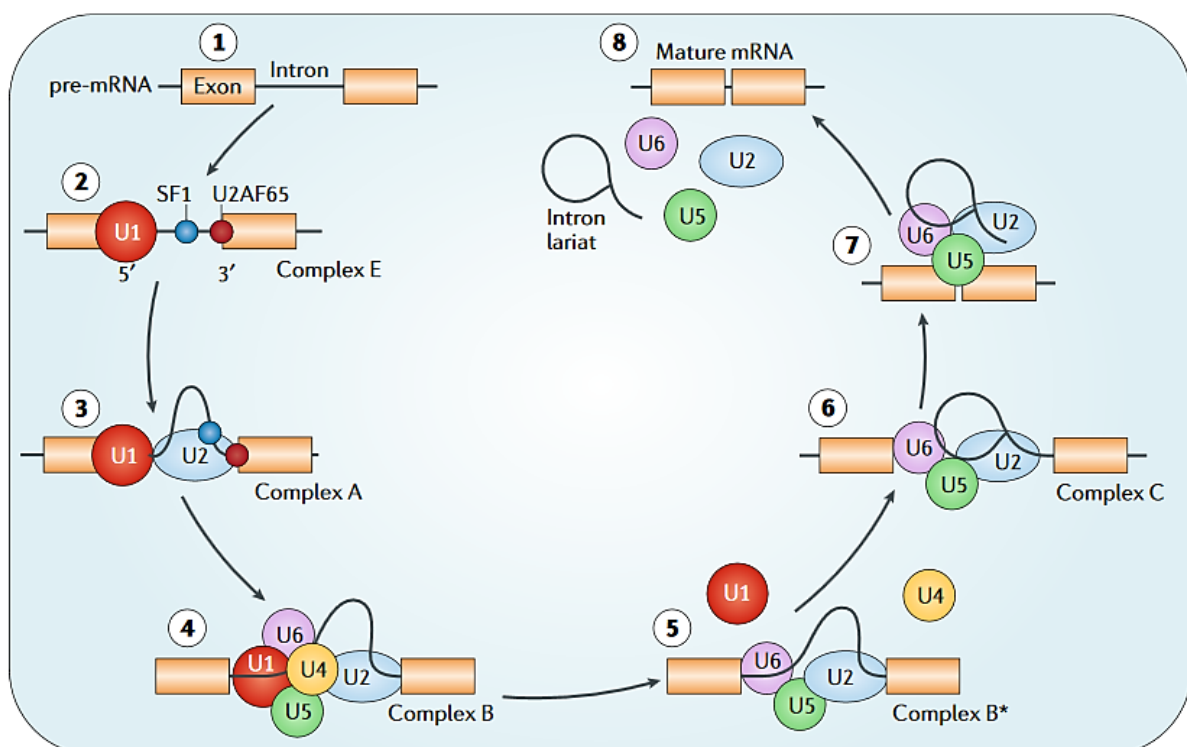
Figure 14: Overview of the spliceosome machinery assembly. Abbreviations: 5' splice site (5'SS); 3' splice site (3'SS); branchpoint (BP). From Paschalis et al., 2018.

4. Splicing regulatory sequences

Alternative splicing is also regulated through intronic and exonic *cis*-acting regulatory elements **(Figure 15)** (Burd and Dreyfuss, 1994; Fairbrother et al., 2002; Wang et al., 2004). These sequences are directly associated to the regulation of alternative splicing through the binding of specific proteins called RNA-Binding Proteins (RBPs). These *cis*-regulatory elements are classified into four classes, depending on their localization and function: (i) Intronic Splicing Enhancer (ISE), (ii) Intronic Splicing Silencer (ISS), (iii) Exonic Splicing Enhancer (ESE) and (iv) Exonic Splicing Silencer (ESS).
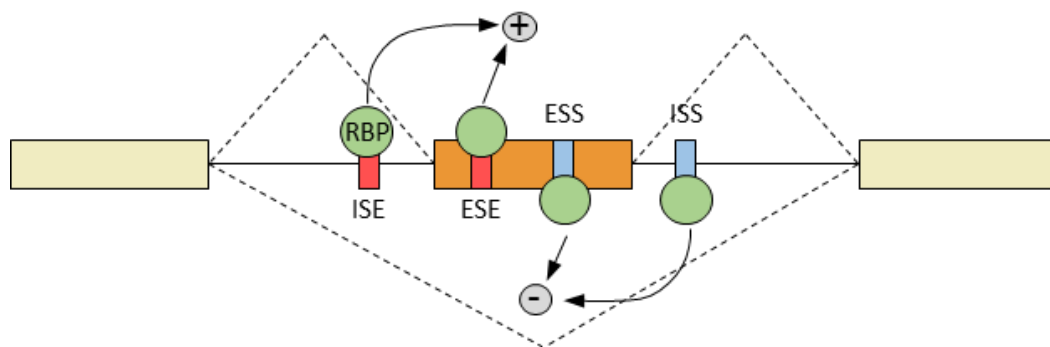
Figure 15: Schematic model of splicing regulation through RBPs. Abbreviations: RNA-binding proteins (RBPs); exonic splicing enhancer (ESE); exonic splicing silencer (ESS); intronic splicing enhancer (ISE); intronic splicing silencer (ISS). Adapted from Kornblihtt et al., 2013.

5. RNA-binding proteins

RNA binding proteins (RBPs) affect alternative splicing decisions by binding enhancer/silencer sites and interacting with the core spliceosome components. Each RBP usually binds specific RNA motifs of about 4 to 6 nucleotides, which have been precisely mapped in the human transcriptome for more than one hundred RBPs using high-throughput sequencing of RNA isolated from crosslinking immunoprecipitation experiments (CLIP-seq) (Hafner et al., 2010; Licatalosi et al., 2008; Rossbach et al., 2014; Ule et al., 2005; Van Nostrand et al., 2016). RBP expression is tightly regulated, depending on the cellular context. Some RBPs are only expressed in a certain tissue type or depending on the developmental stage, consistent with the importance of alternative splicing on cell-type function and development.

6. Coupling transcription and splicing

The RNA polymerase II (POL2) transcribes genes as precursor-mRNAs (pre-mRNAs). Afterwards, these pre-mRNAs molecules are converted to mRNA by multiple post-transcriptional modifications consisting in the 5' capping, the 3' polyadenylation and the splicing of introns. Nearly all human genes are spliced and approximately 80% of splicing reactions occur on the nascent RNA (Girard et al., 2012; Wang et al., 2008). Transcription and splicing occur concomitantly, thereby POL2 plays an important role on splicing decision (Beyer and Osheim, 1988; Cramer et al., 2001; Roberts et al., 1998; Tennyson et al., 1995). In addition, POL2 is able to recruit splicing factors and RNA processing factors through its C-ter domain (Misteli and Spector, 1999). This domain, called carboxy-terminal domain (CTD), is a highly phosphorylated region, which is required for cotranscriptional splicing (Hirose et al., 1999; Zeng and Berget, 2000). Moreover, POL2 elongation rate can directly influence splicing changes. Splice site selection mostly occurs cotranscriptionnally and the RNA POL2 kinetic can influence splicing decision (Cramer et al., 1999; Kadener et al., 2001).

A model has been proposed in which a fast elongation rate results in a competition between multiple splice sites (weak and strong ones) leading to the skipping of the weaker exon **(Figure 16)**. Whereas slowing down the RNA POL2 usually favors exon inclusion because only the first exon (weak) is already transcribed when the splicing machinery starts to operate (Cáceres and Kornblihtt, 2002; Kornblihtt, 2005; de la Mata et al., 2003). However, in some cases, slow elongation can favor RBP recruitement to a splicing silencer, thereby favoring exon skipping. (Dujardin et al., 2014).
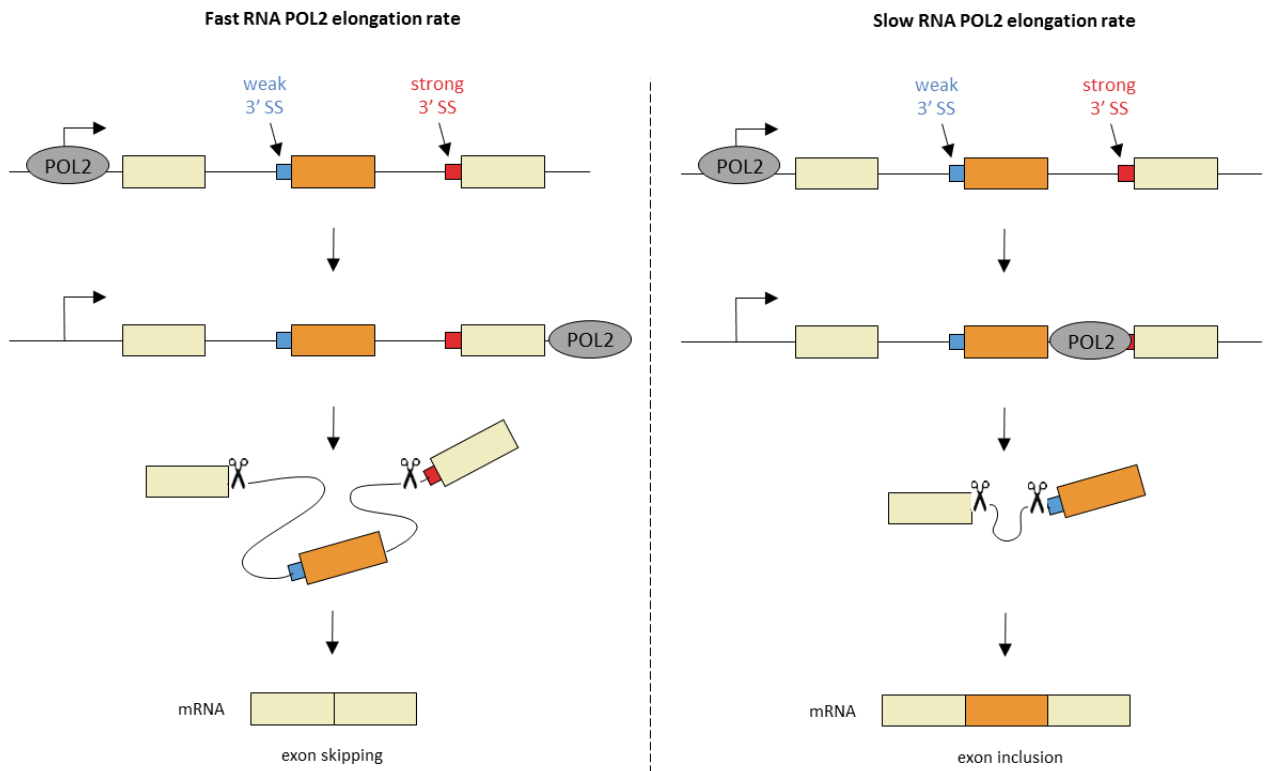
Figure 16: Putative model of the alternative splicing regulation through modulation of the RNA polymerase II elongation rate. Adapted from Kornblihtt et al., 2005.

Transcription factors typically bind response elements (RE) to orchestrate gene transcription regulation. Because of the coupling between transcription and splicing, transcriptional stimuli and transcription factors can impact alternative splicing (Auboeuf et al., 2002, 2007; Rambout et al., 2018). In addition to the regulation of splicing by controlling the RNA polymerase II elongation rate (Saldi et al., 2016), transcriptions factors can also serve as a scaffold to recruit splicing co-factors co-transcriptionally (Markus et al., 2006). As I will discuss below, this has been shown in the case of EWS-FLI1 (Sanchez et al., 2008a; Selvanathan et al., 2015)

Moreover, it is now well established that chromatin modifications can also impact on alternative splicing (Luco et al., 2010, 2011). These effects can be mediated by interactions between chromatin proteins and splicing factors through adaptor proteins. For instance it has been demonstrated that the chromatin binding protein MRG15 binds the histone tails H3K36me2,3 to recruit the RNA binding protein PTBP, which controls alternative splicing outcomes (Luco et al., 2010, 2011).

Furthermore, Allemand and colleagues have purified the U2 snRNP spliceosome complex and performed mass spectrometry analysis (Allemand et al., 2016). They highlighted the presence of the SWI/SNF complex but also proteins involved in the histone post-translational modifications (PTMs), thus confirming the role of epigenetic regulators in alternative splicing (Batsché et al., 2006).

Finally, lncRNAs can modulate chromatin structure through protein interactions to control alternative splicing and maintain cell-specific splicing programs (Gonzalez et al., 2015).

C. Alternative splicing as a new hallmark of cancer

Alternative splicing has been involved in all the main cellular hallmarks of cancer, including angiogenesis, survival, invasion and metastasis. Therefore, researchers propose to add alternative splicing as a new hallmark of cancer (Ladomery, 2013; Oltean and Bates, 2014).

1. Cell survival

Programmed cell death or apoptosis is a physiological process that is common to most organisms. *BCL2L1* gene (also known as Bcl-x) produces two distinct mRNAs, through alternative splicing, with antagonistic functions (Boise et al., 1993). The long isoform (Bcl-x$_L$), which contains all Bcl-2 homolgy (BH) domains, has a pro-survival function whereas the short one (Bcl-x$_S$), which lacks the first two BH domains, activates apoptosis **(Figure 17)**. More interestingly, both isoforms have a specific cell-type expression. Bcl-x$_S$ is mostly expressed in cells with a high turnover rate such as T-cells whereas Bcl-x$_L$ is expressed in adult brain where cell regeneration is low (Boise et al., 1993). In addition, antiapoptotic isoform Bcl-x$_L$ is overexpressed in cancers in comparison with normal tissue counterparts (Trisciuoglio et al., 2017). Therefore, targeting alternative splicing of *BCL2L1* gene would be a promising target to trigger tumor cell death. However, numerous splicing factors have been identified as splicing modulator of Bcl-x gene, highlighting the complexity of this process (Bielli et al., 2014; Garneau et al., 2005; Paronetto et al., 2007; Revil et al., 2007, 2009; Zhou et al., 2008). This is a wonderful example of how physiological processes can be hijacked by the splicing machinery to modify the balance between survival and apoptosis, thus inducing tumorigenic features.

Bcl-x pre-mRNA

Bcl-xS mRNA Bcl-xL mRNA

Bcl-x proteins

proapoptotic
high turnover rate

antiapoptotic
low turover rate
expressed in cancer
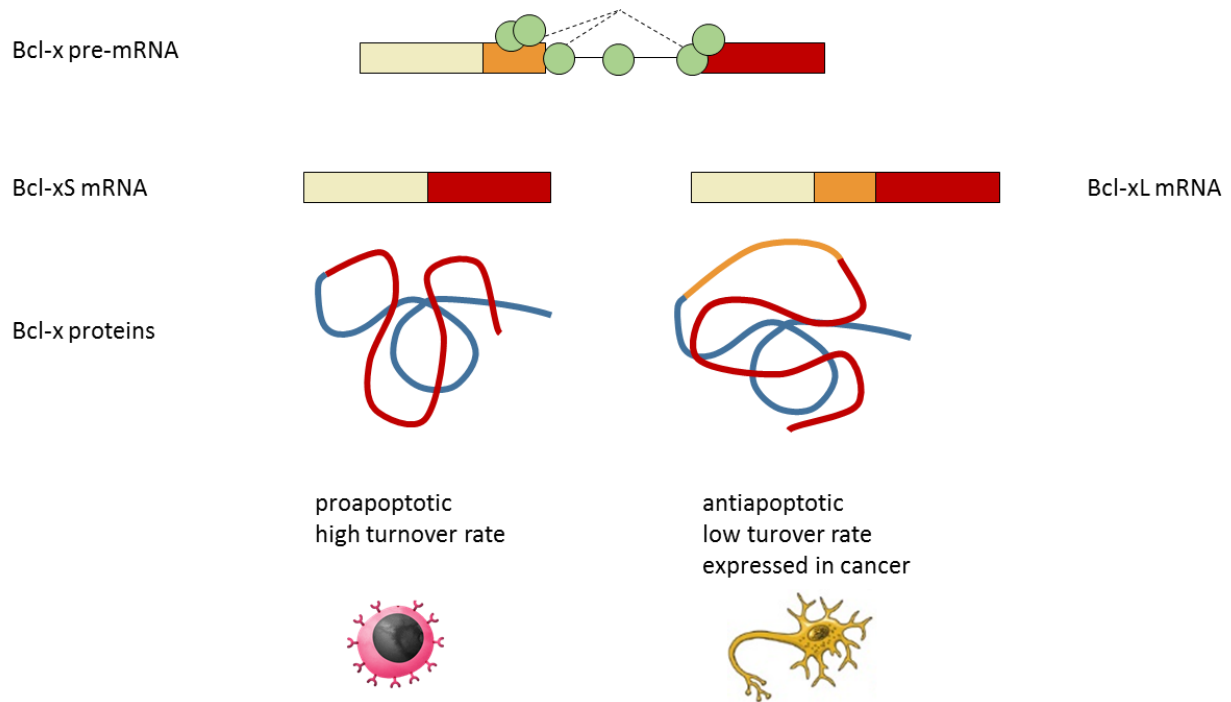
Figure 17: Bcl-x isoforms have distinct cellular functions. Several RBPs have been found to regulate Bcl-x alternative splicing, thus producing distinct isoforms with opposite functions. Adapted from Pio Ruben et al., 2009.

2. Epithelial-to-mesenchymal transition

The first observation of epithelial-to-mesenchymal transition (EMT) was made by Elizabeth Hay in 1968 during the chick embryo development (Hay, 1968). The term epithelial-to-mesenchymal "transformation" was introduced later on (Bolender and Markwald, 1979; Hay, 1995) and replaced by "transition" due to its transient and reversible properties. EMT is a highly complex process that defines a massive epigenetic and transcriptomic reprogramming of a polarized immotile epithelial cell towards a mesenchymal cell with migratory and invasive capacities (Kalluri and Weinberg, 2009).

During embryonic development, EMT is necessary for the formation of the organs structures (Thiery et al., 2009). For instance, epithelial cells from the neuroepithelium generate neural crest cells. These cells gain motility and acquire invasive abilities to spread from the neural tube, hence leading to dissemination to various parts of the embryo where they can differentiate into specialized cells (Acloque et al., 2009; Ahlstrom and Erickson, 2009). This physiological program requires complex and multiple signals such as microenvironment, transcriptional/post-transcriptional regulators and epigenetic factors (Garg, 2013; Nieto et al., 2016; Serrano-Gomez et al., 2016).

In addition, this process is often hijacked by cancer cells, in order to gain plasticity and to switch from a highly proliferative state to a pro-metastatic state (Cano et al., 2000; Thiery, 2002; Tsai and Yang, 2013). Furthermore, EMT has been widely implicated as therapeutic resistance mechanism in cancer such as in prancreatic and colon cancer (Arumugam et al., 2009; Kim et al., 2015; Wang et al., 2009).

Cells have the specificity to undergo EMT and conversely MET to switch back and forth between an epithelial to a mesenchymal state **(Figure 18)**. During this highly plastic process, cells express specific markers of each cell state, one of the most known is the cell-cell adhesion molecule E-cadherin (gene name: *CDH1*) (Boyer and Thiery, 1993; Kalluri and Weinberg, 2009; Zeisberg and Neilson, 2009). Over the past decade, most studies focused on transcription factors EMT inducers such as SNAI1 (Cano et al., 2000), ZEB2 (Comijn et al., 2001) and TWIST1 (Yang et al., 2004). Nevertheless, accumulating evidence suggest that alternative splicing regulators appears to have a major role in governing cell plasticity and control of EMT process (Bonomi et al., 2013; Braeutigam et al., 2014; Chaudhury et al., 2016; Pradella et al., 2017; Shapiro et al., 2011; Warzecha and Carstens, 2012; Yang et al., 2016).
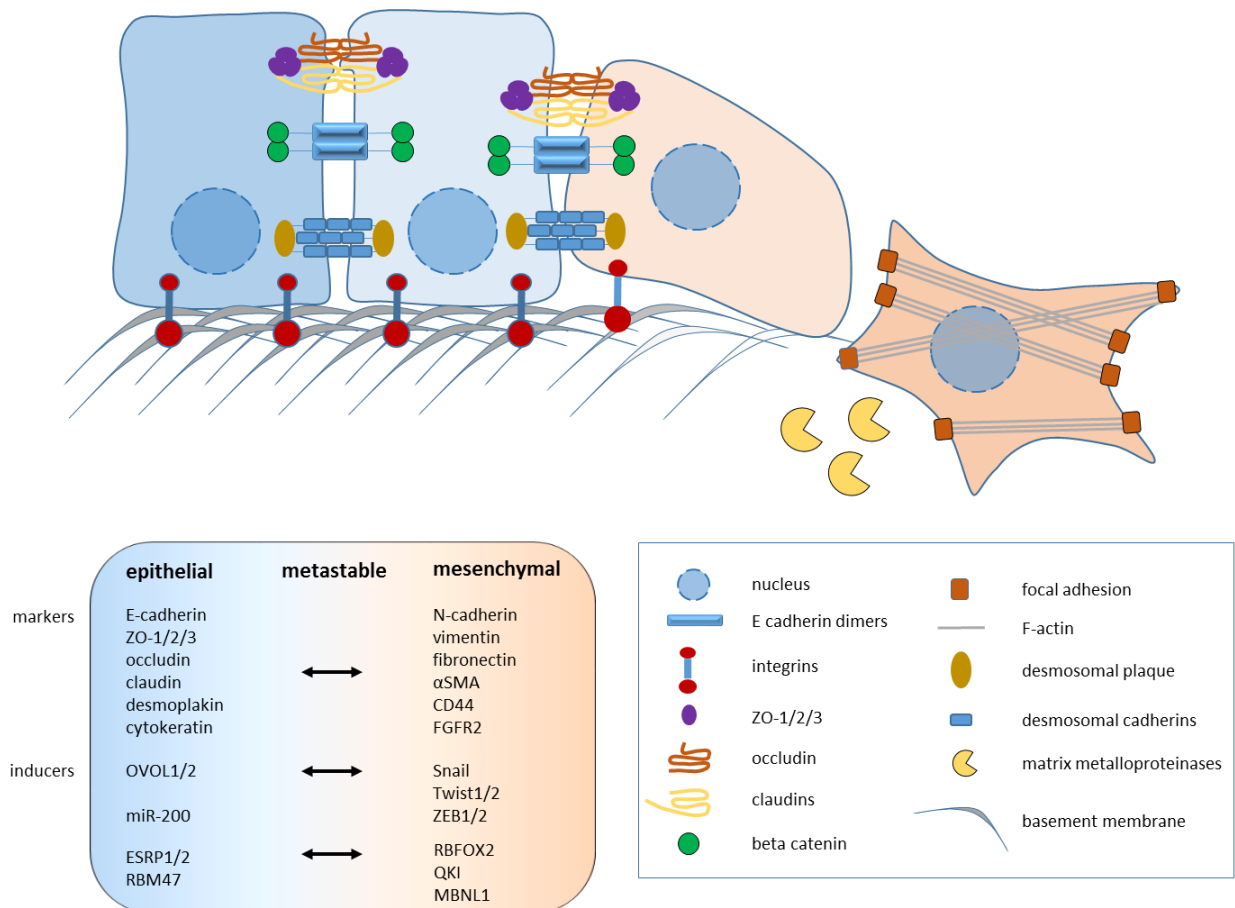
Figure 18: Schematic overview of epithelial-to-mesenchymal transition. Epithelial-to-mesenchymal transition involves functional transition from a polarized epithelial cell towards a mesenchymal-migrating cell. Cells expressing the two sets of markers indicate that the process is partial and cells exhibiting an intermediate phenotype that we can call "metastable".

One of the best examples of cell type-specific RBPs is the epithelial splicing regulatory proteins 1 and 2 (ESRP1 and ESRP2) that are epithelial cell-type specific splicing regulators. They are involved in EMT and contribute to drive the epithelial-phenotype (Bebee et al., 2015; Warzecha et al., 2010). ESRPs proteins bind UG-rich motifs to promote splicing with a position-dependent mechanism. ESRPs proteins regulate exon skipping or exon inclusion by binding upstream or downstream intron, respectively (Dittmar et al., 2012). In contrast, RBFOX2, a master splicing regulator, is more expressed in mesenchymal than epithelial cells. It controls a mesenchyme-specific splicing program and is involved in EMT and late mesoderm differentiation (Braeutigam et al., 2014; Fici et al., 2017; Mallinjoud et al., 2014; Shapiro et al., 2011; Venables et al., 2013a, 2013b).

In particular, Shapiro and colleagues have wonderfully demonstrated by using an EMT-inducible model, based on TWIST overexpression, that several splicing factors were directly involved in the regulation of this process (Shapiro et al., 2011). They have shown that ESRP1 splicing factor is differentially expressed depending on the EMT cell status and its ectopic expression in mesenchymal cells shifted cell phenotype and abrogated migratory capacities. Additionally, depletion of RBFOX2 splicing factor in mesenchymal cells also conferred numerous epithelial characteristics but the phenotype switch was slighter. This study established a specific EMT splicing program governed by ESRP1 and RBFOX2 splicing factors. These factors play a determinant role in epithelial and mesenchymal phenotype, respectively, and may be important for tumor dissemination. Aditionally, Ranieri and colleagues have demonstrated that FGFR2 spliced variant was able to induce EMT in human keratinocyte cell line. Ectopic expression of a specific isoform of FGFR2 in this epithelial cell line modified cell morphology, migratory capacities and tumorigenic properties (Ranieri et al., 2015). Altogether, this highlighted that splicing factors, as well as specific isoforms, are directly involved in the regulation of key cellular processes including EMT.

3. Splicing as an oncogenic driver

One of the best demonstrations of the oncogenic role of splicing is probably the serine/arginine-rich splicing factor 1 (SRSF1). Its localization is mainly nuclear but it has the ability to shuttle between the nucleus and the cytoplasm depending on its phosphorylation state (Cáceres et al., 1998; Gui et al., 1994). *SRSF1* overexpression is often observed in cancer (Ghigna et al., 1998; Gout et al., 2012; Karni et al., 2007) and leads to cell transformation in immortalized fibroblasts (Anczuków et al., 2012; Karni et al., 2007). Oncogenic potential of SRSF1 is thought to be a cumulative effect of its various functions on alternative splicing. Several oncogenic splicing events induced by SRSF1 factor have been extensively described in this context. For instance, SRSF1 binds exonic enhancer and aberrantly regulates splicing isoform of macrophage stimulating 1 receptor (MSTR1, also known as Ron), hence producing RonΔ11 isoform, which lacks the cassette exon 11 of 147bp (Collesi et al., 1996; Ghigna et al., 2005). This results in a critical change in Ron receptor conformation and function. Indeed, RonΔ11 isoform leads to a constitutive activation of Ron receptor, which is sufficient to confer motility and migratory capacities (Collesi et al., 1996; Ghigna et al., 2005).

Morever, several splicing factors are recurrently found mutated in cancers (Anczuków and Krainer, 2016). For instance, splicing factor 3b subunit 1 (*SF3B1*) is the most frequently mutated splicing factor in cancer and is present in around 50% of patients with myelodysplastic syndromes (Malcovati et al., 2011; Yoshida et al., 2011). SF3B1 mutant protein recognizes an alternative branchpoint compared to the wild-type SF3B1 leading to the use of a cryptic 3' splice site usage located upstream the canonical branchpoint (Alsafadi et al., 2016; Darman et al., 2015; DeBoever et al., 2015).

Transcriptome wide analysis studies highlighted that alternative splicing programs are massively modified in cancer (Sebestyén et al., 2016). In addition, because alternative splicing alteration in cancer is negatively correlated to the number of mutations, it has been proposed that splicing alteration may represent an independent oncogenic process, which could acts as potential driver (Climente-González et al., 2017).

More recently, a comprehensive analysis established the splicing landscape of 32 tumor types from The Cancer Genome Atlas (TCGA) data. They found that tumors have up to 30% more splicing events as compared to matched normal tissues (from TCGA and GTEx database: https://www.gtexportal.org/) and novel unannotated splicing events is a common feature in cancer (Kahles et al., 2018). In addition, they observed that some tumors exhibited massive numbers of splicing alteration and called this phenomenon "syndeothripsis", in analogy to chromothripsis (Stephens et al., 2011). However, the splicing burden observed in those tumors is not yet well understood. Furthermore, they identified a subset of recurrent tumor-specific neojunctions that have the potential to generate neopeptides with high MHC-I-affinity. Therefore, this study provides a global overview of splicing pattern alterations in cancer and new insights on how alternative splicing could be used to develop immune-based tools for diagnosis, prognosis and therapy.

Altogether, this illustrates how splicing complexity contributes to tune crucial biological processes involved in cell physiology and cancer. Comprehensive analysis of tumor splicing landscape is crucial to develop new targeted therapies.

D. Splicing regulation by FET, ETS and FET-ETS proteins

1. Splicing regulation by FET proteins

The FET family proteins are involved in several physiological processes such as transcription, RNA metabolism and splicing (Svetoni et al., 2016). FET members have been reccurently identified as chromosomal translocation partners as in Ewing sarcoma or liposarcoma (Delattre et al., 1992; Rabbitts et al., 1993). For instance, FUS protein can bind GGUG-containing RNAs to recruit splicing factors such as SC35 and TASR to alter alternative splicing (Yang et al., 1998). EWS protein regulates the RNA polymerase II elongation rate and interacts with the core spliceosome component U1C, *via* its low complexity domain, to induce alternative splicing (Bertolotti et al., 1998; Knoop and Baker, 2000, 2001; Paronetto et al., 2011; Sanchez et al., 2008a).

2. ETS transcription factors as splicing modulators

Several reports have demonstrated that ETS transcription factor family is implicated in post-transcriptional processes such as splicing or mRNA degradation. For instance, Spi-1-/PU.1 is able to bind RNA *via* its DNA-binding domain to induce alternative splicing (Guillouf et al., 2006; Hallier et al., 1996, 1998). However, effect of ERG subfamily (ERG, FLI1 and FEV) on splicing has not been documented yet. Nonetheless, it has recently been shown that ERG is abe to control mRNA degradation through interaction with RBPs (Rambout et al., 2016). Altogether, these observations suggest that ETS proteins should not be considered only as transcription factors and might be implicated in RNA-splicing.

3. Splicing regulation by EWS-FLI1 fusion protein

In addition to its gene expression regulation activity, EWS-FLI1 influences post-transcriptional decisions through splicing mechanisms. Knoop and colleagues have shown that the fusion protein interacts with the splicing factor U1C, a component of the U1 snRNP, which is necessary for the early stages of spliceosome formation (Knoop and Baker, 2000). This observation reveals that EWS-FLI1 may have the ability to alter splicing. Shortly after, they demonstrated that EWS-FLI1 induces alternative splicing of a specific reporter through interference with the heterogeneous nuclear ribonucleoprotein A1 (hnRNP A1) (Knoop and Baker, 2001).

EWS-FLI1 interacts with RNA-polymerase II (POL2), which influences alternative splicing outcomes (Cramer et al., 1999; Roberts et al., 1998; Yang et al., 2000). Sanchez and colleagues have demonstrated that EWS-FLI1 oncoprotein decreases POL2 elongation rate on the cyclin D1 gene (*CCND1*), therefore promoting a splicing isoform called cyclin D1b that corresponds to the use of an intronic polyadenylation site and thus an alternative last exon. EWS had opposite effects when compared to EWS-FLI1 on cyclin D1 elongation rate and alternative splicing. Authors suggested a potential role of this CCND1 isoform on Ewing sarcoma oncogenesis. Indeed, the cyclin D1b isoform is expressed at higher level in Ewing tumors as compared to normal tissues. It plays a role on cell growth and was previously shown to be more oncogenic than the canonical cyclin D1a isoform (Sanchez et al., 2008a, 2008b).

More recently, EWS-FLI1 was described binding RNA, thus suggesting a direct role in alternative splicing (Erkizan et al., 2015). Few months later, the RNA binding motif of EWS-FLI1 was published using Cross-Linking ImmunoPrecipitation sequencing (CLIP-seq) (Selvanathan et al., 2015). These observations are surprising because the RNA binding domain of EWSR1 is not present in the fusion protein and FLI1 has never been shown to directly bind RNA. Using liquid chromatography–mass spectrometry (MS), enrichment in RNA-processing proteins as EWS-FLI1 interactors was observed. Several proteins of the early and late components of the spliceosome machinery (U1 and U5) directly interact with the fusion protein supporting the regulatory role of EWS-FLI1 on splicing **(Figure 19)**. Additionally, they have shown by exon array that EWS-FLI1 regulates exon usage of more than 200 genes especially the γ-TERT isoform that have an increased telomerase activity. Collectively, these results point out the ability of the EWS-FLI1 fusion protein to impact on alternative splicing programs.
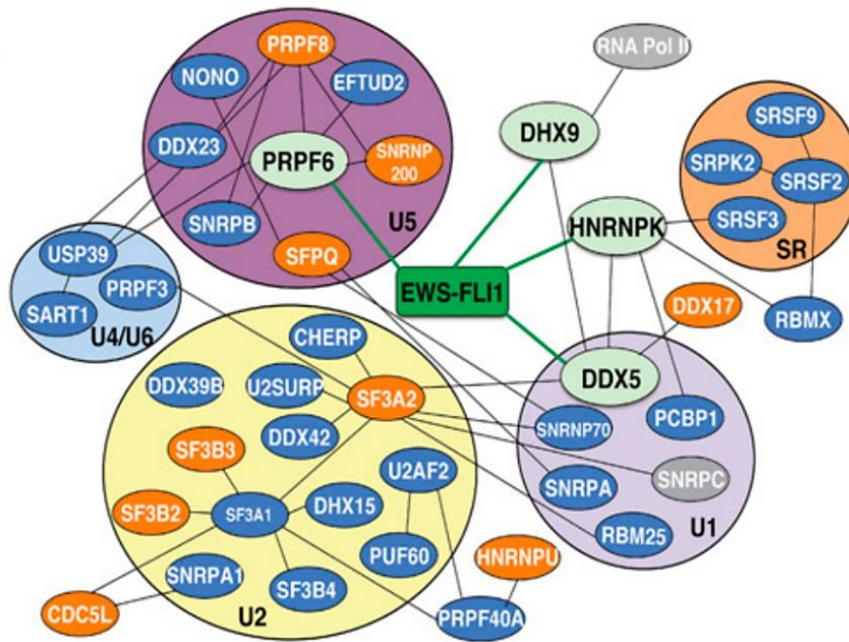
Figure 19: Mass spectrometry analysis identified spliceosome-associated proteins as EWS-FLI1 binding partner. Direct EWS-FLI1 partners that have been validated by ELISA are indicated in green. Indirect partners validated using co-immunoprecipitation are in orange. Non validated binding partners are in blue and gray nodes represent protein partners that have not been identified by MS but predicted to interact with EWS-FLI1. From Selvanathan et al., 2015.

Almost twenty years ago, Knoop and Baker suggested in their paper that the splicing alteration induced by the fusion protein might be mediated *via* the FLI1 moiety because the wild-type EWS did not show any activity on this reporter (Knoop and Baker, 2001). This hypothesis is fascinating because, since decades, EWS is known to play a major role on splicing and to induce a large set of alternative splicing events (Paronetto et al., 2011; Rappsilber et al., 2002; Sanchez et al., 2008a; Tan and Manley, 2009; Wu and Green, 1997; Yang et al., 1998). That is why EWS-FLI1 function on splicing is mostly attributed to its EWS part. Our study put the lights on an unexpected function of ETS and especially the ETS moiety of FET-ETS fusions in alternative splicing regulation.

III.   Splicing analysis

A. Technologies

In 1977, Frederick Sanger developed the first sequencing technology allowing identification of each nucleotide from a DNA sequence (Sanger et al., 1977). Three years later, he is awarded of the Nobel Prize in Chemistry. Almost fifteen years and more than 1 billion dollars were necessary to fully sequence the first Human genome in 2003 (International Human Genome Sequencing Consortium, 2004). Few years later, appeared the second generation of sequencers or Next Generation Sequencing (NGS) technologies (Margulies et al., 2005). NGS technologies are massive parallel systems that allow sequencing of billions of nucleotides in one single run. Plenty of applications have been developed to decipher the cellular complexity **(Figure 20)**.
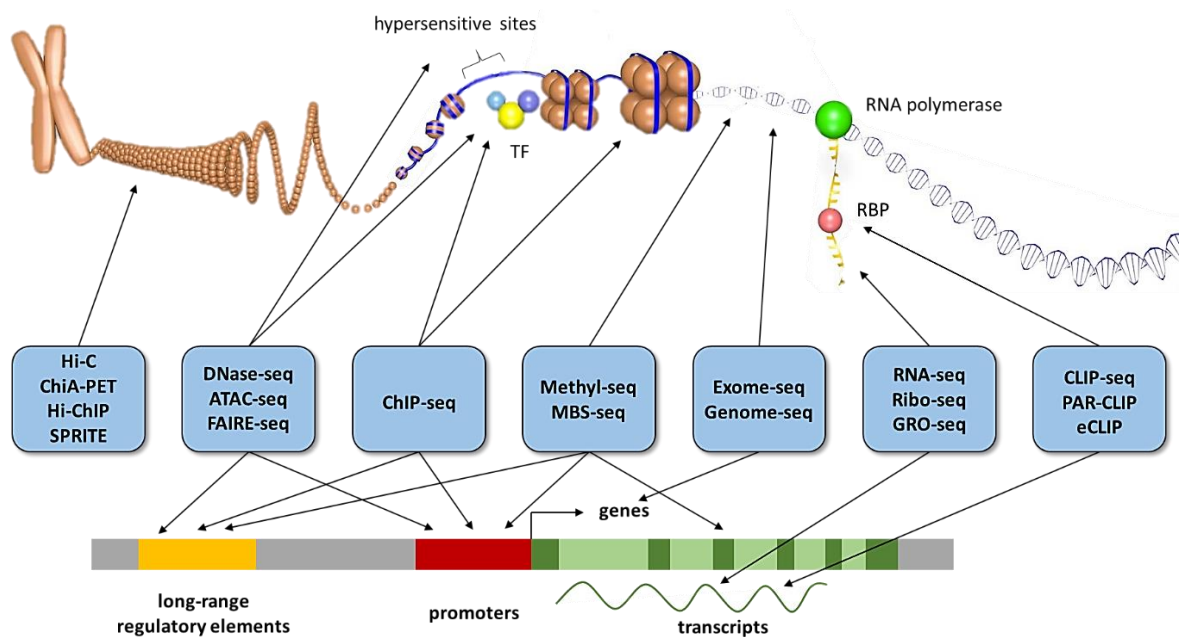
Figure 20: Overview of a subset of next generation sequencing applications. Adapted from ENCODE. Adapted from ENCODE.

1. RT-PCR

This old fashion method is still routinely used to validate a specific splicing event because it is reliable and fast to set up. RT-PCR allows detection of multiple transcripts by using specific primers located in constitutive exons flanking your splicing event of interest. Then, PCR products are run on an agarose gel in order to separate both transcripts according to their size **(Figure 21)**. The limit of this method is that the design of specific primers is required for each splicing event.
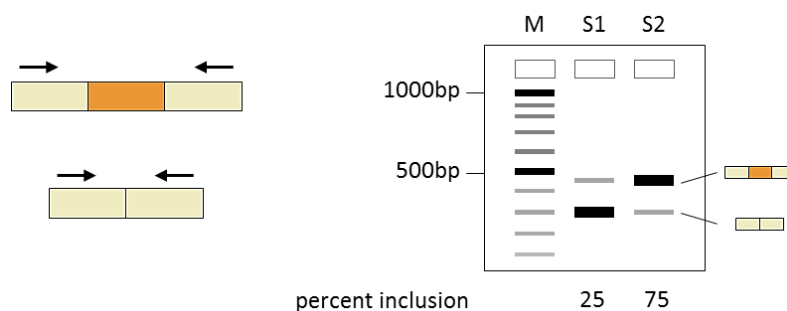
Figure 21: Schematic overview of alternative splicing analysis using RT-PCR method. Abbreviations: DNA ladder (M); PCR products of sample 1 (S1) and sample 2 (S2).

2. Exon-arrays

Microarray technology is widely used to assess mRNA expression levels. Affymetrix, one of the most popular companies in the microarray field, released Exon arrays in order to study exon expression, hence alternative splicing. For each known exons (from databases but also from algorithmic predictions), several probes were designed and spotted on the chip. For instance, the Human GeneChip® Exon 1.0 ST Array is composed of 1.4 million of probe set distributed along the human coding genome (Affymetrix: Exon Array Design Datasheet http://www.affymetrix.com/support/technical/datasheets/exon_arraydesign_datasheet.pdf). Exon-level analyses provide relevant information on exon expression in order to detect isoform variation along the genome. However, limitation of exon arrays is that they cannot detect probes, which are not included in the design.

3. RNA-seq

With the development of hight-throuhput sequencing and the explosion of bioinformatics tools, knowledge and discoveries have considerably increased in the past decade, especially on splicing (Djebali et al., 2012). Studying transcriptome is crucial to identify and to interpret molecular mechanisms that govern a specific process or a phenotype. With the advance of NGS, it is now possible to identify the complete set of transcripts in a given cell without any preconceived ideas. This method is called RNA-sequencing (RNA-seq). Bainbridge and colleagues published the first paper on whole-transcriptome analysis of a prostate cancer cell line using high-throughput sequencing and basic local alignment search tool (BLAST) (Bainbridge et al., 2006). This revolutionary tool allows to quantify rare RNA molecules and to capture the transcriptome landscape of a cell. Numerous protocols have been developed to study a subpopulation of RNA molecules such as mRNA, miRNA, ncRNA. I will focus on the rest of the manuscript on mRNA sequencing. Most NGS technologies, illustrated in **Figure 22**, are based on short-reads sequencing (from 50 to 300bp per read). This is the biggest limitation of these technologies because computational approaches to reconstruct full-length transcripts from short reads sequencing is challenging. With recent advances, in particular from Pacific Biosciences (PacBio) company, it is now routinely feasible to sequence full-length mRNA molecules. This is a huge step to better characterize unknown genomes, due to improvement in *de novo* assembly.

Additionally, the "dark" genome (repetitive elements, extreme GC sequences, centromeres, etc...) is now in the spotlight of this novel generation of sequencers. To date, only Oxford Nanopore technology developed a machine that directly sequence RNA molecules without any pre-amplification (Garalde et al., 2018).

In general, sequencing libraries are made of a high quality extraction with a quality metrics called RNA integrity number (RIN) usually above seven. Retro-transcription of mRNA molecules is performed using poly(T) oligonucleotides and 3' ends of cDNA are adenylated to perform ligation of adapters. Then, clonal PCR amplification is performed on a solid surface (*e.g.* flow-cell for Illumina technology) or in liquid (*e.g.* emulsion beads for 454 technology). Sequencing is carried out and each nucleotide is identified *via* fluorescence measurement (*e.g.* reversible terminator) or optical measurement (*e.g.* pyrosequencing).
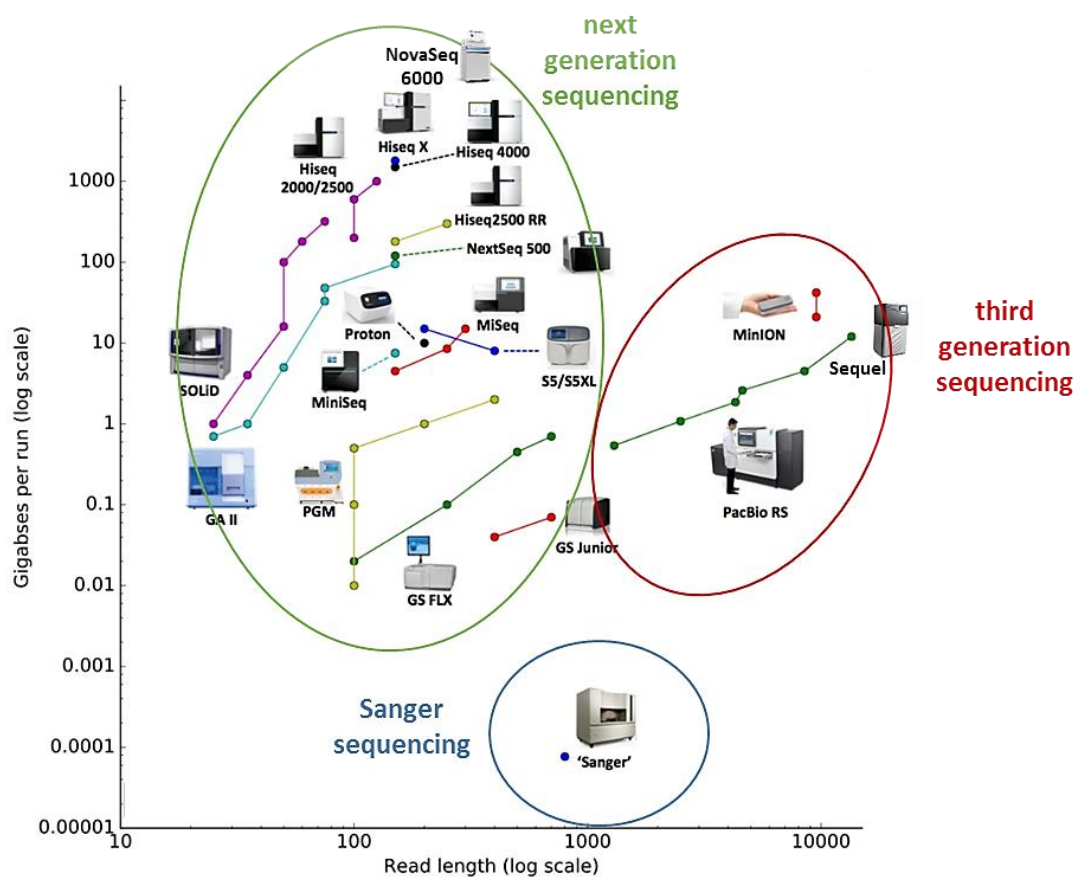
Figure 22: Sequencing technologies: from 1977 to 2018. Adapted from Lex Nederbragt 2016.

4. Illumina

Illumina sequencers are the most used across the world due to their low effective cost and high sequencing depth. Formerly marketed by Solexa company, a wide-range of sequencers exists, from the pioneer Genome Analyser to the novel NovaSeq system. This sequencing method closely looks like the Sanger synthesis method, but uses special fluorescently labelled terminator nucleotides, which allow the chain termination process to be reversed (Bentley et al., 2008).

Double stranded template cDNA molecules are denatured and hybridized on a flow-cell coated with oligonucleotides corresponding to adapters. Clusters of cDNA are generated by clonal bridge amplification. The sequencing reaction is performed by adding all elements for DNA elongation process. The first chain extension is performed; a specific fluorescently terminator reversible nucleotide is incorporated and the reaction stops. Secondly, all non-incorporated elements are washed out and the incorporated nucleotide is identified by fluorescent imaging. Thirdly, the terminator is chemically cleaved from the incorporated nucleotide and another sequencing cycle begins **(Figure 23)**. These three steps are repeated over one hundred times (depending on the read length) and all images are computationally converted into a final nucleotide sequence corresponding to the initial template DNA.
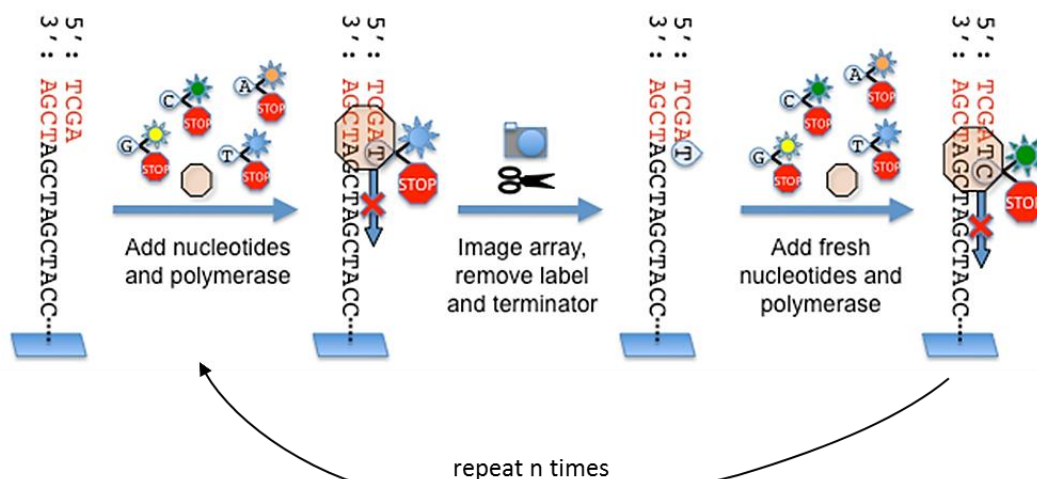
Figure 23: Schematic principle of Illumina sequencing. Adapted from Anderson et al., 2010.

5. Pacific Biosciences

PacBio has developed the Single molecule real time (SMRT) approach. This technology is based on a chip that contains over one million of holes, called zero-mode waveguide (ZMW). Long-life DNA polymerase is covalently attached to the bottom of each ZMW and a unique cDNA molecule penetrates into the hole to be elongated. Each DNA incorporation is recorded in real-time by a fluorescent signal. PacBio sequencer offers full-length mRNA sequencing and fast runs (about 10 hours) but has a high error rate (from 10% to 15%). Nonetheless, this limitation is bypassed due to the very low systematic error, meaning that errors are randomly distributed along the read length. DNA polymerase can elongate multiple times the same unique cDNA (due to the circularization of the molecule), hence a circular consensus sequence (CCS) is extracted with less than 1% error **(Figure 24)** (Goodwin et al., 2016). The revolutionary aspect of PacBio technology is the full-length mRNA sequencing, which gives rise to the complete transcriptome landscape of a cell. Each exon can be specifically assigned to a transcript allowing to precisely know what is the exon composition of each transcript.
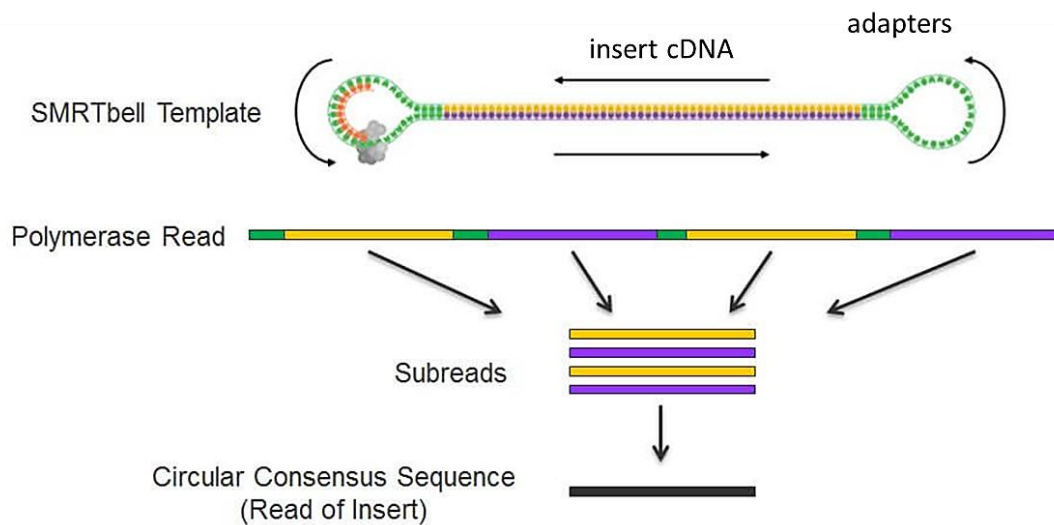
Figure 24: Schematic principle of PacBio Circular Consensus Sequence (CCS) strategy. Template cDNA is circularized and the polymerase with high processivity can read the template multiple times. Then, a consensus sequence is built based on these subreads to remove sequencing errors. Adapted from https://www.pacb.com.

B. Bioinformatics

All these advances in next generation sequencing require a lot of computational programs, storage spaces and bioinformaticians. There are more than 200 different tools referenced in the OmicTool database (https://omictools.com/; July 2018) to study alternative splicing. The large diversity and the number of available tools reflect the complexity to analyze such data as alternative splicing.

1. Pipeline

Here is an overview of the pipeline that I set up during my thesis to analyze alternative splicing from raw reads to an excel sheet with a list of alternative splicing events that can be used at the bench for biological validation and further experiments **(Figure 25)**.
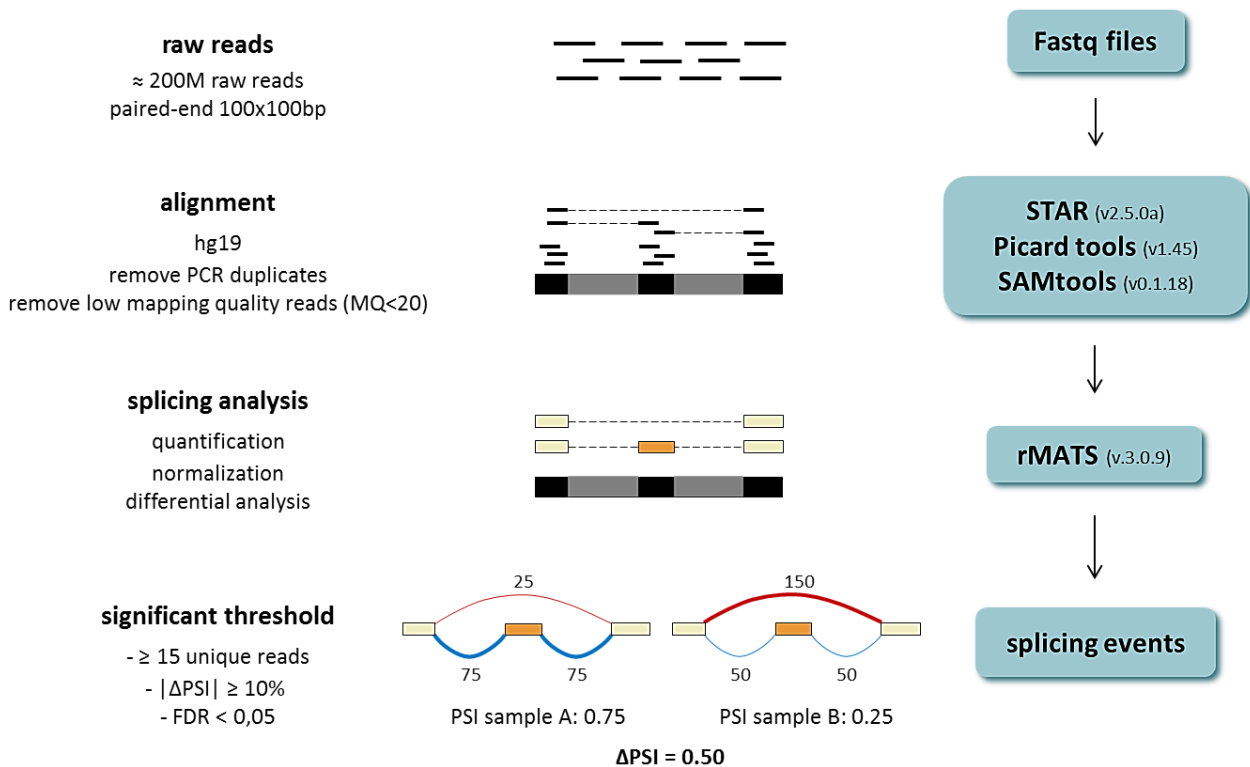
Figure 25: Overview of the pipeline to study alternative splicing. Abbreviations: million reads (M); mapping quality (MQ); percent of spliced-in (PSI); false discovery rate (FDR).

2. Quality control

Analysis of splicing in RNA-seq data mainly relies on junction reads (reads that span two exons), which directly arise from a splice transcript. In order to capture a sufficient number of junction reads on low expressed transcripts, it is recommended for mRNA-seq to sequence around 100 million of reads on paired-ends library of 100bp. After the sequencing step, there are several metrics to assess the quality of the RNA-seq experiment. FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc) is a tool developed by the Babraham lab in Cambridge, UK. It gives a full report including numerous metrics such as the base quality score across all your reads, the nucleotide frequency per base, the length distribution and overrepresented sequences. This first quality check is necessary to identify potential problems during library preparation or sequencing such as contamination. In addition, it is important to look at numbers provided by the aligner (after the mapping step). Generally, for a human RNA-seq, we expect more than 80% of uniquely mapped reads and less than 25% of PCR duplicate reads.

3. Alignement

In order to quantify alternative splicing events, reads need to be aligned on the genome using spliced aligners because of the junction reads. We distinguish spliced mappers (TopHat, STAR, GNSAP) from non-spiced mappers (Bowtie, BWA), which are more suitable for prokaryotic RNAseq. Two main aligners take the spotlight in the RNA world are TopHat (Trapnell et al., 2009) and STAR (Dobin et al., 2013). The first one use an "exon-first" approach that consists in mapping the whole reads against the reference genome. Then TopHat looks for spliced alignments with the remaining reads. On the other hand, STAR is based on the "seed-and-extend" method that cuts reads in k-mers and then compare them to the reference genome. Once a k-mer is uniquely aligned in the genome, the k-mer is extended until splice junction is found **(Figure 26)**. This last method increases the percent of mapped reads on the genome and has low false positive-rate. However, STAR needs high computational resources (around 30GB of RAM).
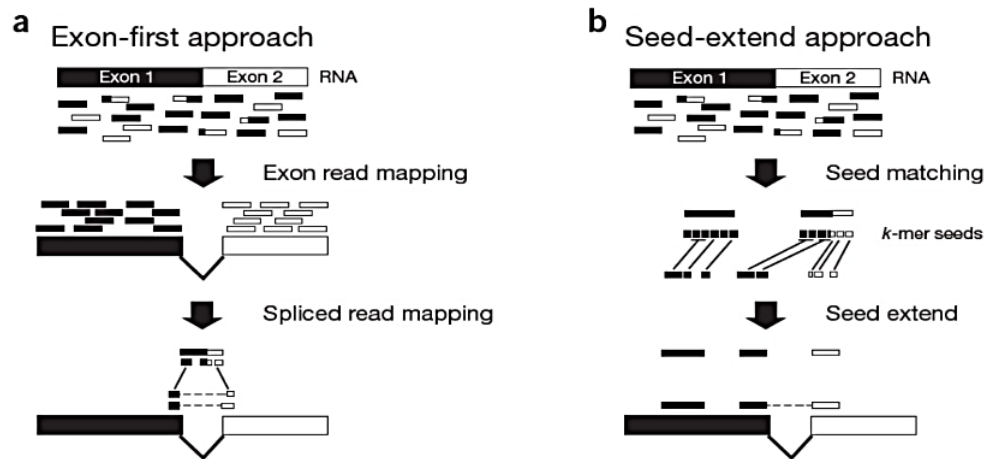
Figure 26: Overview of the two mapping methods. (a) Exon-first approach; (b) seed-extend approach. From Garber et al 2011.

4. Quantification

As for the alignment step, there are many ways to quantify transcripts abundance. Here, I will focus on the two main algorithms to detect differentially spliced isoforms in RNA-seq data. Isoform-reconstruction-based tools, such as Cufflinks (Trapnell et al., 2010) relies on "overlap graph" to distinguish incompatible reads that must originate from different isoforms. Transcript assembly is performed by implementing Dilworth theorem (i.e. "the number of mutually incompatible reads is the same as the minimum number of transcripts needed to explain all the fragments"; (Trapnell et al., 2010)). Then, transcript abundance is estimated for each predicted mRNA molecules. The strength of this method is that you can identify all expressed transcripts and their abundance in a given cell or in a specific condition. Nevertheless, limitations are the lack of indication on the exons/junctions involved and, as discussed before, short-reads sequencing implies complex algorithm methods to reconstruct entire transcripts and false positive rate remains high. During my thesis, I have found that approximately 20 to 30% of differentially spliced transcripts identified by Cufflinks were false positives.

To surpass this limitation, exon-centered methods have been developed such as rMATS (Shen et al., 2014). Instead of having a full view of the transcript composition, rMATS focuses on local scale such as exons and junctions. RNA-seq reads mapping to exon-exon junctions are used to calculate the relative inclusion of each exon in two samples (or more). Then, using Markov chain method, rMATS computes Bayesian probabilities that the splicing difference for a given exon do not exceed a threshold. This method is extremely powerful and the number of false positives is low. Nonetheless, the main limitation of this method is how to interpret these events. There is no information in which transcript belongs the differentially expressed exon. Public databases are available to identify functional domains associated to exons, such as Ensembl (https://www.ensembl.org). As for RT-PCR, by using RNA-seq reads it is possible to calculate the splicing frequency of a specific exon. The percent of spliced-in (PSI) value is a metric that represents the relative percentage of inclusion of a given exon **(Figure 27)**.
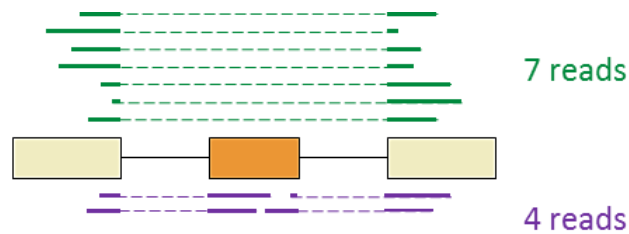
$$\text{PSI} = \frac{\#\ inclusion\ reads}{(\#\ inclusion\ reads + \#\ skipping\ reads)} = \frac{4}{(4+7)} = 0.36$$

Figure 27: Quantification of an alternative splicing event using the percent of spliced in (PSI) value.

5. Visualization

Even with this enormous collection of available tools, splicing analysis is still complex to interpret and you have to dig into the data to understand the biological interpretations of the analysis. Vizualization of the alternative complexity is an important aspect of the analysis and dedicated tools have been developed to allow the visualization of RNA-seq data. One way to have a more concrete interpretation of these data is to directly visualize them on a genome browser such as integrative genome viewer (IGV) (Robinson et al., 2011). Using IGV, user can display several RNA-seq data and scroll over the genome or in a specific region of interest. It allows to observe directly a splicing event and to understand the exon structure of a specific region. Several features are displayed on the viewer: (i) individuals reads, (ii) junction reads and (iii) read coverage **(Figure 28)**. In addition, junction reads of a specific event can be visualized using sashimi-plot allowing a direct quantification of the junction reads number **(Figure 29)** (Katz et al., 2015).
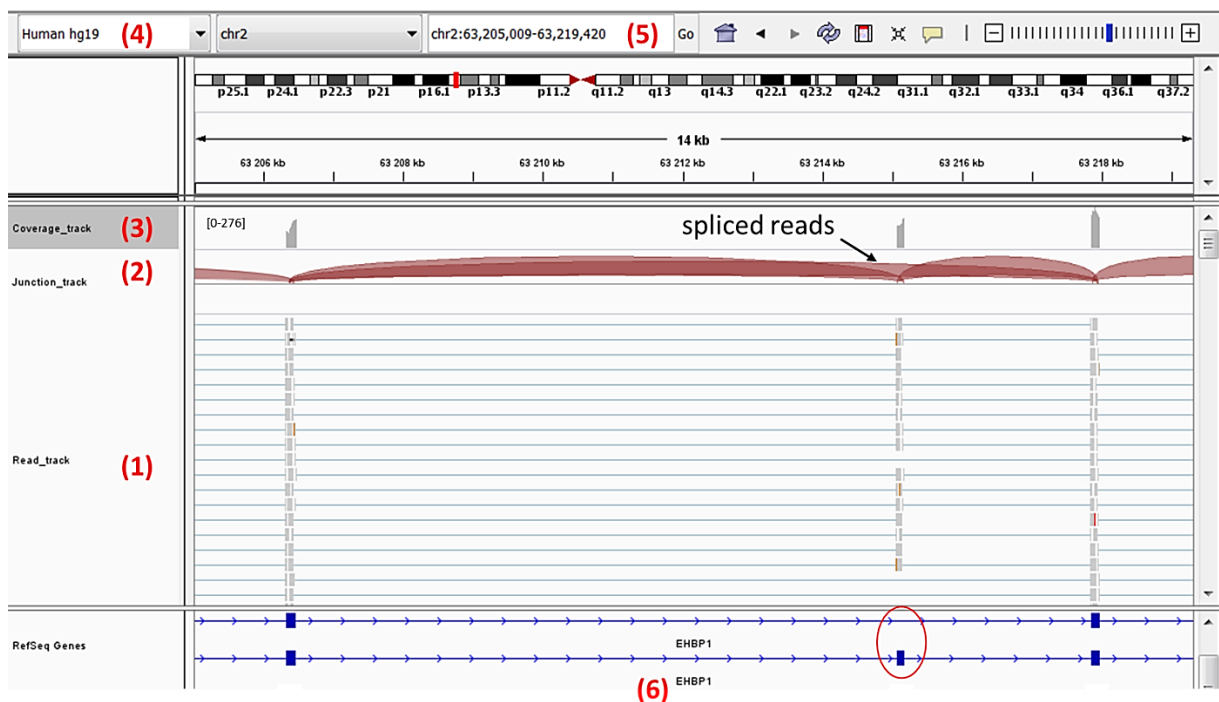
Figure 28: IGV screenshot of a genomic region in EHBP1 gene: (1) Read track representing sequencing reads. Each gray box is a read and grey lines are spliced alignments; (2) junction track is represented by an arc of the circle more or less intense depending on the quantity of reads; (3) cumulative read distribution along the genomic region; (4) reference genome; (5) genomic coordinates and (6) gene of interest. Arrows represent spliced reads. Alternatively spliced exon is framed in red.
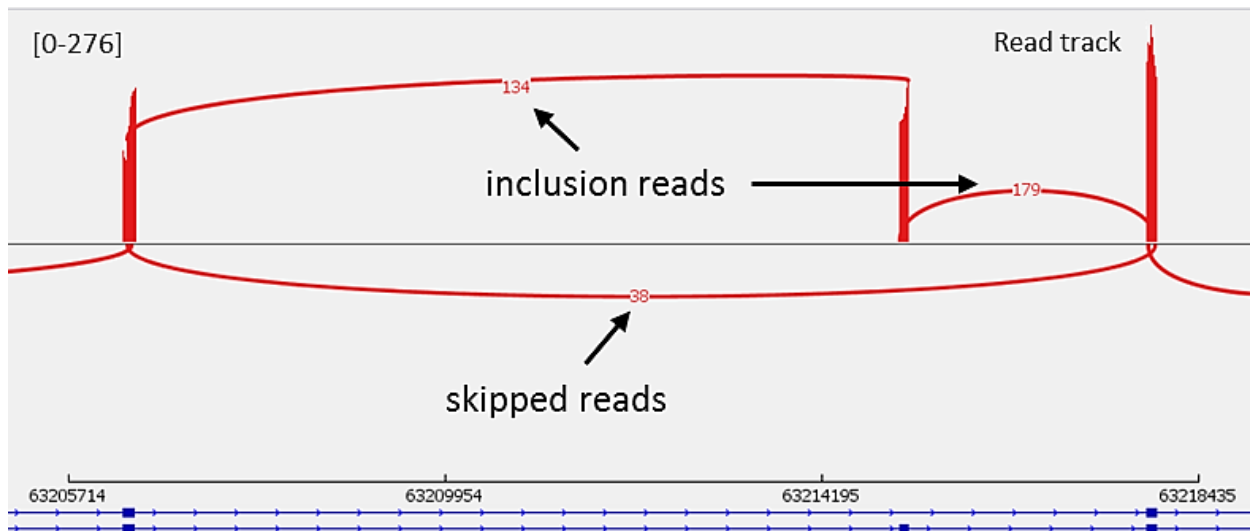
Figure 29: Sashimi plot of a genomic region in EHBP1 gene . Reads coverage is represented as well as junction reads. For instance, there is 38 reads supporting the exon-skipping transcript.

RESULTS

At the beginning of my thesis, we decided to focus on the fusion protein EWS-FLI1, which started to be clearly implicated on alternative splicing regulation. One paper came out in early 2015 about analysis of EWS-FLI1 proteome and its implication on splicing. Using immunoprecipitation experiments followed by mass spectrometry, Selvanathan et al., have confirmed that EWS-FLI1 oncoprotein interacts with several components of the spliceosome machinery. In addition, they performed exon array analysis on multiple Ewing sarcoma cell lines to identify differential exon expression patterns altered by EWS-FLI1. They highlighted ten genes that undergo differential alternative splicing pattern upon EWS-FLI1 depletion. However, using RNA-seq data produced from the lab, we could not fully reproduce their results. On these ten genes, four (40%) did not show any expression (no reads coverage in bam files) in one of the two conditions (EWS-FLI1-expressing cells or EWS-FLI1-depleted cells). For instance, Selvanathan and colleagues have identified TERT gene as differentially spliced. In our data, we could not detect any reads on TERT gene in EWS-FLI1-depleted cells on three different cell lines. Three of them (30%) have clearly no significant differential expression upon EWS-FLI1 depletion and three of them (30%) were validated in our cohort. Altogether, this indicates the complexity to study alternative splicing and the difference between technologies such as exon array and RNA-seq, which can explain the observed discrepancy.

Nevertheless, we decided to tackle this question by identifying differential splicing events regulated by EWS-FLI1 using RNA-seq experiments on Ewing sarcoma cell line. In addition, deciphering the mechanism underlying this complex process may provide useful targets for therapeutic purpose. This work was part of a collaboration between Franck Dequiedt lab and Olivier Delattre lab. Our study combines biological aspects, computational approaches and mechanistic insights to characterize the impact of ERG proteins and therefore EWS-ETS fusions on splicing. Importantly, we investigated the functional impact of splicing induced by EWS-ETS fusions in Ewing sarcoma.

The manuscript below will be shortly submitted in Cancer Discovery journal.

For complete transparency, contributions were done as follows: Experiments on HeLa cell lines (K.G., C.O.G), Experiments on Ewing sarcoma cell lines (O.S., A.C., J.P., M.M.A., K.L.), Computational analysis (O.S.), Computational support (C.O.G., B.S., O.M., S.G.L.), Figures (O.S., K.G., B.S.), Writing original draft (O.S., M.D., O.D., F.D.), Supervision (M.D., O.D., F.D.), Funding Acquisition (M.D., O.D., F.D.).

A novel function of Erg transcription factors in alternative splicing is altered in Ewing's sarcoma fusions

Olivier Saulnier[1,2], Katia Guerdi[3,4], Alina Chakraborty[5,6], Marie-Ming Aynaud[1], Joséphine Pineau[1], Tina O'Grady[3,4], Benjamin Sadacca[7,8], Karine Laud[1], Xavier Rambout[3,4], Olivier Mirabeau[1], Sandrine Lalami-Grossetete[1], Martin Dutertre[5,6,9,*], Olivier Delattre[1,2,9,*] and Franck Dequiedt[3,4,9,*]

[1]Institut Curie Research Center, SIREDO Oncology Center, Paris-Sciences-Lettres Research University, INSERM U830, Laboratory of Biology and Genetics of Cancers, Paris, France; [2]Université Paris Diderot, Sorbonne Paris Cité, France; [3]University of Liège, Interdisciplinary Cluster for Applied Genoproteomics, Liège, Belgium; [4]University of Liège, GIGA-Molecular Biology in Diseases, Liège, Belgium; [5]Institut Curie, PSL Research University, CNRS UMR3348, Orsay, France; [6]Université Paris Sud, Université Paris-Saclay, CNRS UMR3348, Orsay, France; [7]RT2Lab Team, Translational Research Department, Institut Curie, PSL Research University, Paris, France; [8]Institut de Mathématiques de Toulouse; UMR5219 Université de Toulouse; CNRS UPS IMT, Toulouse, France
[9]These authors contributed equally
[*]Correspondence: martin.dutertre@curie.fr (M.D.), olivier.delattre@curie.fr (O.D.), fdequiedt@uliege.be (F.D.)

ABSTRACT

Oncogenic alterations of *ERG* transcription factor family genes are observed in a variety of cancers including Ewing sarcomas and prostate cancers, and are thought to act through aberrant transcription regulation. Here, we show that ERG and FLI1 interact with spliceosomal proteins and are associated with nascent RNA. Moreover, ERG can induce alternative splicing of a reporter minigene. Analyses of RNA-seq data further show that upstream sequences of ERG-regulated exons are enriched in RBFOX2 binding sites. ERG interacts with RBFOX2 via its C-terminal domain and they both collaborate to regulate a common set of splicing events. EWS-FLI1 which contains this C-ter domain also binds RBFOX2 but, contrarily to wild-type ERG proteins, mostly opposes RBFOX2 splicing function by inhibiting its binding to its target pre-mRNAs. EWS-FLI1 hence alters the epithelial to mesenchymal transition splicing program regulated by RBFOX2. Finally, we show that EWS-FLI1-induced *ADD3* splicing contributes to the phenotype of Ewing cells.

SIGNIFICANCE

In addition to their role as transcription factors, we show that ERG proteins regulate the splicing landscape of cells through interaction with the master splicing regulator RBFOX2. Interference with splicing programs accounts for important oncogenic effects of EWSR1-FLI1, the oncoprotein of Ewing sarcoma.

Keywords: ERG – alternative splicing – RBFOX2 – EWS-FLI1 – Ewing sarcoma

INTRODUCTION

ERG subfamily proteins (ERG, FLI1 and FEV) belong to the ETS transcription factor family, one of the largest families of transcription factors in metazoans which is defined by a highly conserved DNA-binding ETS domain(Sharrocks, 2001). ERG proteins are involved in oncogenic fusion translocations in multiple cancers including Ewing's sarcoma (Zucman et al., 1993), a highly aggressive bone and soft tissue tumor. The primary oncogenic event of Ewing sarcoma is a reciprocal chromosomal translocation between genes of the *FET* family (*FUS*, *EWSR1* and *TAF15*) and *ERG* subfamily(Delattre et al., 1992). In particular, the EWSR1-FLI1 fusion has been reported to occur in approximately 90% of Ewing sarcoma cases(Delattre et al., 1992). Because it includes the C-terminal half of the FLI1 protein, which contains the ETS DNA-binding domain, fused to the low complexity region of EWSR1, the EWS-FLI1 fusion has mostly been studied as an oncogenic transcription factor. The EWS-FLI1 oncoprotein specifically binds to GGAA microsatellites and induces chromatin opening and activation of a large set of de novo enhancers(Boulay et al., 2017; Gangwal et al., 2008; Guillon et al., 2009). The neomorphic property of EWS-FLI1 requires the phase transition ability of wild-type EWS to hijack a chromatin remodeling machinery and to drive an aberrant transcriptional program(Boulay et al., 2017). In addition to its transcriptional regulatory activity, EWS-FLI1 has recently been shown to influence alternative splicing through interactions with core components of the spliceosome or by regulating the RNA polymerase II elongation rate(Knoop and Baker, 2000; Sanchez et al., 2008a; Selvanathan et al., 2015). To date, this function is mainly attributed to the EWS moiety. However, our latest observations led us to challenge this view. Indeed, we recently demonstrated that in addition to their role in transcription, ERG proteins also impact post-transcriptional processes such as mRNA stability(Rambout et al., 2016).

We found that ERG is recruited to specific transcripts via interaction of its C-terminal activation domain (CTAD) with RNA-binding proteins (RBPs). Once recruited to its target mRNAs, ERG recruits key components of the decay machinery to induce their degradation.

On these grounds, we hypothesized that ERG proteins might control other post-transcriptional processes, including splicing. In this study, we identified and characterized a new function of Erg proteins in alternative splicing of pre-messenger RNAs. We showed that the C-terminal domain of Erg proteins interacts with the splicing regulator RBFOX2 to regulate a common splicing program. Interestingly, EWS-FLI1 also interacts with RBFOX2 but inhibits its binding to pre-mRNA targets, and thus antagonizes a large-scale mesenchymal splicing program controlled by RBFOX2. Among these spicing targets, we identified a splicing isoform of ADD3 that represses the mesenchymal cell phenotype. This study reveals a new splicing regulatory role for ERG family proteins that is altered in FET-ETS oncogenic fusions, thereby contributing to splicing mis-regulation and cell phenotype changes in Ewing sarcoma.

RESULTS

ERG interacts with spliceosome components and controls splicing

To unravel potential post-transcriptional functions of ERG, we curated two protein-protein interaction databases(Snel et al., 2000; Stark et al., 2006) (BioGRID, https://thebiogrid.org and STRING, https://string-db.org) and identified 97 unique interactors for ERG **(Supplementary Table 1)**. The known association of ERG with transcriptional co-regulators and chromatin remodeling factors led us to anticipate enrichment of such proteins in this list. To our surprise, gene enrichment and prioritization analysis using the ToppGene suite revealed that a significant proportion of ERG partners (27%; 26 out of 97; FDR = 7.77E-22) are categorized as "*mRNA splicing, via spliceosome*" (GO:0000398) **(Figure 1A)**. Distribution of GO molecular function terms also highlighted a highly significant enrichment in proteins categorized as "RNA binding" (GO:0003723) **(Figure S1A)**. These proteins included core components of snRNP particules as well as snRNP-associated factors, spliceosome-associated hnRNPs and pre-mRNA processing factors. We confirmed interactions of ERG with SF3B2 and SF3B3, two core components of the SF3B complex by immunoblot analysis of endogenous ERG immunoprecipitation **(Figure 1B)**.

Furthermore, we demonstrated that ERG was also able to interact with core components of the spliceosome machinery, such as SF3A1 (another component of the SF3 complex), U1-70K (a core component of U1 snRNP) and U2AF65 (the larger subunit of U2AF) **(Figure 1B)**. Interestingly, most of these interactions were maintained when lysates were treated with RNAse, indicating that RNA is not necessary for the interactions tested. Altogether, these findings provide a framework to further investigate the potential function of ERG in splicing.

If ERG is a *bona fide* splicing regulator, it should be found in association with nascent RNA, because splicing is largely cotranscriptional(Kornblihtt, 2007). To test this, we prepared chromatin and other high molecular weight components (HMW) fraction from nuclei of HeLa cells. RNA- and DNA-associated proteins were then extracted from the insoluble HMW fraction by treating sequentially the pellet with RNAse and benzonase, according to a previously described protocol(Damianov et al., 2016). ERG could be extracted from the HMW fraction by the RNAse treatment, indicating that it was associated with nascent RNA **(Figure 1C)**. As expected, ERG was also found in the DNA-associated chromatin fraction, as it was solubilized with an additional benzonase treatment **(Figure 1C)**. Consistent with these findings, we found that FLI1, an another member of the ERG subfamily, was also present in both the DNA-associated and the RNA-associated fractions in HL-60 cells **(Figure S1B)**. Altogether, these data indicate that a proportion of ERG or FLI1 interacts with spliceosomal proteins and is associated with chromatin-bound RNA, supporting its role in pre-messengerRNA splicing.

To formally test this, we then used a splicing reporter assay in which an MS2 binding site was inserted downstream of exon 7 in a minigene that contains exons 6 to 8 of the *SMN2* gene(Sun et al., 2012). FLAG-tagged ERG proteins were expressed with an additional N-terminal tag derived from the MS2 bacteriophage coat protein (CP) to enable tethering to the minigene transcript(Rambout et al., 2016). Compared to MS2-CP alone, we observed that tethering of ERG, FLI1 or FEV significantly increased inclusion of exon 7 **(Figure 1D)**. Importantly, ERG proteins were unable to promote inclusion of exon 7 when not fused to MS2-CP, indicating that their ability to control exon inclusion is strictly dependent on their recruitment to the target pre-messengerRNA. We next asked which domain of ERG was responsible for this effect. We generated deletion mutants of ERG lacking either the PNT, ATAD, ETS or CTAD domain fused to MS2-CP, and tested them on the *SMN2* minigene splicing assay.

Among these constructs, the mutant lacking the CTAD domain was the only one showing a reduced splicing activity indicating that this particular domain, which is present in all ERG members, is important for its ability to promote exon inclusion (**Figure S1C**). In addition, because the mutant lacking the ETS domain is unable to bind DNA(Siddique et al., 1993), these results also suggest that the function of ERG in splicing is independent of its ability to activate transcription. Altogether, the above findings led us to investigate the ability of ERG to globally modulate the cellular splicing landscape.

We downregulated *ERG* expression in HeLa cells using a previously validated siRNA(Rambout et al., 2016) and analyzed splicing changes by RNA-seq. Among the ERG subfamily members, HeLa cells only express ERG, and, *ERG* mRNA levels was strongly decreased upon siERG transfection **(Figure 1E)**. Differential analysis of splicing events using the rMATS software(Shen et al., 2014) revealed splicing regulations upon ERG depletion that were very consistent across three biological replicates **(Figure 1F)**. The most frequent splicing event observed was the regulation of alternatively spliced exons (ASEs) with more than 400 ASEs being spliced-in (228 exons) or spliced-out (182) in control cells when compared to ERG-depleted cells **(Figure S1D; Supplementary table 2)**. GO term analysis revealed that ERG-regulated cassette exons were enriched in genes involved in "RNA processing" (GO: 0006396), "mRNA metabolic process" (GO: 0016071) and "cell cycle" (GO: 0007049) **(Figure S1E).**

In addition, we performed differential analysis of mRNA expression levels between control cells and ERG-depleted cells. We identified 2106 genes exhibiting changes in abundance in ERG-knocked-down cells, including 945 (45%) up- and 1160 (55%) down-regulated. Interestingly, we found no significant overlap between differentially spliced (DSG) and expressed (DEG) genes whether we looked at the gene or functional enrichment level **(Figure 1F, S1F, supplementary table 3).** Altogether**,** these observations suggest that ERG has separate roles in transcription and splicing, thereby regulating different sets of genes and potentially different cellular functions.

A

mRNA metabolic process (35)

mRNA splicing, via spliceosome (26)

RNA splicing (27)

DNA metabolic process (29)

negative regulation of gene expression (35)

chromosome organization (29)

-log10 (FDR)

B

Input
IP : CTL
IP: ERG
RNase A

WB : SF3A1

WB : SF3B3

WB : SF3B2

WB : U70-K

WB : U2AF65

WB : ERG

C

α - U2AF65

α - ERG

α - GAPDH

α - Histone H3

Total extracts · RNA-associated fraction · DNA-associated fraction

D

□ FLAG
■ FLAG-MS2

Relative percentage of EA7 inclusion

FLAG   FLI1   ERG   FEV

*   **   ***

FLAG   FLI1   ERG   FEV

MS2   - +   - +   - +   - +

E6 | E7 | E8

E6 | E8

E

ERG relative expression

siCTRL   siERG

***

F

siCTRL   siERG

1.5
0.5
0
-0.5
-1
-1.5

G

DEG   DSG

2059   47   375

n.s

75

Figure 1: ERG transcription factor interacts with spliceosomal proteins and controls splicing. (A) Distribution of enriched GO biological process terms from a list of curated protein interactors of ERG. Number of ERG-interactors in each GO are indicated in brackets (B) Western blot analysis of endogenous ERG immunoprecipitations from RNAse-treated or untreated lysates, using antibodies against the indicated spliceosome components (SF3A1, SF3B2, SF3B3, U70-K, UAF65). An anti-ERG antibody was used as control. (C) Immunoblot analysis of ERG in total, RNA- and DNA-associated HMW fractions from HeLa cells. U2AF65, GAPDH and Histone H3 specific antibodies were used as control for fraction purity (D) RT-PCR analysis of SMN2 minigene exon 7 inclusion. Samples are RNA from HeLa cells transfected with the SMN2-MS2 minigene reporter and either FLAG- or FLAG-MS2-CP-tagged version of FLI1, ERG and FEV. Results shown are means $\pm$ s.e.m (n=3 independent experiments) relative to FLAG-MS2-CP alone. *P<0.05; **P<0.01; ***P<0.001 by two-tailed unpaired Student's t test. (E) ERG mRNA expression levels following specific siERG treatment using normalized reads counts from RNA-seq experiments. Results shown are means $\pm$ s.e.m (n=3 independent biological replicates) relative to siCTRL experiment. *P<0.05; **P<0.01; ***P<0.001 by two-tailed unpaired Student's t test. (F) Heatmap of Z-scores of PSI values from differentially spliced exons between HeLa cells transfected either with siCTRL and HeLa siERG. (G) Overlap between differentially expressed genes (DEG) identified using htseq-count (FDR<0.05 and FC>2) and differentially spliced genes (DSG) identified by rMATS (with at least 15 reads supporting the event, FDR<0.05 and |ΔPSI>10%).

ERG and RBFOX2 control a common splicing program

To elucidate the specificity and mechanisms of splicing regulation by ERG, we performed motif enrichment analysis on ERG-regulated ASEs. Using a compilation of 110 known RNA-binding proteins (RBPs) binding sites from the literature(Anderson et al., 2012; Ray et al., 2013), we scanned sequences 250bp upstream and downstream of ERG-regulated ASEs. Enrichment was scored relative to background exons generated from non-regulated exons. The GCAUG motif, specific for the RBFOX family stood out from this analysis as significantly enriched RBP upstream of ERG-spliced-out ASEs **(Figure 2A)**. RBFOX proteins, which include RBFOX1, RBFOX2 and RBFOX3, are master splicing regulators in various cell types(Kuroyanagi, 2009). Interestingly, RBFOX proteins exhibit a position-dependent effect on splicing, preferentially splicing-in or splicing-out exon when bound downstream or upstream of the ASE, respectively(Jangi et al., 2014; Sun et al., 2012; Yeo et al., 2009). Altogether, these observations suggest that ERG might collaborate with Rbfox proteins to control alternative splicing.

To explore this possibility, we examined the effects of knocking down *RBFOX2* on alternative splicing in HeLa cells. Transfection of a specific siRNA resulted in efficient depletion of *RBFOX2* mRNA expression levels **(Figure S2A)**. In agreement with previous studies(Jangi et al., 2014; Zhang et al., 2008), *RBFOX2* knockdown mostly resulted in the regulation of ASEs **(Figure S2B)**. In total, we identified nearly 400 exons that were differentially spliced, with 233 spliced-in and 166 spliced-out in control cells as compared to RBFOX2-depleted cells (**Supplementary table 4**). Comparison with our dataset of ERG-regulated ASE revealed that a highly significant fraction (one third, p<10E-50) of RBFOX2-regulated ASEs were also sensitive to ERG depletion **(Figure 2B)**. Strikingly, 96% of the ASEs that were altered following RBFOX2 depletion were similarly regulated after ERG depletion **(Figure 2C)**, as illustrated for *ATP5SL* and *ITGA6* gene **(Figure 2D)**. Altogether, our observations strongly suggest that ERG and RBFOX proteins cooperate to regulate an similar alternative splicing program within the same protein complex.

A

RBFOX2 RNA-map

ERG-regulated exons

B

p < 10^{E}-50

C

D

ATP5SL

ITGA6

PSI   0,35   0,82   0,74

PSI   0,61   0,32   0,12
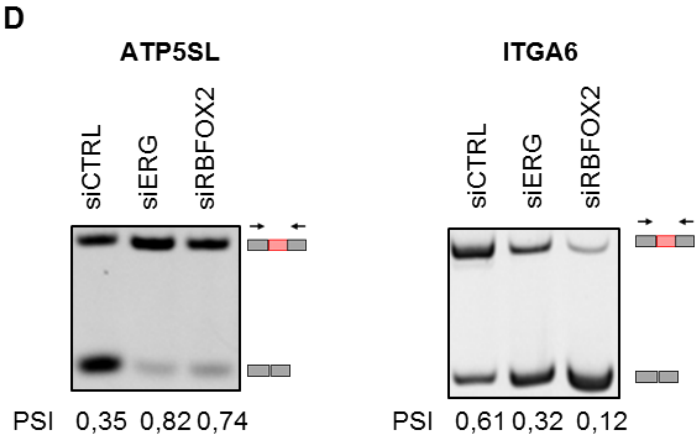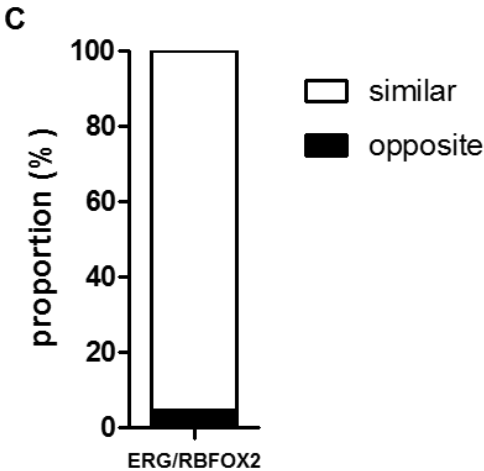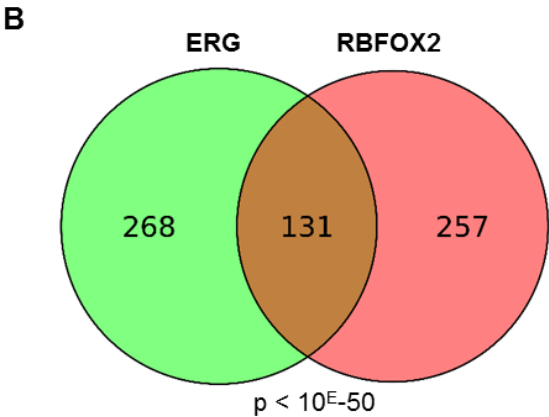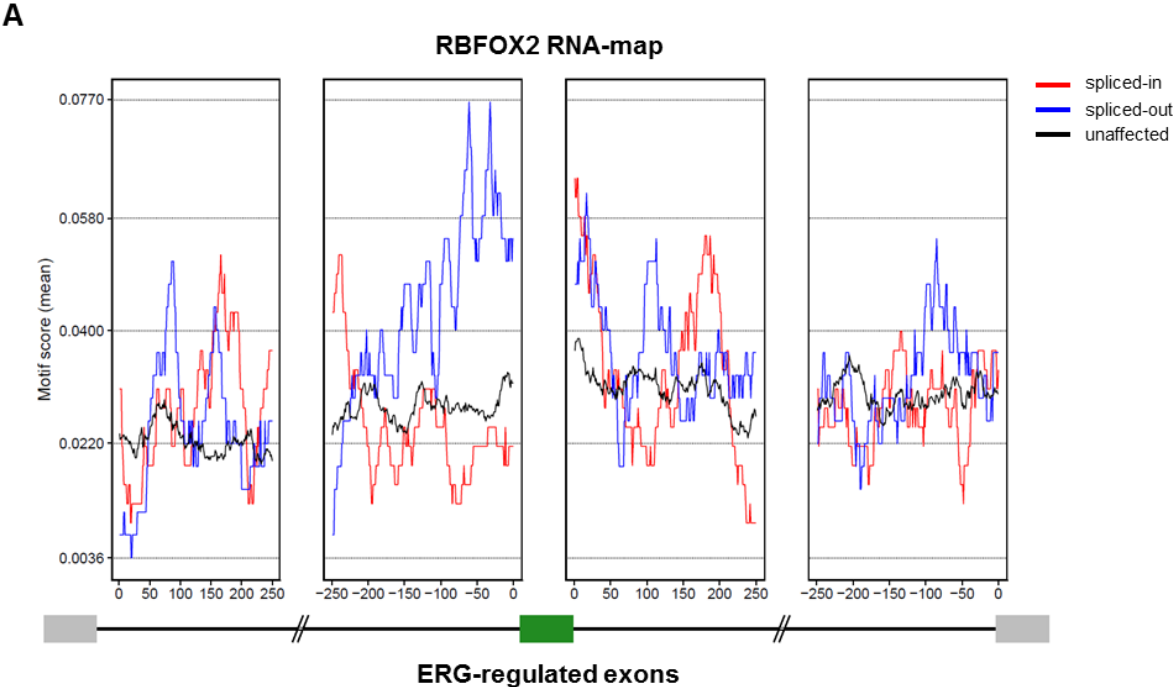
Figure 2: ERG regulates a RBFOX2-dependent splicing program. (A) RBFOX motif enrichment analysis upstream and downstream of ERG-regulated exons in HeLa cells. Red and blue lines represent intronic RBFOX motif scores around ERG spliced-in or spliced-out exons, respectively. Unaffected exons are generated from exons not modulated following ERG knockdown (|ΔPSI<5% and FDR>0.85) (B) Overlap between differentially spliced exons following ERG or RBFOX2 knockdown in HeLa cells. (C) Proportion of common target exons shown in (B)(n=131) categorized as "similar" or "opposite" depending on whether delta PSI values vary in respectively the same or opposite direction in siERG and siRBFOX2 datasets (D) RT-PCR analysis of two representative exons similarly spliced by ERG and RBFOX2. Samples are RNA from HeLa cells transfected by control siRNA or specific ERG or RBFOX2 siRNAs.

ERG and RBFOX2 collaborate on splicing through protein interaction

To test whether ERG and RBFOX proteins may interact with each other, we first conducted co-immunoprecipitation experiments. We expressed FLAG-tagged Erg subfamily proteins together with HA-tagged RBFOX1 or RBFOX2 in HeLa cells. Anti-FLAG immunoprecipitation followed by western blot analysis revealed that RBFOX1 **(Figure S3A)** and RBFOX2 **(Figure 3A)** strongly interact with ERG, FLI1 and FEV. Next, we looked at the interaction between endogenous proteins in HeLa cells. We observed that RBFOX2, the only member expressed in these cells (**Figure S3B**), readily co-immunoprecipitates with endogenous ERG, in a RNA-independent manner (**Figure 3B**). Using FLAG-tagged ERG mutants lacking individual domains, we examined the RBFOX2 region interacting with ERG. Co-immunoprecipitation experiments showed that the mutants lacking the PNT (pointed), ATAD (Amino terminal transcriptional activation) or ETS (DNA binding) domains still interacted with RBFOX2. In contrast, deletion of the CTAD (Carboxy terminal activation) region abolished interaction with RBFOX2 **(Figure 3C)**. Protein interaction assays based on complementation of the *Gaussia* luciferase (gPCA)(Remy and Michnick, 2006) confirmed that the CTAD is required for interaction with RBFOX2 and demonstrated that both proteins are in close proximity *in vivo* **(Figure S3C)**. Finally, we mapped the region in RBFOX1 **(Figure S3D)** and RBFOX2 **(Figure 3D)** that interacts with ERG by using their N-terminal, C-terminal of RRM domains in a GST pull-down experiment. This approach highlighted that the C-terminus of RBFOX1/2, which is required for their splicing function(Sun et al., 2012; Ying et al., 2017), mediated the interaction with ERG. Altogether, these data demonstrated a physical interaction between ERG and RBFOX2. Interestingly, the CTAD domain of ERG, which is important to induce exon inclusion in the tethering assay **(Figure S1C)**, is also the RBFOX2-interacting domain, suggesting that the interaction between ERG and RBFOX2 may mediate their collaboration to control splicing.
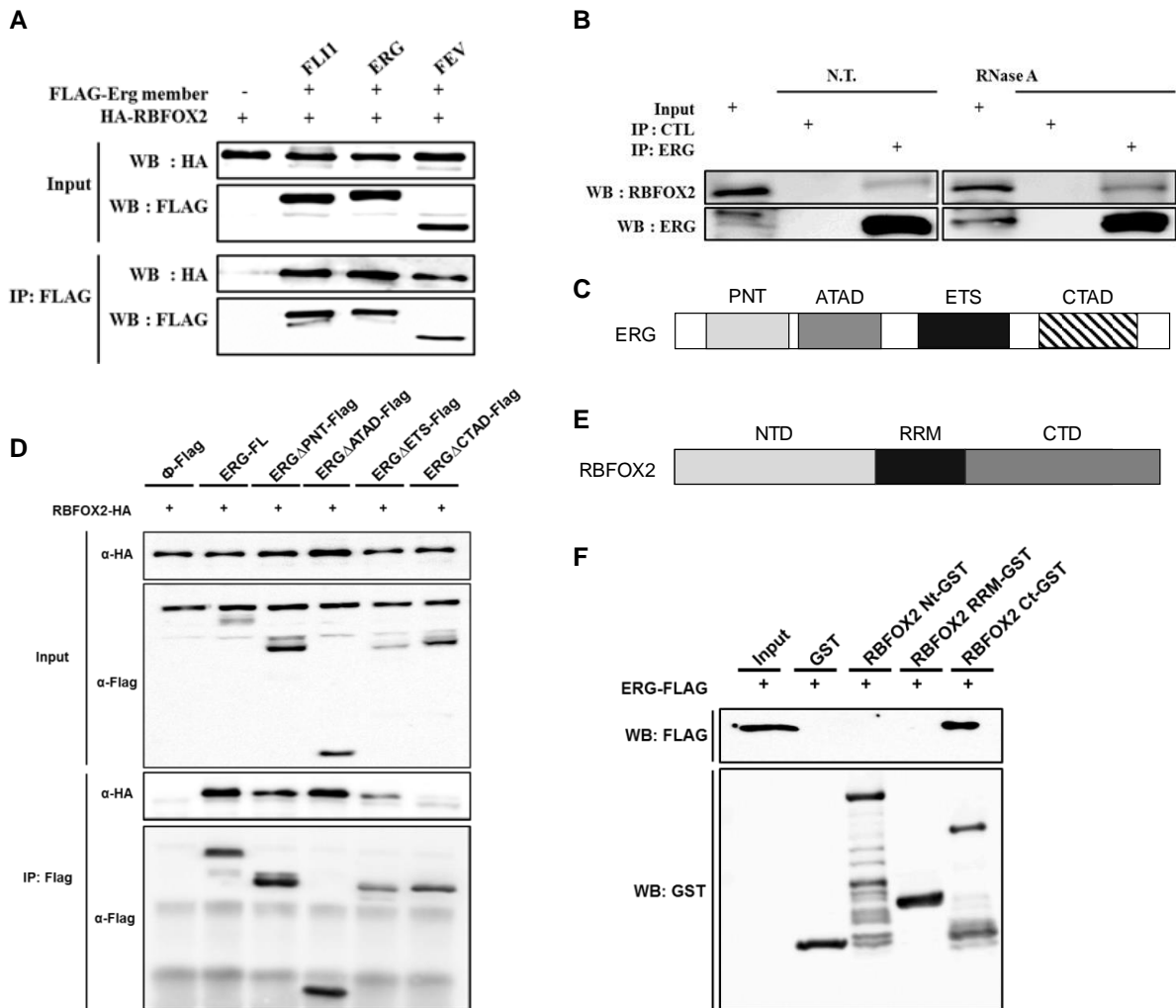
Figure 3: ERG physically interacts with RBFOX2 through its C-terminal domain. (A) Immunoprecipitation of FLAG-tagged Erg members and anti-FLAG and anti-HA western blotting. Samples are lysates from HEK-293 cells transfected with HA-RBFOX2 alone or with either of the FLAG-tagged ERG members. (B) Immunoprecipitation of ERG followed by anti-ERG and anti-RBFOX2 western blotting. Samples are control (CTL) and anti-ERG immunoprecipitates from HeLa cell lysates treated (+) or not (-) with RNAse-A. (C) Schematic ERG domains. (D) Immunoprecipitation of FLAG-tagged ERG deletion mutants and anti-FLAG and anti-HA western blotting. Samples are lysates from HEK-293 cells transfected with HA-RBFOX2 alone or with either of the FLAG-tagged ERG mutants. (E) Schematic RBFOX2 domains. (F) GST pull down assays using GST alone or GST-tagged RBFOX2 domains (Nt: N-terminus; RRM or Ct: C-terminus) and anti-FLAG and anti-GST western blotting. Samples are lysates from HEK-293 cells transfected with FLAG-ERG.

Identification of splicing alterations induced by EWS-FLI1 in Ewing sarcoma

The C-terminal domain of ERG family proteins, which mediates ERG interaction with RBFOX2 and is required for its splicing activity, is present in Ewing sarcoma-associated FET-ETS fusions, such as EWS-FLI1. While EWS-FLI1 is known to affect a large-scale splicing program in Ewing sarcoma cells(Selvanathan et al., 2015), its function in splicing is only attributed to its EWS moiety(Sanchez et al., 2008a). Our results suggest that the FLI1 moiety might also contribute to the EWS-FLI1 splicing function. First, we tested whether EWS-FLI1 is indeed able to interact with RBFOX2. Overexpression of FLAG-tagged versions of EWS, FLI1 or EWS-FLI1 together with a HA-tagged RBFOX2 in HeLa cells followed by anti-FLAG immunoprecipitation revealed a strong interaction of RBFOX2 with both FLI1 and EWS-FLI1 **(Figure 4A)**. Notably, only a much weaker but reproducible interaction was observed between the wild-type EWS protein and RBFOX2. The demonstration of an interaction between EWS-FLI1 and RBFOX2 prompted us to further investigate the potential role of RBFOX2 in EWS-FLI1-mediated splicing regulation.

We used the previously described Ewing sarcoma cell line ASP14(Carrillo et al., 2007) that harbors a doxycycline-inducible shRNA targeting the *EWS-FLI1* fusion transcript. RNA-seq experiments were performed after seven days of doxycycline (DOX) treatment (D7), when EWS-FLI1 was depleted, as compared to a non-depleted condition (D0) **(Figure 4B)**. We identified a global splicing alteration pattern upon EWS-FLI1 depletion that was consistent in three biological replicates **(Figure 4C)**. As previously observed for ERG, the most frequent splicing events was ASEs **(Figure S4A)** with more than 1300 exons regulated upon EWS-FLI1 depletion and a majority of them were spliced-out in presence of EWS-FLI1 (57%, 781/1360) **(Supplementary table 5)**. Our RNA-seq analysis was validated by RT-PCR analysis of a randomly chosen subset of splicing events, which showed a strong correlation of regulations (delta PSI, percent of spliced-in) between both methods **(Figure 4D)**.

We also conducted rescue experiments after washing the DOX from the media at day 7, therefore allowing EWS-FLI1 re-expression. RNA-seq experiments on intermediate (day 10) and final (day 22) time points were performed, for comparison of differentially spliced exons with EWS-FLI1-depleted cells (day 7) using rMATS **(Figure S4B)**. Remarkably, most ASE, being either spliced-in (ΔPSI>0) or splice-out (ΔPSI<0) in presence of EWS-FLI1, were reverted to their basal level after re-expression of EWS-FLI1 **(Figure 4E)**.

GO term enrichment analysis on EWS-FLI1 ASEs highlighted an enrichment in GO biological function terms "regulation of mRNA splicing, via spliceosome" (GO:0048024), "microtubule-

based process" (GO:0007017), and "cytoskeleton organization" (GO:0007010) **(Figure S4C).**
This observation suggested a different function of ERG and EWS-FLI1 on alternative splicing.
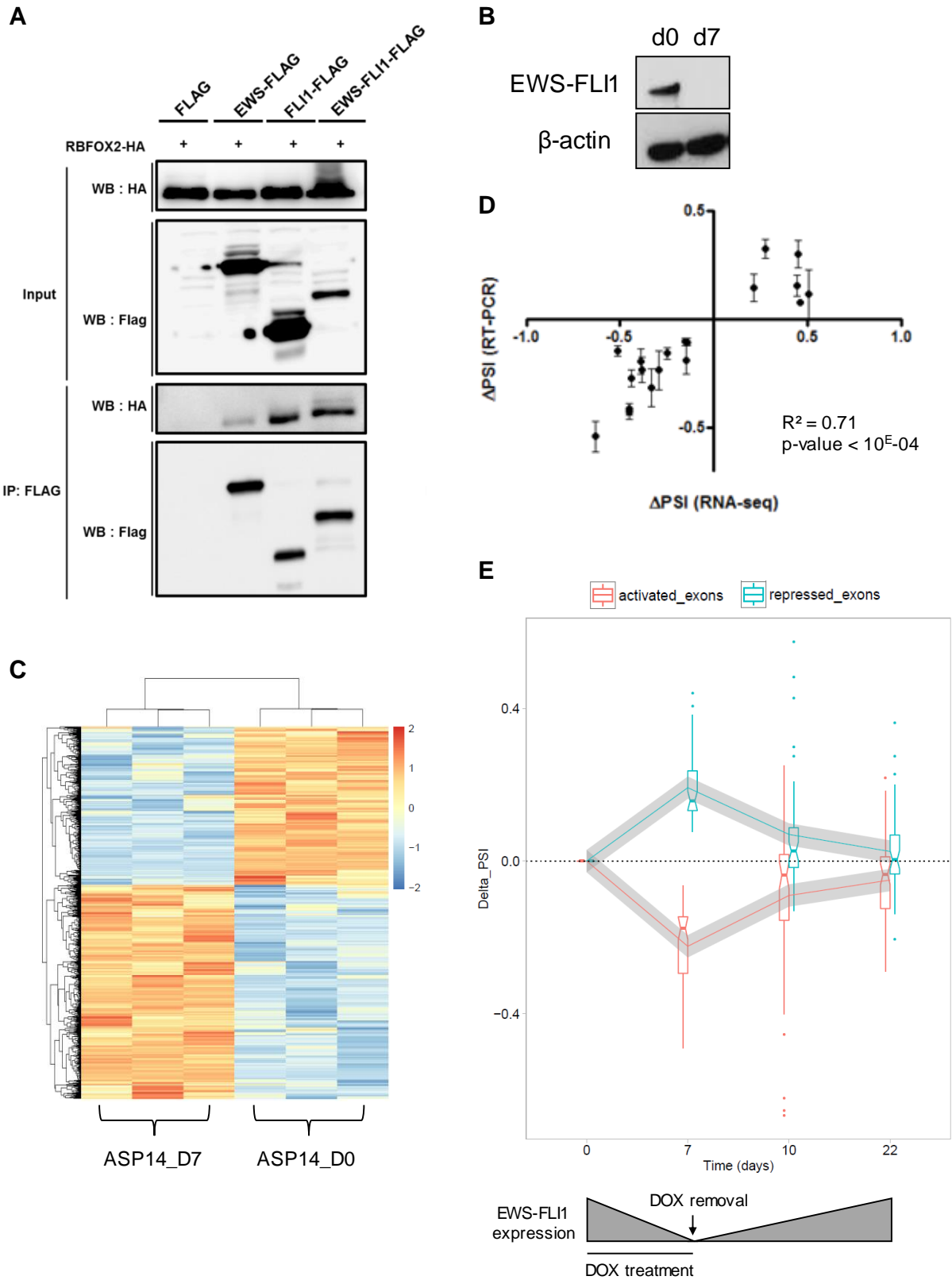
Figure 4: Transcriptome-wide analysis of splicing changes following knockdown of EWS-FLI1. (A) Immunoprecipitation of FLAG-tagged EWS, FLI1 or EWS-FLI1 and anti-Flag and anti-HA western blotting. Samples are lysates from HEK-293 cells transfected with HA-RBFOX2 together with the FLAG empty vector or with FLAG-tagged EWS, FLI1 or EWS-FLI1. (B) Western blotting of EWS-FLI1 in ASP14 cells at day0 or day7 of doxycycline treatment. Actin was used as loading control. (C) Heatmap of Z-scores of PSI values from differentially-spliced exons between ASP14 cells at day0 or day7 of doxycycline treatment. (D) Correlation of ΔPSI values analyzed by RT-qPCR or RNA-seq from a subset of exons showing differential splicing between ASP14 cells at day0 and day7 of doxycycline treatment. (E) Effects of EWS-FLI1 re-expression on EWS-FLI1-dependent splicing events. Delta PSI values of differentially spliced genes at day0 and day7 of doxycycline treatment in ASP14 cells. Doxycycline was removed at day7 and ΔPSI of EWS-FLI1-dependent splicing were analyzed after 3 (day10) and 14 days (day22) after doxycycline removal. Alternatively spliced exons were recovered to their basal splicing pattern after EWS-FLI1 rescue expression (day17). Splicing analysis was performed with rMATS using day7 as reference data (EWS-FLI1-depleted cells).

EWS-FLI1 and RBFOX2 have opposite roles on splicing

Next, we searched for enriched RBP motifs in upstream and downstream sequences of EWS-FLI1-regulated ASE. This analysis showed a significant enrichment of the RBFOX-binding motif **(Figure 5A)**, which highlighted the significance of our findings on ERG and RBFOX2 in HeLa cells. To our surprise, the RBFOX motif was enriched upstream of EWS-FLI1 spliced-in exons, instead of spliced-out exons in the case of ERG **(Figure 2A)**. These observations suggested that EWS-FLI1 and RBFOX2 may functionally interact in an antagonistic manner rather than in a collaborative manner, as observed for ERG and RBFOX2.

To further test this hypothesis, we analyzed splicing alterations following RBFOX2 depletion in Ewing sarcoma cell line ASP14 **(Figure S5A)**. We detected, using rMATS, more than 700 splicing events regulated upon RBFOX2 depletion **(Supplemental table 6)** and as observed in HeLa cells, we found that the most frequent splicing event altered after RBFOX2 depletion was ASEs (74,0%, 568/768) **(Figure S5B)**. As expected, from our data above, we found a highly significant overlap between EWS-FLI1 and RBFOX2-regulated exons **(Figure 5B)**. However, in contrast to the similar splicing effect observed between ERG and RBFOX2 **(Figure 2C)**, we found that more than half of the common ASE between EWS-FLI1 and RBFOX2 were oppositely regulated **(Figure 5C)**, as illustrated and confirmed by RT-PCR analysis of two representative examples, *ADD3* and *MICAL3* gene **(Figure 5D)**. These observations thus indicate that although EWS-FLI1 and RBFOX2 control common ASE, they have opposite effects on most of these splicing targets.

To understand how EWS-FLI1 might antagonize RBFOX2-dependent splicing function, we first examined RBFOX2 expression in EWS-FLI1-expressing *versus* non-expressing ASP14 cells. We found that RBFOX2 protein expression level did not change following DOX-mediated downregulation of EWS-FLI1 (data not shown), thus excluding the trivial explanation that EWS-FLI1 might repress RBFOX2 expression. Next, we performed RNA immunoprecipitation experiments of endogenous RBFOX2 in the ASP14 cell line following inhibition of EWS-FLI1. Quantitative PCR on two common splicing targets of EWS-FLI1 and RBFOX2 revealed that binding of RBFOX2 to pre-mRNAs was strongly increased in EWS-FLI1-depleted cells **(Figure 5E)**. Collectively, our data suggest that EWS-FLI1 antagonizes RBFOX2 alternative splicing function on a subset of its targets by inhibiting RBFOX2 binding to these targets.

A

RBFOX2 RNA-map

EWS-FLI1-regulated exons

B

EWS-FLI1

RBFOX2

1074   218   313

$p < 10^{E}-50$

C

similar

opposite

EWS-FLI1/RBFOX2

D

	EWS-FLI1		RBFOX2	
	+	−	+	−
ADD3				
PSI	0,29	0,05	0,17	0,68
MICAL3				
PSI	0,49	0,94	0,45	0,01

E

ASP14
ASP14+DOX

MICAL3   ADD3
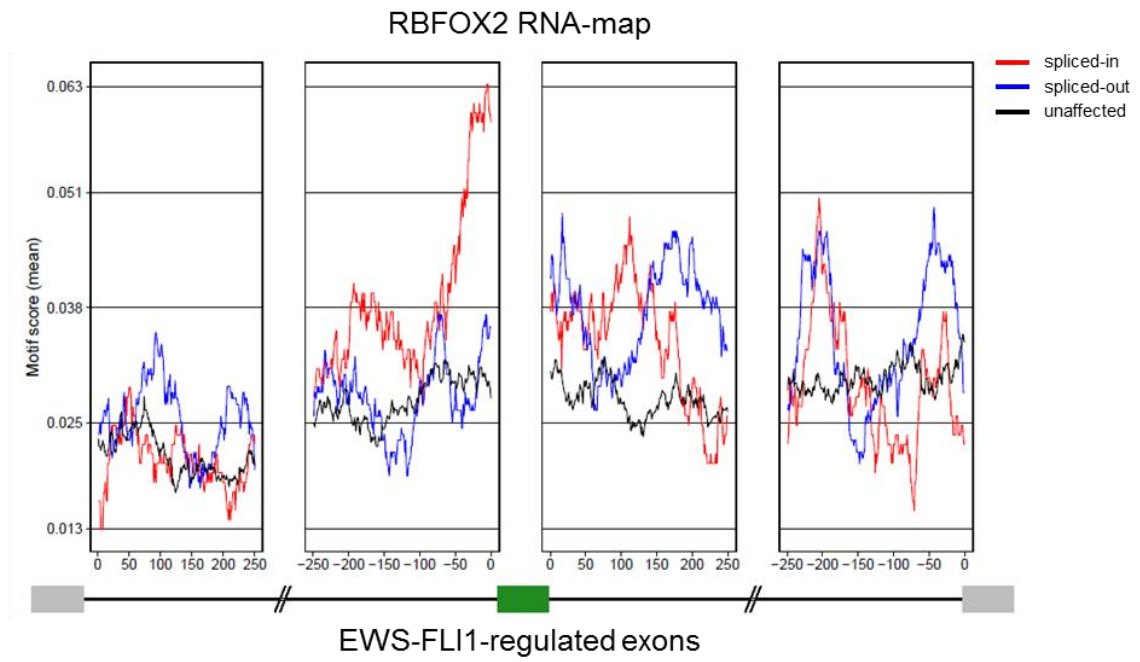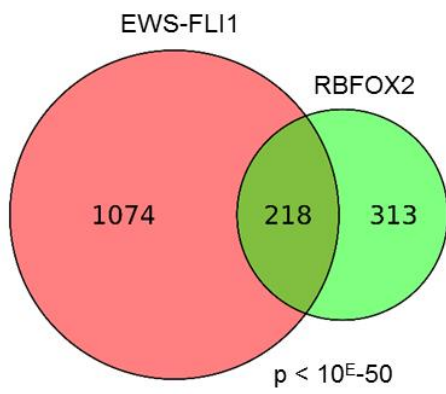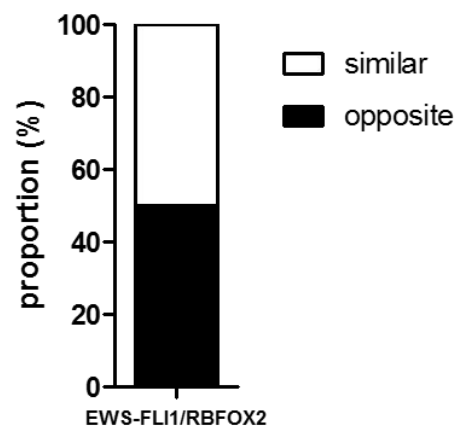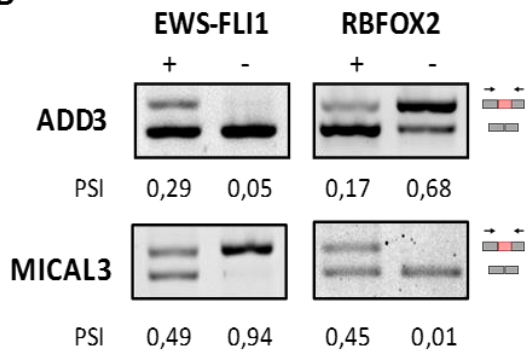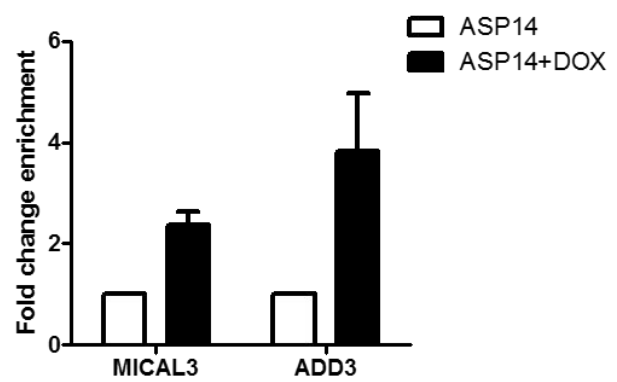
Figure 5: Antagonistic effects of EWS-FLI1 and RBFOX2 depletion on splicing. (A) RBFOX motif enrichment analysis upstream and downstream of EWS-FLI1-regulated exons in ASP14 cells. Red and blue lines represent intronic RBFOX motif scores around EWS-FLI1 spliced-in or spliced-out exons, respectively. Unaffected exons are generated from exons not modulated following ERG knockdown (|ΔPSI<5% and FDR>0.85) (B) Overlap between differentially spliced exons following EWS-FLI1 or RBFOX2 knockdown in ASP14 cells. (D) Proportion of common target exons shown in (B)(n=218) categorized as "similar" or "opposite" depending on whether ΔPSI values vary in respectively the same or opposite direction in EWS-FLI1 and RBFOX2 knockdown ASP14 cells. (E) RT-PCR analysis of two representative exons oppositely spliced by EWS-Fli1 and RBFOX2. Samples are RNA from ASP14 cells treated for 7 days with doxycycline to downregulate EWS-FLI1 expression or with specific RBFOX2 siRNA. (F) RNA immunoprecipitation experiments using anti-RBFOX2 antibodies followed by qPCR to detect MICAL3 and ADD3 transcripts. Samples are RNA from EWS-FLI1 expressing (ASP14) or depleted (ASP14+DOX) ASP14 cells.

ADD3 splicing induced by EWS-FLI1 is a phenotypic-driver in Ewing sarcoma

To address the functional relevance of these findings, we previously performed functional analysis on EWS-FLI1 splicing targets and observed a significant enrichment in cytoskeleton-related genes **(Figure S4C)**. Interestingly, RBFOX2 has been shown to be a regulator of epithelial to mesenchymal transition (EMT)(Braeutigam et al., 2014; Pradella et al., 2017; Shapiro et al., 2011), a process that plays a key role in Ewing sarcoma biology(Chaturvedi et al., 2012, 2014; Franzetti et al., 2017; Wiles et al., 2013). We found a highly significant overlap between EWS-FLI1-regulated ASE, RBFOX2-regulated ASE and EMT-regulated ASEs (the latter is a public siplicing EMT signature(Yang et al., 2016)) **(Figure 6A)**. This observation suggests that alternative splicing is an additional layer of EMT-regulation in Ewing sarcoma. To establish the functional relevance of EWS-FLI1 antagonism with RBFOX2 splicing regulation, we turned our attention towards the *ADD3* gene. ADD3 is an EMT-linked protein that is known to play a role in actin cytoskeleton remodeling(Kiang and Leung, 2018). In addition, exon 14 of *ADD3* is antagonistically regulated by EWS-FLI1 and RBFOX2, *i.e.* EWS-FLI1 spliced-in exon 14 whereas RBFOX2 spliced-out exon 14. We validated the regulation of *ADD3* exon 14 splicing by EWS-FLI1 depletion using either DOX treatment or siRNAs transfection in the Ewing sarcoma cell lines ASP14 and MHH-ES1, respectively **(Figure 6B)**. Specific depletion of *ADD3* exon 14 transcript in these two Ewing sarcoma cell lines **(Figure S6A)** was strongly efficient as we completely abolished *ADD3* exon 14-containing mRNA. Immunofluorescence staining of F-actin demonstrated a dramatic cytoskeleton remodeling with the presence of numerous actin stress fibers and increased cell area, characteristic of mesenchymal cells(Vallenius, 2013) **(Figure 6C, 6D)**. Concordant with these observations, cell invasion capacities were strongly increased after inhibition of exon 14 transcript in a three-dimensional type I collagen spheroid assay **(Figure 6E)**. Previous studies showed a similar switch to a mesenchymal phenotype following EWS-FLI1 depletion(Chaturvedi et al., 2012, 2014; Franzetti et al., 2017; Wiles et al., 2013). Thus, we conclude that exon 14-containing isoform of ADD3 plays a role in the repression of the mesenchymal phenotype in Ewing sarcoma cells. Importantly, depletion of the ADD3 exon 14-containing isoform was not associated with a reduction in *EWS-FLI1* expression (**Figure S6B**). In addition, RNA-seq analysis of a collection of Ewing tumors showed that the relative inclusion levels of *ADD3* exon 14 was significantly correlated with *EWS-FLI1* expression levels (p=0.017; R^2=13.1%). Finally, we performed k-means clustering algorithm to separate Ewing sarcoma tumors according to their relative inclusion levels of *ADD3* exon 14.

We observed that low levels of *ADD3* exon 14 inclusion in Ewing sarcoma tumors were significantly associated with poor survival **(Figure S6C)**. This observation indicate that tumors expressing low levels of ADD3 exon 14-containing isoform promote a EMT-like process with migratory and invasive advantages that may favor Ewing sarcoma metastatic spreading. Altogether, these data strongly suggest that the regulation of *ADD3* splicing by EWS-FLI1 via inhibition of RBFOX2 binding to the ADD3 pre-mRNA is important for Ewing sarcoma biology.
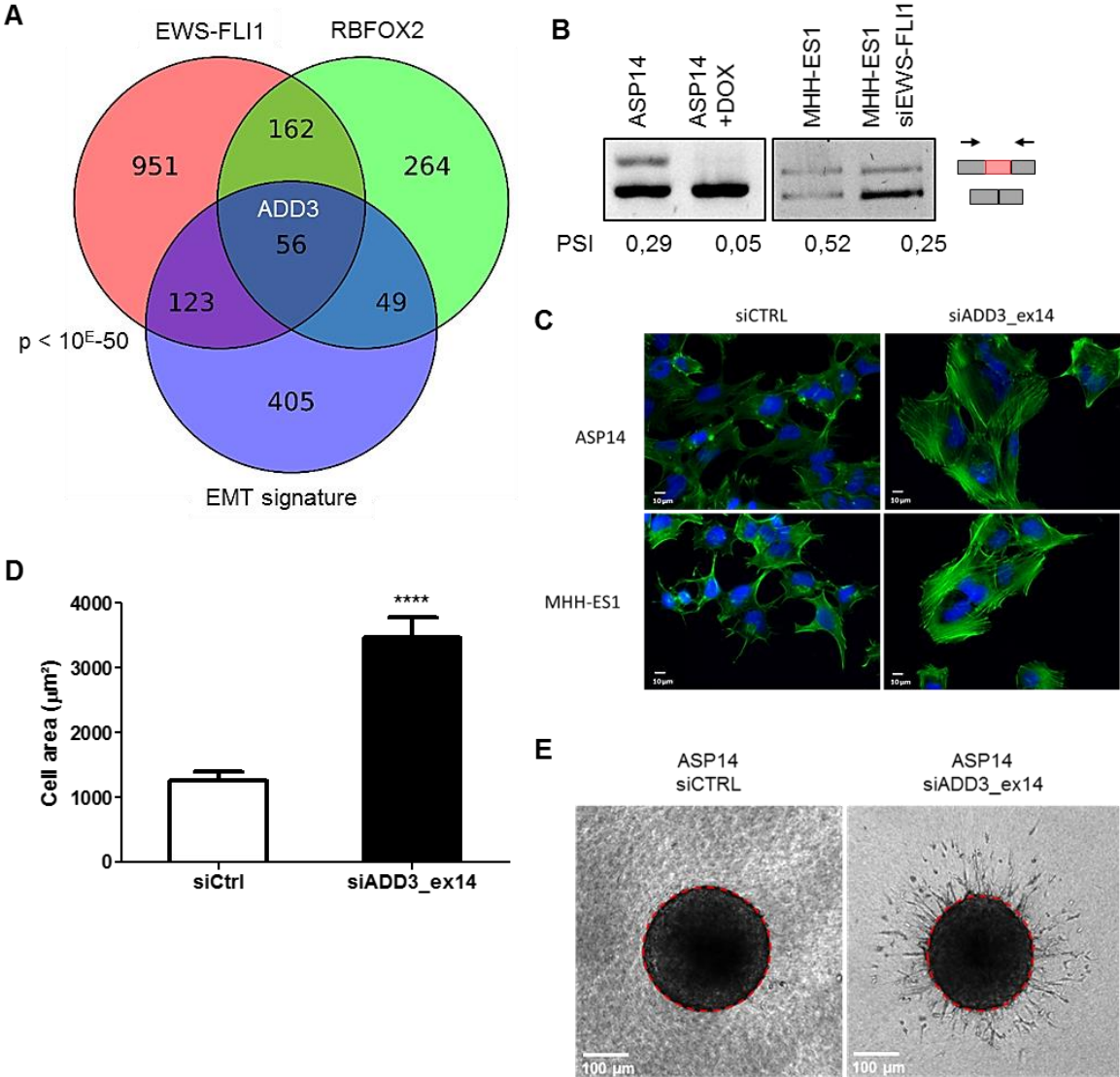
Figure 6: Splicing of ADD3 exon 14 isoform participates in the phenotype of Ewing sarcoma cells. (A) Overlap between EMT-associated splicing events in H358 human epithelial cells (Yang *et al.,* 2016) and EWS-FLI1- or RBFOX2-dependent exons in ASP14 cells. (B) RT-PCR analysis of ADD3 exon 14 splicing in two Ewing sarcoma cell lines after EWS-FLI1 depletion. Samples are RNA from ASP14 cells non treated (ASP14) or treated with doxycycline for 7 days or MHH-ES1 cells transfected with a control siRNA or a specific EWS-FLI1 siRNA. (C) Immunofluorescence of actin filaments stained with phalloidin (green channel) and DAPI (blue channel) of ASP14 and MHH-ES1 Ewing sarcoma cell lines treated with either siCTRL or a siRNA against the exon14-containing ADD3 isoform (siADD3_ex14). (D) Measurement of cell area of ASP14 cells treated with either siCTRL or a siRNA against the exon14-containing ADD3 isoform (siADD3_ex14). Data are represented as mean +/− s.e.m. ****P<0.0001. (E) Three dimensional type-I collagen multicellular spheroid invasion assay of ASP14 cell line treated with siCTRL or a siRNA against the exon14-containing ADD3 isoform (siADD3_ex14). Red dotted lines represent the initial spheroid perimeter.

DISCUSSION

Alternative splicing is often dysregulated in cancer and might be considered as a potential oncogenic driver event37. Several direct and indirect mechanisms have been linked to splicing regulation by transcription factors, including the regulation of the RNA polymerase II elongation rate or the regulation of splicing factors mRNA levels7,38. The present data are consistent with emerging evidences showing that proteins lacking canonical RNA binding domains, including transcription factors, also regulate splicing outcomes through interactions, whether with RNA or with RNA binding proteins39,40.

In this study, we reported that ERG subfamily proteins control the cellular splicing landscape by interacting with several core components of the spliceosome machinery as well as the master splicing regulator RBFOX2. We demonstrated that both, ERG and RBFOX2, control a common splicing program and collaborate on pre-mRNA via protein-interaction through the C-terminal domain of ERG. Because the ETS domain of ERG does not overlap the C-terminal domain and is required for its transcriptional activity16, these findings suggest that the splicing function of ERG is independent of its ability to regulate gene expression. Interestingly, the C-terminal domain of ERG is involved in several cancers as fusion partner gene, including Ewing sarcoma and prostate cancer3,41. Thus, we investigated the potential implication of EWS-ETS oncogenic fusions in this context.

Alternative splicing is a common EWS-ETS-mediated feature in Ewing sarcoma7,8. In this study, we describe an additional mechanism of alternative splicing regulation by EWS-ETS fusion proteins. We have shown that EWS-FLI1 also interacts with RBFOX2 and regulates a common splicing program. However, EWS-FLI1 regulated these pre-mRNA targets in an opposite manner as compared to RBFOX2. Using RNA immunoprecipitation experiments, we showed that RBFOX2 binding to its targets is increased after depletion of EWS-FLI1 in Ewing sarcoma cells. We hypothesized that EWS-FLI1 sequestrate RBFOX2 through its ability to form protein aggregates via the prion-like domain of EWS. Our results suggest that EWS-ETS oncoproteins alter the splicing landscape of Ewing sarcoma by interfering with RBFOX2 function. This observation is consistent with the previously described dominant negative function of EWS-ETS fusions on the activity of wild-type ETS proteins42.

Importantly, we demonstrated that EWS-FLI1 fusion inhibits RBFOX2 binding to ADD3 pre-mRNA to promote spliced-in of exon 14, which represses the mesenchymal cell phenotype. Furthermore we demonstrated that cell migratory and invasive capacities were increased following ADD3 exon 14 spliced-out. We observed that Ewing sarcoma tumors that spliced-out exon 14 of ADD3 were associated with poor survival, suggesting a potential impact of this isoform on Ewing sarcoma tumor invasion. Interestingly, it has already been described that spliced-out of ADD3 exon 14 is associated with metastasis in mice[43] and has been found overexpressed in cancer[44,45]. Structure prediction of the encoded peptide by spliced-in of exon 14 leads to a 32 amino acids sequence predicted for a low complexity region and a small coiled coil. This sequence is inserted right upstream the MARCKS domain, a functional domain required for ADD3 interaction with both spectrin and actin. Further experiments need to be performed to decode this function and especially on the spectrin-actin network assembly.

Altogether, the identification of ERG splicing function expands the spectrum of splicing abnormalities in cancer and provides further evidence that transcription factors affect splicing outcomes. Our study gives new relevant perspectives on ERG-rearranged proteins function in different cancers and on how these cancers might exhibit splicing alterations. In addition, we propose that splicing dysregulation is an important feature of Ewing sarcoma biology and participates in the cellular plasticity.

SUPPLEMENTAL FIGURES

Supplementary Figure S1
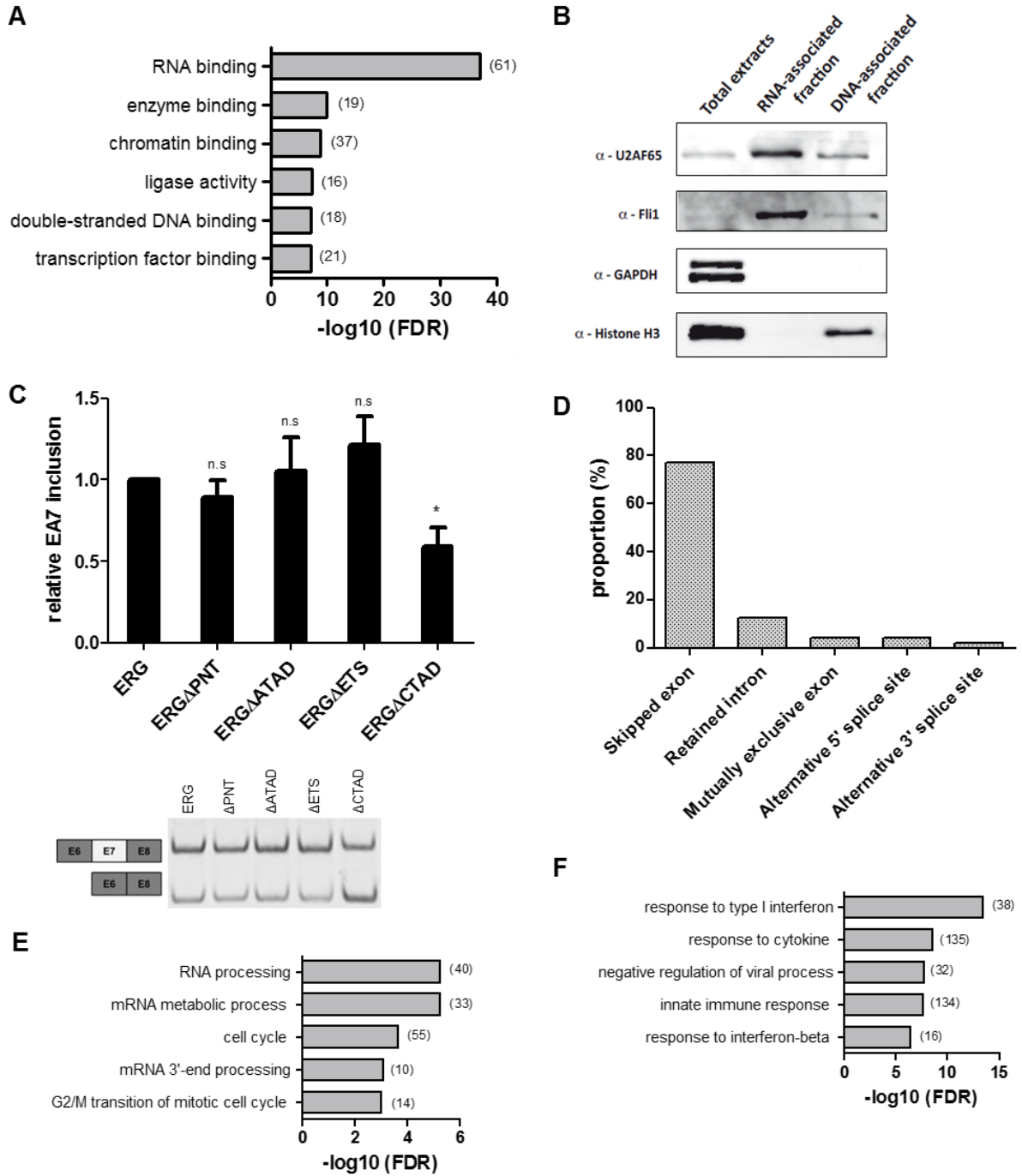
Supplementary Figure S2
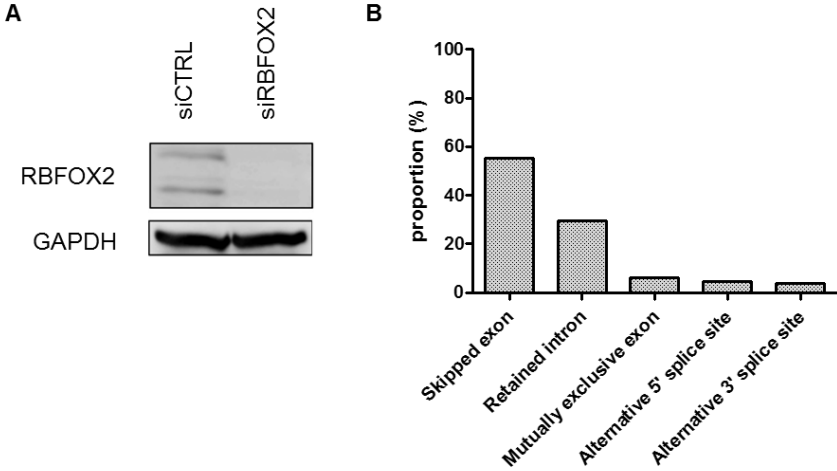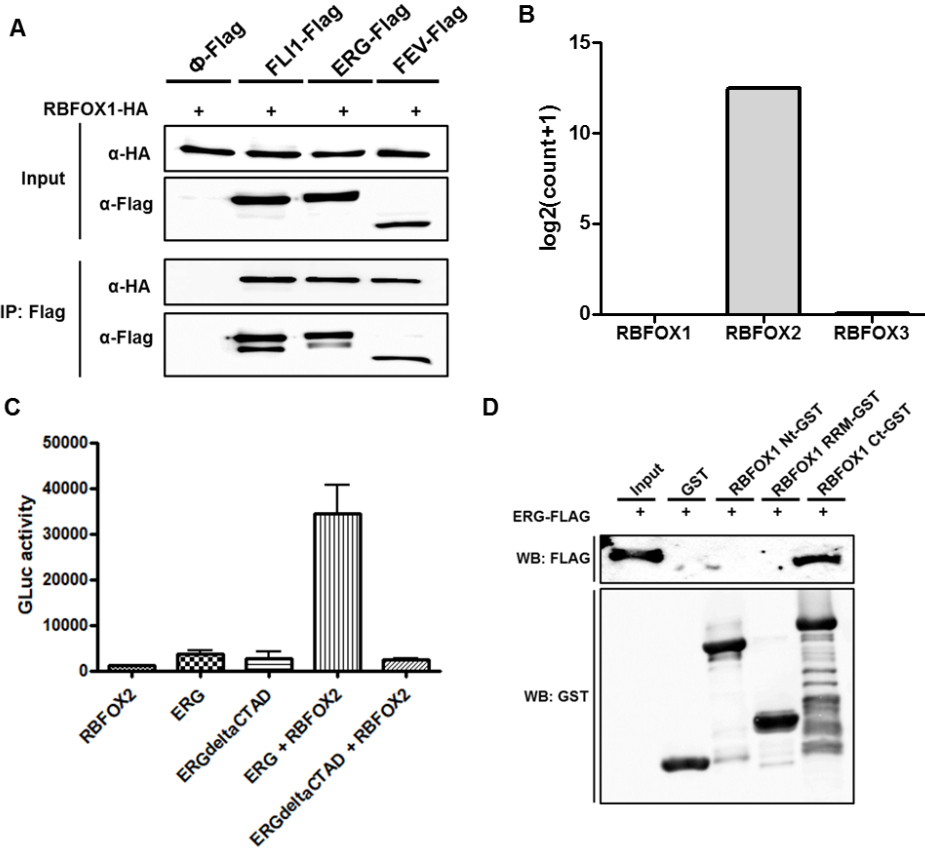
A

B

Supplementary Figure S3

A

B

C

D

Supplementary Figure S4

A

B

C

Supplementary Figure S5

A

B

Supplementary Figure S6

A

siADD3_ex14   -   +

ASP14

PSI   17,8   0

siADD3_ex14   -   +

MHH-ES1

PSI   18,2   0

B

C

ADD3 os

High exon inclusion (7)
Intermediate exon inclusion (12)
Low exon inclusion (24)

0,04   0,8   0,09
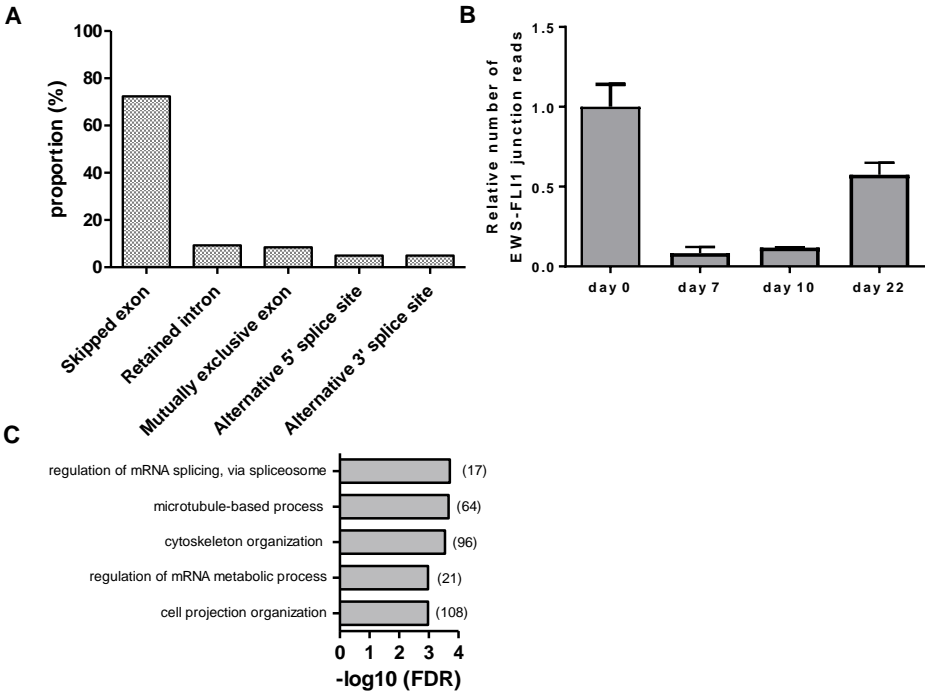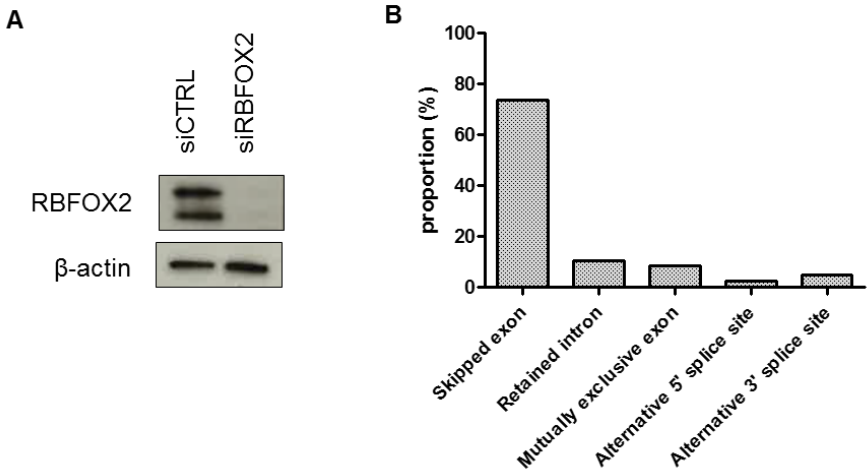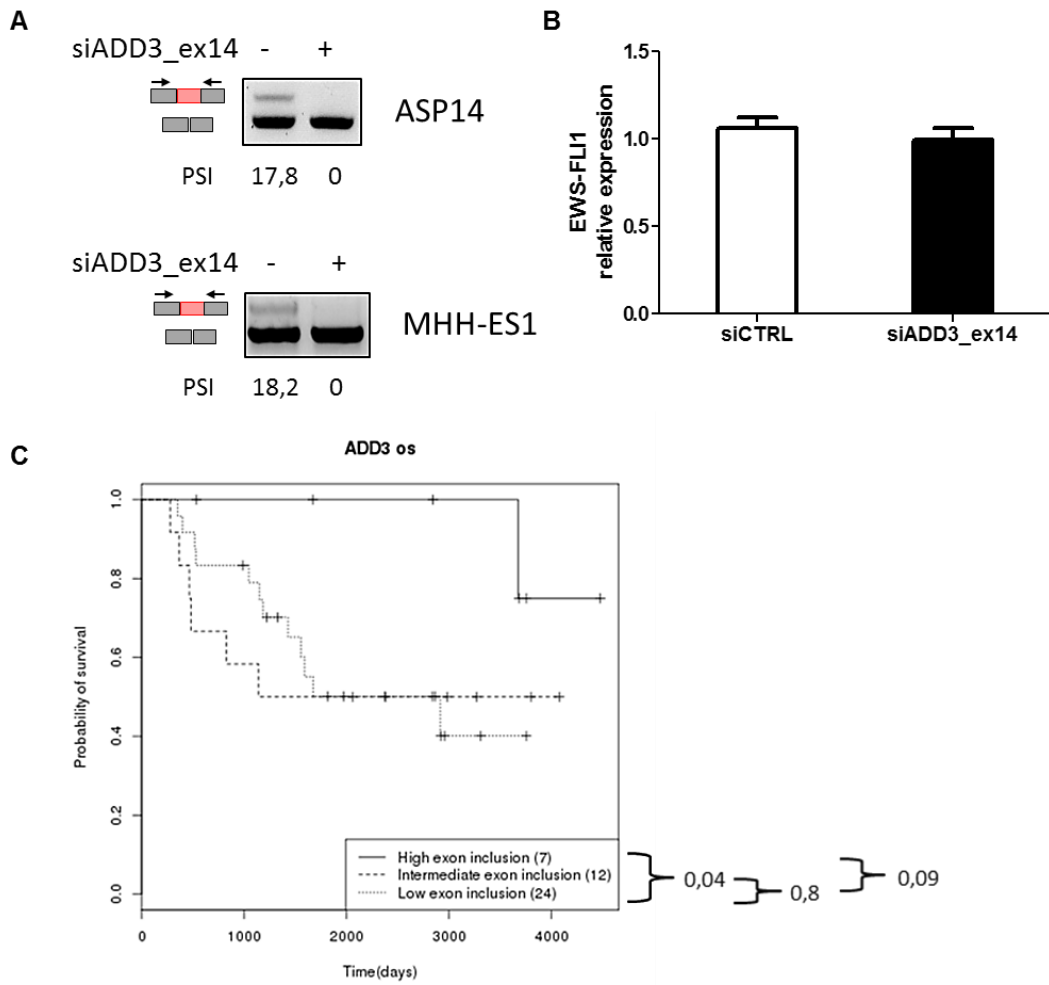
Supplementary Figure S1: ERG transcription factor interacts with spliceosomal proteins and controls splicing. (A) Distribution of the most significantly enriched GO molecular function terms from a list of curated proteins known as ERG-interactors (from the BioGRID and STRING databases)**.** Number of ERG-interactors in each GO are indicated in brackets (B) Immunoblot analysis of FLI1 in total, RNA- and DNA-associated HMW fractions from HL60 cells. U2AF65, GAPDH and Histone H3 specific antibodies were used as control for fraction purity. (C) RT-PCR analysis of SMN2 minigene exon 7 inclusion. Samples are RNA from HeLa cells transfected with the SMN2-MS2 minigene reporter and either FLAG- or FLAG-MS2-CP-tagged version of the indicated ERG deletion mutants. Results shown are means $\pm$ s.e.m (n=3 independent experiments) relative to FLAG-MS2-CP alone. *P<0.05; n.s: non-significant by two-tailed unpaired Student's t test. (C) Types and proportion of significant differential splicing events identified after siRNA-mediated ERG inhibition in HeLa cells. (E) Distribution of top significant GO biological function terms from genes containing ERG-regulated exons. Number of spliced genes regulated by ERG in each GO are indicated in brackets. (F) Distribution of top significant GO biological function terms from differentially expressed genes by ERG. Number of genes regulated by ERG in each GO are indicated in brackets.

Supplementary Figure S2: (A) Western blotting of RBFOX2 and GAPDH, as control. Samples are total lysates from HeLa cells transfected with control (siCTL) or specific RBFOX2 siRNA (siRBFOX2). (B) Types and proportion of RBFOX2-dependent splicing events in HeLa cells.

Supplementary Figure S3: Erg proteins physically interacts with RBFOX proteins (A) Immunoprecipitation of FLAG-tagged EWS, FLI1 or EWS-FLI1 and anti-Flag and anti-HA western blotting. Samples are lysates from HEK-293 cells transfected with HA-RBFOX1 together with the FLAG empty vector or with FLAG-tagged EWS, FLI1 or EWS-FLI1. (B) Expression levels of *RBFOX1, 2 and 3* mRNA in HeLa cells from RNA-seq analysis. (C) Protein interaction assay between RBFOX2 and ERG full-length or an ERG mutant lacking the CTAD domain, using the *gaussia princeps* luciferase complementation method. Results are means.± s.e.m from one representative experiment out of 3. (D) GST pull down assays using GST alone or GST-tagged RBFOX1 domains (Nt: N-terminus; RRM or Ct:C-terminus) and anti-FLAG and anti-GST western blotting. Samples are lysates from HEK-293 cells transfected with FLAG-ERG.

Supplementary Figure S4: Transcriptome-wide analysis of splicing changes following knockdown of EWS-FLI1. (A) Types and proportion of EWS-FLI1-dependent splicing events in ASP14 cells. (B) Relative number of EWS-FLI1 junction reads using STAR-Fusion on RNA-sequencing datasets. Samples are total lysates from ASP14 cells treated with doxycycline for the indicated time points. (C) Distribution of top significant GO biological function terms from genes containing EWS-FLI1-regulated exons. Number of spliced genes regulated by EWS-FLI1 in each GO are indicated in brackets.

Supplementary Figure S5: (A) Western blotting of RBFOX2 and GAPDH in ASP14 cells transfected with control (siCTL) or specific RBFOX2 siRNA (siRBFOX2). Actin was used as loading control. (B) Types and proportion of RBFOX2-dependent splicing events in ASP14 cells.

Supplementary Figure S6: Splicing of ADD3 exon 14 isoform participates in the phenotype of Ewing sarcoma cells (A) RT-PCR analysis of levels of exon14-containing ADD3 transcripts in ASP14 and MHH-ES1 Ewing sarcoma cell lines. Samples are RNA from ASP14 or MHH-ES1 cells transfected with a control siRNA (siCTL) or a siRNA specific of ADD3 exon14 (siADD3_ex14). (B) RT-qPCR analysis of EWS-FLI1 expression. Samples are RNA from ASP14 treated with control siRNA (siCTRL) or treated with a specific siRNA of ADD3 exon 14 (siADD3_ex14). (C) Kaplan-Meier curve of ADD3 exon 14 PSI values in 43 Ewing tumors showed significant differences between tumors expressing high or low ADD3 exon 14 transcript levels. Tumors were separated by k-means clustering. Number of tumors in each subgroup is indicated in brackets.

METHODS

Cell culture and transfection

Cell lines were obtained from American Type Culture Collection (ATCC) and were routinely checked by PCR for the absence of mycoplasma. Ewing sarcoma cell line ASP14 was generated as previously described (Carrillo et al., 2007). HeLa, HEK293T and ASP14 cell lines were cultured at 37°C, in 5% CO_2 with DMEM (Gibco) supplemented with 10% FBS (Eurobio) and 1% antibiotics (v/v) (penicillin and streptomycin (Gibco)).

Induction of EWSR1-FLI1 specific shRNA was performed by adding 1 µg/mL of doxycycline in the medium *ex-tempo*. After seven days of treatment, doxycycline was removed and cells were washed three times to stop the shRNA induction.

Transfection of siRNAs (see sequences Supplemental table 6) was performed using Lipofectamine RNAiMAX Reagent according to the manufacturer's instructions (Thermo Fisher Scientific) and collected 48 hours post-transfection.

Plasmids were transfected with the Lipofectamin 2000 Transfection System according to the manufacturers' protocol (Thermo Fisher Scientific) and collected 48 hours post-transfection.

HA-FLAG-tagged ERG inducible stable HeLa cell line was generated as follows: HeLa cells were transduced with pLenti CMV rtTA3 Hygro (Addgene) (10 MOI) and selected with 100 µg/ml Hygromycin B Gold (InvivoGen). Next, rtTA3 HeLa cells were transduced with pLenti CMVtight Blast expressing HA-FLAG-tagged ERG (50 MOI) and selected with 10 µg/ml blasticidin (InvivoGen) (Rambout et al., 2016).

Subcellular fractionation

HeLa cells were washed with cold PBS, harvested and lysed with CLB buffer (Cytoplasmic Lysis Buffer; 10mM (m/v) Tris HCl pH7.9, 340 mM (m/v) Sucrose, 3 mM (m/v) CaCl2, 0,1 mM (m/v) EDTA, 2 mM (m/v) MgCl2, 1 mM (m/v) DTT, 0.5% (v/v) NP40, cOmplete Protease Inhibitor Cocktail (Roche) and Halt Phosphatase Inhibitor Cocktail (Thermo Fisher Scientific)) on ice for 5 min. Cytoplasmic fraction was removed by centrifugation at 3,500x g for 15 min at 4°C. The pellet was washed several times with CLB wash buffer (CLB buffer without NP40) and lysed with NLB buffer (Nuclear Lysis Buffer ; 20 mM (m/v) HEPES pH7.9, 10 % (v/v) Glycerol, 3 mM (m/v) EDTA, 150 mM (m/v) KOAc, 1.5 mM (m/v) MgCl2, 1 mM (m/v) DTT, 0.1 % (v/v) NP40, cOmplete Protease Inhibitor Cocktail (Roche), Halt Phosphatase Inhibitor Cocktail and Protector RNase Inhibitor (Roche)). Soluble and HMW fractions were separated by centrifugation at 15,000x g for 30 min at 4°C. The pellet was washed several times with NLB washing buffer (NLB buffer without NP40), centrifuged at 3,500x g for 5 min and lysed with NIB buffer (Nuclease Incubation Buffer ; 150 mM (m/v) HEPES pH7,9, 10 % (v/v) Glycerol, 150 mM (m/v) KOAc, 1.5 mM (m/v) MgCl2, 1 mM (m/v) DTT,

cOmplete Protease Inhibitor Cocktail and Halt Phosphatase Inhibitor Cocktail). Next, the lysate was divided into two equal portions to treat one of them with RNase A (Thermo Fisher Scientific) for 30 min at RT. Then, lysates were centrifuged at 20,000x g for 30 min and the supernatants were recovered to form the RNA-associated chromatin fractions traited or not with RNAse A (Thermo Fisher Scientific). Finally, pellets were incubated at 25°C on a rotator with 5U/µl of Benzonase (Sigma) to solubilize the DNA-associated fraction.

Immunoprecipitation

For endogenous co-immunoprecipitation, HeLa cells were lysed in IPLS buffer (Immuno-Precipitation Law Salt ; 50 mM (m/v) Tris-HCl pH 7.5, 0.5 mM (m/v) EDTA pH8, 0.5% (v/v) NP-40, 10% (v/v), glycerol, 120 mM (m/v) NaCl, cOmplet Protease Inhibitors (Roche) and Halt Phosphatase Inhibitors (Thermo Scientific). Lysates were incubated with Protein G magnetic beads (Millipore) for 1h at 4°C for precleared step and incubated at 37°C for 30 min with or without RNase A (200 ug/ml) (Thermo Scientific). Then, lysates were incubated for 2h at 4°C with Protein G magnetic beads (Millipore) and anti-ERG antibody or rabbit anti-IgG antibody (Santa Cruz). Beads were washed 4 times with ILPS buffer. Immunoprecipitates were boiled in Laemmli buffer and analyzed by SDS-PAGE and Western blot according to standard procedures and developed with the ECL detection kit (GE Healthcare Bio-Sciences, Uppsala, Sweden).

Western blot

Cells were washed once with cold PBS and scrapped on ice with lysis buffer (20mM Tris HCl pH8 ; 1% NP40 ; 150nM NaCl) supplemented with protease inhibitor cocktail (Sigma). Then, cells were pelleted and quantified using Bradford protein assay (ref). Proteins extracts were separated by sodium dodecyl sulfate–polyacrylamide gel electrophoresis and transferred onto nitrocellulose membrane. Blots were incubated with primary antibodies followed by anti-IgG horseradish peroxidase-conjugated. Proteins were detected using enhanced chemiluminescence (Pierce) and images were acquired with ChemiDoc™ Gel Imaging System (Bio-Rad). Primary antibodies are listed in Supplemental Table 7.

MS2 tethering assay

MS2 tethering assays were performed using HeLa cells co-transfected with control MS2-CP or various MS2-CP-tagged constructs, together with SMN2-MS2 (Hua et al., 2008). This minigene was gifted from Krainer's lab (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, USA) (Sun et al., 2012). Reverse transcription (RT)-PCR amplifications were performed (see above) using forward and reverse primers of SMN2-MS2 minigene (Supplemental Table 7).

GST pull-down assay

HA-FLAG-tagged ERG was produced in inducible stable HeLa cell lines, and GST-tagged RBFOX1 or GST-tagged RBFOX2 construction domains were produced in BL21 Star E.coli (Invitrogen). GST-fusion protein production and GST pull-down procedure were performed as described in (Dequiedt et al., 2005) using Glutathione-Superflow Resin (Clontech).

Immunofluorescence

Cells were grown on Lab-Tek CC2™ chamber slide, thoroughly washed with PBS and fixed with 4% PFA at room temperature for 15 minutes. Cells were washed with PBS, permeabilized with 0.2% Triton X-100 in PBS for 30 minutes and blocked in 1% BSA 0.1% Triton X-100 in PBS for 30 minutes. The Lab-Tek chamber slides were incubated for one hour with Phalloidin conjugated with Alexa 488 (1:100) and mounted with ProLong™ containing DAPI (1/10000).

Three-dimensional collagen spheroid assay

Spheroids were formed using 2,000 cells plated in Ultra Low Attachment plates (ULA, Corning) in 50 µL of culture medium. After three days of natural aggregation by gravity, spheroids were embedded in 50 µL of collagen type-I solution (5mg/ml collagen, 1X MEM, 0.025M Hepes, 0.017M NaOH) and allowed to polymerize at 37°C for 30 minutes in a humidified air incubator.

RNA extraction and RT-PCR

RNA extractions were performed on 1µg of RNA using the Nucleospin II kit (Macherey-Nagel) and reverse-transcribed using the High-Capacity cDNA Reverse Transcription kit (Applied Biosystems). Next, cDNA molecules were amplified by PCR performed using the AmpliTaqGold DNA Polymerase kit with Gold Buffer and MgCl2 (Applied Biosystems). Oligonucleotides were purchased from MWG Eurofins Genomics (Supplementary data). Reactions were run on an ABI/PRISM 7500 instrument and analyzed using the 7500 system SDS software (Applied Biosystems) Then, PCR reactions were loaded on 2% agarose gel electrophoresis with SYBR Safe DNA Gel Stain (1/10000, invitrogen). The migration took place with the 6X DNA Loading Dye and the GeneRuler 100bp Plus DNA Ladder (Fermentas). Gels were finally observed and photographed on a UV lamp and images were analyzed by densitometry using the ImageJ software. Primers used are listed in Supplemental table 7.

RNA sequencing and data processing

Every RNA samples were evaluated for integrity using BioAnalyzer instrument (Agilent). All samples displayed excellent quality (RNA Integrity Number above 9). Libraries were performed using the TruSeq Stranded mRNA Library Preparation Kit. Equimolar pool of libraries were sequenced on a Illumina HiSeq 2500 machine using paired-ends reads (PE, 2x101bp) and High Output run mode allowing to get 200 millions of raw reads per sample. Raw reads were mapped on the human reference genome hg19 using the STAR aligner (v.2.5.0a)[46]. PCR-duplicated reads and low

mapping quality reads (MQ<20) were removed using Picard tools and SAMtools, respectively. We next used rMATS (v3.0.9)[17], an event-based tool, to identify differentially spliced events using RNA-seq data. Five distinct alternative splicing events were analyzed using rMATS: skipped exons (SE), alternative 3' splice sites (A3SS), alternative 5' splice sites (A5SS), mutually exclusive exons (MXE) and retained introns (RI). Briefly, rMATS uses a counts-based model, to calculate percent of spliced-in (PSI) value among replicates, using both spliced reads and reads that mapped the exon body. We used three different thresholds to identify differentially spliced events between two groups: splicing event has to be (i) supported by at least 15 unique reads, (ii) |ΔPSI| > 10%; (iii) FDR < 0.05.

Gene expression analysis was performed as follows: aligned reads were counted using htseq-count[47] and normalized according to the DESeq size factors method[48]. We used 2-fold change and FDR < 0.05 as the determination of differentially expressed genes. Statistical analysis and plots were performed inside R environment version 3.1.0. P-value thresholds are depicted as follows: *: $p<0.05$; **: $p<0.01$; ***: $p<0.001$. Statistical analysis of GO terms enrichment was performed on the ToppFun online software. Fastq files are available in GEO database under the accession number XXXX.

Motif enrichment analysis

To identify the RBPs involved in the regulation of the skipped exons, we extent the software rMAPS. This tool identifies the binding positions of RBPs around skipped exons. The purpose of rMAPS is to identify known RBP motifs that are significantly enriched in differentially regulated exons between two sample groups as compared to control (background) events. rMAPS analyzes each set of 300 nt length sequences, with a sliding window of 50 nt, and counts the number of times the motif matches each sequence. The resulting "enrichment score" is then used to compare local enrichment in the window between significant exons and background exons by the Wilcoxon rank sum test. This process results in a set of 250 highly correlated p-values, which rMAPS summarizes by the minimum (raw) p-value.

We extended rMAPS by proposing a method to identify intervals significantly enriched for a given RBP. Here, "significantly" means that with high probability, the proportion of false positives (or False Discovery Proportion: FDP) among any of the selected intervals does not exceed a user-defined threshold. This method is based on the concept of post-hoc inference, as introduced by Goeman and Solari (Statistical Science, 2011) and further studied by Blanchard, Neuvial and Roquain (https://arxiv.org/abs/1703.02307). Importantly, the user may choose the threshold on the FDP post hoc, ie after the data analysis. Compared with rMAPS, this approach reduces the number of identified false positives and allows the identification of their precise binding site. We chose to call significant all the intervals with a FDP less than 25%.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: O. Saulnier, K. Guerdi, M. Dutertre, O. Delattre, F. Dequiedt
Development of methodology: O. Saulnier, K. Guerdi, A. Chakraborty, B. Sadacca
Acquisition of data: O. Saulnier, K. Guerdi, A. Chakraborty, M-M. Aynaud, J. Pineau, K. Laud
Analysis and interpretation of data: O. Saulnier, C. O'Grady, B. Sadacca
Writing, review, and/or revision of the manuscript: O. Saulnier, M. Dutertre, O. Delattre, F. Dequiedt
Administrative, technical, or material support: B. Sadacca, S. Lalami-Grossetete, O. Mirabeau, X. Rambout
Study supervision: M. Dutertre, O. Delattre, F. Dequiedt

REFERENCES

1. Sharrocks, A. D. The ETS-domain transcription factor family. *Nat. Rev. Mol. Cell Biol.* **2,** 827–837 (2001).
2. Zucman, J. *et al.* Combinatorial generation of variable fusion proteins in the Ewing family of tumours. *EMBO J.* **12,** 4481–4487 (1993).
3. Delattre, O. *et al.* Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours. *Nature* **359,** 162–165 (1992).
4. Gangwal, K. *et al.* Microsatellites as EWS/FLI response elements in Ewing's sarcoma. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 10149–10154 (2008).
5. Guillon, N. *et al.* The oncogenic EWS-FLI1 protein binds in vivo GGAA microsatellite sequences with potential transcriptional activation function. *PloS One* **4,** e4932 (2009).
6. Boulay, G. *et al.* Cancer-Specific Retargeting of BAF Complexes by a Prion-like Domain. *Cell* **171,** 163-178.e19 (2017).
7. Sanchez, G. *et al.* Alteration of cyclin D1 transcript elongation by a mutated transcription factor up-regulates the oncogenic D1b splice isoform in cancer. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 6004–6009 (2008).
8. Selvanathan, S. P. *et al.* Oncogenic fusion protein EWS-FLI1 is a network hub that regulates alternative splicing. *Proc. Natl. Acad. Sci. U. S. A.* **112,** E1307-1316 (2015).
9. Knoop, L. L. & Baker, S. J. The splicing factor U1C represses EWS/FLI-mediated transactivation. *J. Biol. Chem.* **275,** 24865–24871 (2000).
10. Rambout, X. *et al.* The transcription factor ERG recruits CCR4–NOT to control mRNA decay and mitotic progression. *Nat. Struct. Mol. Biol.* **23,** 663–672 (2016).
11. Stark, C. *et al.* BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* **34,** D535-539 (2006).
12. Snel, B., Lehmann, G., Bork, P. & Huynen, M. A. STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Res.* **28,** 3442–3444 (2000).
13. Kornblihtt, A. R. Coupling transcription and alternative splicing. *Adv. Exp. Med. Biol.* **623,** 175–189 (2007).
14. Damianov, A. *et al.* Rbfox Proteins Regulate Splicing as Part of a Large Multiprotein Complex LASR. *Cell* **165,** 606–619 (2016).
15. Sun, S., Zhang, Z., Fregoso, O. & Krainer, A. R. Mechanisms of activation and repression by the alternative splicing factors RBFOX1/2. *RNA* **18,** 274–283 (2012).
16. Siddique, H. R., Rao, V. N., Lee, L. & Reddy, E. S. Characterization of the DNA binding and transcriptional activation domains of the erg protein. *Oncogene* **8,** 1751–1755 (1993).

17. Shen, S. *et al.* rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl. Acad. Sci. U. S. A.* **111,** E5593–E5601 (2014).

18. Ray, D. *et al.* A compendium of RNA-binding motifs for decoding gene regulation. *Nature* **499,** 172–177 (2013).

19. Anderson, E. S. *et al.* The cardiotonic steroid digitoxin regulates alternative splicing through depletion of the splicing factors SRSF3 and TRA2B. *RNA* **18,** 1041–1049 (2012).

20. Kuroyanagi, H. Fox-1 family of RNA-binding proteins. *Cell. Mol. Life Sci. CMLS* **66,** 3895–3907 (2009).

21. Yeo, G. W. *et al.* An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nat. Struct. Mol. Biol.* **16,** 130–137 (2009).

22. Jangi, M., Boutz, P. L., Paul, P. & Sharp, P. A. Rbfox2 controls autoregulation in RNA-binding protein networks. *Genes Dev.* **28,** 637–651 (2014).

23. Zhang, C. *et al.* Defining the regulatory network of the tissue-specific splicing factors Fox-1 and Fox-2. *Genes Dev.* **22,** 2550–2563 (2008).

24. Remy, I. & Michnick, S. W. A highly sensitive protein-protein interaction assay based on Gaussia luciferase. *Nat. Methods* **3,** 977–979 (2006).

25. Ying, Y. *et al.* Splicing Activation by Rbfox Requires Self-Aggregation through Its Tyrosine-Rich Domain. *Cell* **170,** 312-323.e10 (2017).

26. Carrillo, J. *et al.* Cholecystokinin down-regulation by RNA interference impairs Ewing tumor growth. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **13,** 2429–2440 (2007).

27. Braeutigam, C. *et al.* The RNA-binding protein Rbfox2: an essential regulator of EMT-driven alternative splicing and a mediator of cellular invasion. *Oncogene* **33,** 1082–1092 (2014).

28. Shapiro, I. M. *et al.* An EMT-driven alternative splicing program occurs in human breast cancer and modulates cellular phenotype. *PLoS Genet.* **7,** e1002218 (2011).

29. Pradella, D., Naro, C., Sette, C. & Ghigna, C. EMT and stemness: flexible processes tuned by alternative splicing in development and cancer progression. *Mol. Cancer* **16,** 8 (2017).

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

31. Chaturvedi, A. *et al.* Molecular dissection of the mechanism by which EWS/FLI expression compromises actin cytoskeletal integrity and cell adhesion in Ewing sarcoma. *Mol. Biol. Cell* **25,** 2695–2709 (2014).

32. Wiles, E. T., Bell, R., Thomas, D., Beckerle, M. & Lessnick, S. L. ZEB2 Represses the Epithelial Phenotype and Facilitates Metastasis in Ewing Sarcoma. *Genes Cancer* **4,** 486–500 (2013).

33. Franzetti, G.-A. *et al.* Cell-to-cell heterogeneity of EWSR1-FLI1 activity determines proliferation/migration choices in Ewing sarcoma cells. *Oncogene* **36,** 3505–3514 (2017).

34. Yang, Y. *et al.* Determination of a Comprehensive Alternative Splicing Regulatory Network and Combinatorial Regulation by Key Factors during the Epithelial-to-Mesenchymal Transition. *Mol. Cell. Biol.* **36,** 1704–1719 (2016).

35. Kiang, K. M.-Y. & Leung, G. K.-K. A Review on Adducin from Functional to Pathological Mechanisms: Future Direction in Cancer. *BioMed Res. Int.* **2018,** 1–14 (2018).

36. Vallenius, T. Actin stress fibre subtypes in mesenchymal-migrating cells. *Open Biol.* **3,** (2013).

37. Climente-González, H., Porta-Pardo, E., Godzik, A. & Eyras, E. The Functional Impact of Alternative Splicing in Cancer. *Cell Rep.* **20,** 2215–2226 (2017).

38. Das, S., Anczuków, O., Akerman, M. & Krainer, A. R. Oncogenic Splicing Factor SRSF1 Is a Critical Transcriptional Target of MYC. *Cell Rep.* **1,** 110–117 (2012).

39. G Hendrickson, D., Kelley, D. R., Tenen, D., Bernstein, B. & Rinn, J. L. Widespread RNA binding by chromatin-associated proteins. *Genome Biol.* **17,** 28 (2016).

40. Han, H. *et al.* Multilayered Control of Alternative Splicing Regulatory Networks by Transcription Factors. *Mol. Cell* **65,** 539-553.e7 (2017).

41. Tomlins, S. A. *et al.* Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* **310,** 644–648 (2005).

42. Im, Y.-H. *et al.* EWS-FLI1, EWS-ERG, and EWS-ETV1 Oncoproteins of Ewing Tumor Family All Suppress Transcription of Transforming Growth Factor ◻ Type II Receptor Gene. 6

43. Dutertre, M. *et al.* Exon-Based Clustering of Murine Breast Tumor Transcriptomes Reveals Alternative Exons Whose Expression Is Associated with Metastasis. *Cancer Res.* **70,** 896–905 (2010).

44. Eswaran, J. *et al.* RNA sequencing of cancer reveals novel splicing alterations. *Sci. Rep.* **3,** (2013).

45. Langer, W. *et al.* Exon Array Analysis using re-defined probe sets results in reliable identification of alternatively spliced genes in non-small cell lung cancer. *BMC Genomics* **11,** 676 (2010).

46. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).

47. Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinforma. Oxf. Engl.* **31,** 166–169 (2015).

48. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** 550 (2014).

COMPLEMENTARY RESULTS

The following results are not present in the above manuscript and are useful to understand biological results and future perspectives of the present work. These results will be shortly included in a manuscript and submitted to a scientific journal.

QKI is enriched in alternative splicing events regulated by EWS-FLI1

As mentioned in the above manuscript, we performed RNA-seq on multiple Ewing sarcoma cell lines to identify a core-splicing signature EWS-FLI1-dependent. We scanned RBPs binding motifs and found a significant enrichment for RBFOX2 binding motif. Nevertheless, we also found other RBPs motifs significantly enriched in EWS-FLI1-dependent exons suggesting a multiple network involved in the splicing function of EWS-FLI1. For instance, quaking (QKI) protein, which plays also an important role in the epithelial-to-mesenchymal transition, is significantly enriched in EWS-FLI1-regulated exons **(Figure 30A)**. QKI splicing mechanism is well characterized and is similar to RBFOX mechanism. QKI binding motif is enriched upstream of spliced-out exons and downstream of spliced-in exons **(Figure 30B)** (Hall et al., 2013). Strinkingly, as seen for RBFOX2, the RNA-map of QKI in EWS-FLI1 splicing exons is the opposite as what it has been previously described. Indeed, we found that QKI binding motif is highly enriched upstream of EWS-FLI1 spliced-in exons, whereas QKI is known to promote spliced-out of exons by binding upstream intron. These results suggest that QKI participates in the EWS-FLI1 splicing program and may have an opposite role on splicing.
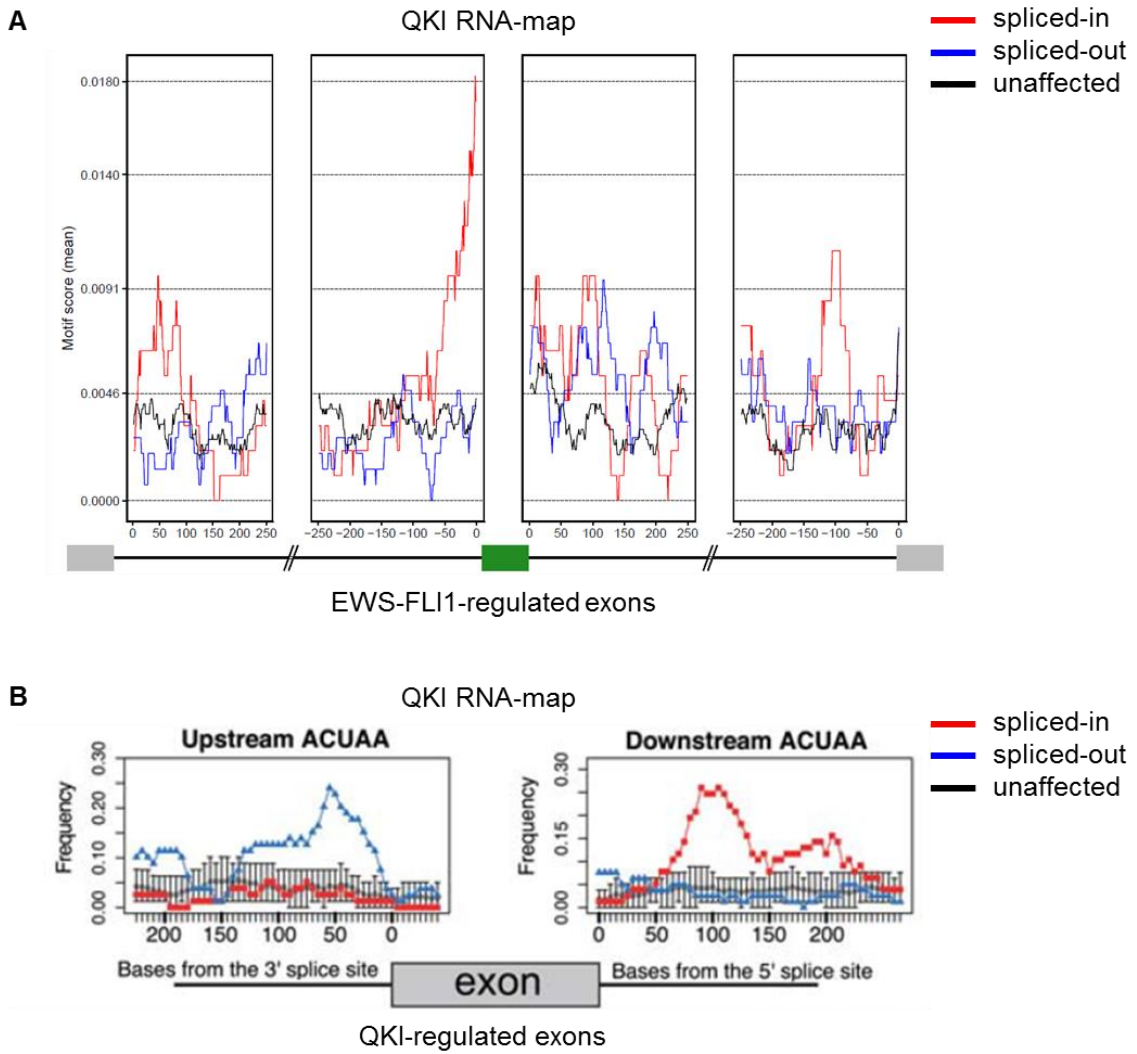
Figure 30: QKI binding motif is enriched in flanking introns of alternative exons regulated by EWS-FLI1. **(A)** RNA-map of QKI upon EWS-FLI1 depletion showing enrichment of QKI binding motif in upstream EWS-FLI1 spliced-in exons. **(B)** Previously described QKI RNA-map showing enrichment upstream of QKI spliced-out exons and downstream of QKI spliced-in exons (Hall et al., 2013.

EWS-FLI1 and QKI regulate a common splicing program and has an opposite role on splicing

To confirm these *in silico* prediction, we performed QKI silencing using specific siRNAs targeting QKI. As expected, 48 hours post-transfection, QKI protein expression was strongly decreased **(Figure 31A)**. Interestingly, it has been described that RBFOX2 and QKI interact together and cooperate to regulate a common splicing program (Brosseau et al., 2014; Hegele et al., 2012; Huttlin et al., 2017; Lim et al., 2006). We performed RNA sequencing on QKI-depleted ASP14 cells and performed rMATS analysis to identify QKI-dependent splicing events. As expected, comparison of alternative splicing events induced by EWS-FLI1, RBFOX2 and QKI highlighted a highly significant overlap **(Figure 31B)** suggesting that these proteins may regulate common splicing program within the same protein complex. To go further, as observed for RBFOX2, we demonstrated that EWS-FLI1 mostly antagonizes QKI splicing function as the majority of splicing events were oppositely regulated **(Figure 31C)**. To confirm these results, we selected a subset of splicing events that were described in the literature as known RBFOX2 or QKI targets. Using RT-PCR, we confirmed that these targets were indeed regulated by RBFOX2 or/and QKI **(Figure 31D)**. In addition, we also showed that QKI and RBFOX2 oppositely regulate alternative splicing as compared to EWS-FLI1 in all exons that we tested. This result confirmed that EWS-FLI1 has an antagonistic effect on QKI (and RBFOX2) splicing function. Furthermore, Katia Guerdi (PhD student of Franck Dequiedt lab) performed co-immunoprecipitation experiments using antibodies against EWS, FLI1 and EWS-FLI1. Surprisingly, western blots revealed that EWS-FLI1 does not interact with QKI, despite interaction between wild-type FLI1 and QKI **(Figure 31E)**. We performed RNA immunoprecipitation of endogenous QKI followed by quantitative PCR on two common splicing targets. Interestingly, we also found QKI binding was increased in EWS-FLI1-depleted cells, suggesting that EWS-FLI1 repress QKI binding to its pre-mRNAs targets **(Figure 31F)**. Altogether, these results suggest that RBFOX2 and QKI may form a protein complex together to regulate a common splicing program. However, in Ewing cells this protein complex could be disrupted by the interaction between EWS-FLI1 and RBFOX2. We hypothesized that despite the interaction between EWS-FLI1 and RBFOX2, the splicing function of RBFOX2/QKI complex is altered due to the loss of QKI partner. Disruption of this protein complex could be mediated by the low complexity region of EWS, which is present in EWS-FLI1 fusion and that is responsible to form protein aggregates, hence sequestrating RBFOX2.
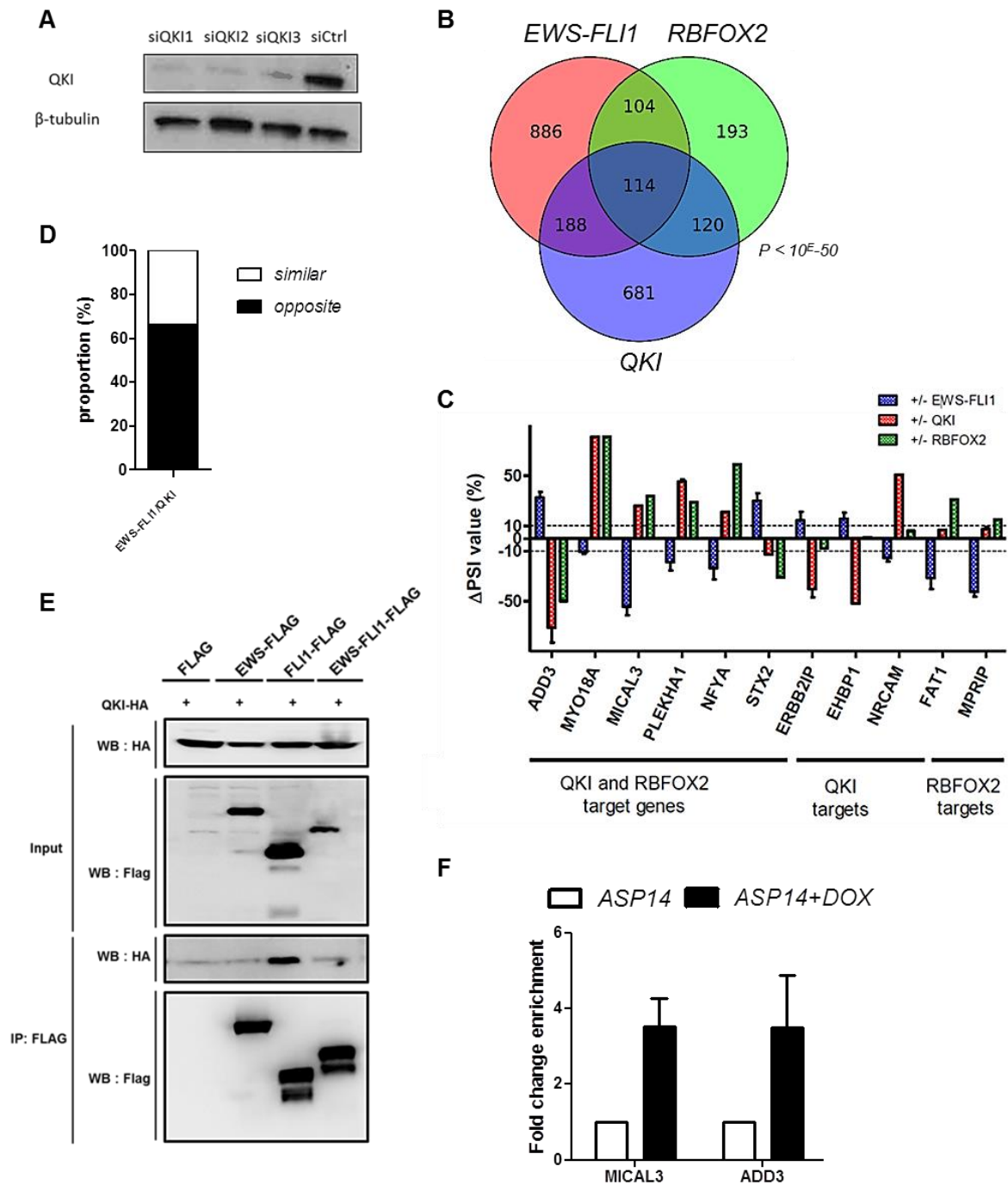
Figure 31: EWS-FLI1 and QKI regulate a common splicing program in an opposite manner. **(A)** Western blot validation of QKI protein knockdown using three independent siRNAs. **(B)** Overlap between alternative exons regulated by QKI, RBFOX2 and EWS-FLI1 using rMATS analysis**. (C)** Proportion of common target exons of EWS-FLI1 and QKI categorized as "similar" or "opposite" depending on whether ΔPSI values vary in respectively the same or opposite direction in EWS-FLI1 and QKI knockdown ASP14 cells. **(B)** RT-PCR analysis of a representative set of exons oppositely spliced by EWS-FLI1 and QKI/RBFOX2. **(E)** Immunoprecipitation of FLAG-tagged version of EWS, FLI1 and EWS-FLI1 followed by anti-HA western blotting. **(F)** RNA immunoprecipitation experiments using anti-QKI antibodies followed by qPCR to detect MICAL3 and ADD3 transcripts. Samples are RNA from EWS-FLI1 expressing (ASP14) or depleted (ASP14+DOX) ASP14 cells.

ADD3 function is potentially impacted by EWS-FLI1 mis-splicing

We then studied the functional impact of the splicing induced by EWS-FLI1 and its potential oncogenic role. We focused on gamma-adducin (ADD3), which is a gene spliced by EWS-FLI1 **(Figure 32A)** and is involved in cytoarchitecture, which is a key process regulated by EWS-FLI1. ADD3 is a membrane cytoskeletal component that shares high structural similarities with alpha and beta adducin (ADD1 and ADD2). Oligomerization are necessary for Adducin activities, such as cell motility and cell organization. All Adducin family members contain an N-terminal globular head domain, a neck domain and a C-terminal tail domain that is homologous to MARCKS protein. The MARCKS-related domain contains a calmodulin binding site and numerous phosphorylation sites that are essential to interact with spectrin and actin. Overall, this domain is fundamental for Adducin functions and may play a role in the spectrin-actin assembly, hence cell architecture. Interestingly, EWS-FLI1-mediated splicing of ADD3 occurs in the exon 14 **(Figure 32B)**, which is right upstream the MARCKS domain. We performed tertiary structure prediction using Simple Modular Architecture Research Tool (SMART) software (http://smart.embl-heidelberg.de) and we found that insertion of cassette exon 14 leads to a 32 amino acids sequence predicted for low complexity region and a small-coiled coil **(Figure 32C)**. We hypothesized that ADD3 exon 14 isoform, which is upregulated by EWS-FLI1, may have a different function on cytoskeleton structure.
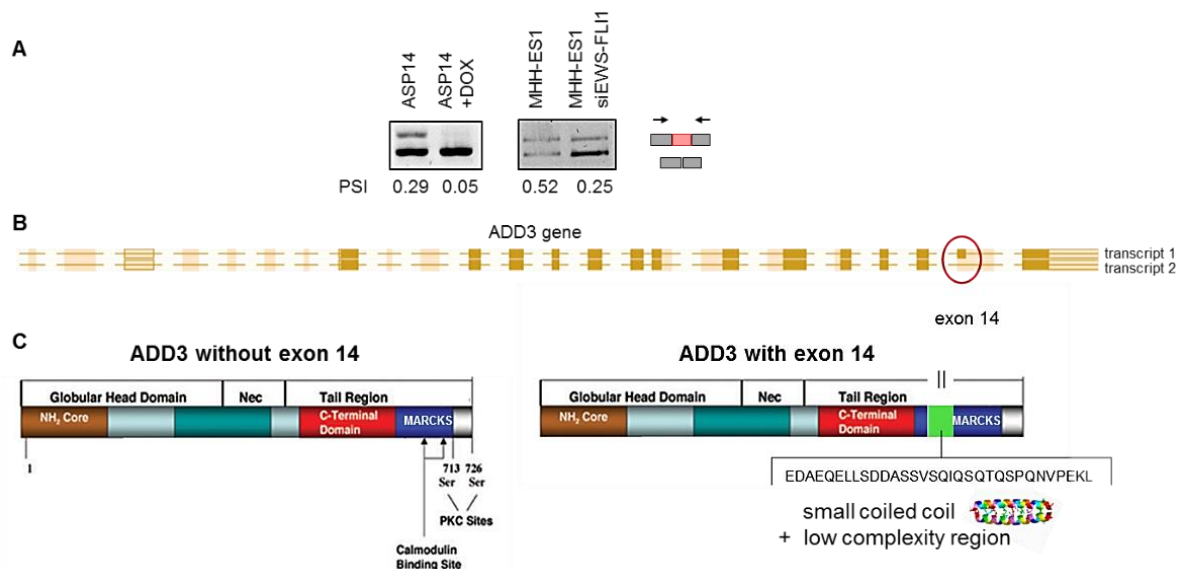
Figure 32: ADD3 is a splicing target of EWS-FLI1. **(A)** RT-PCR of two Ewing sarcoma cell lines upon EWS-FLI1 depletion (doxycycline treatment or siEWS-FLI1) highlighting exon 14 alternative splicing. **(B)** Screenshot of ADD3 genomic structure from Ensembl database. **(C)** Organization of domain of ADD3 protein described in Pariser et al., 2005 and tertiarty structure prediction from SMART software.

ADD3 exon 14 splicing is regulated during EMT

We next wanted to evaluate the splicing of ADD3 exon 14 across human tissues. We performed rMATS analysis on 53 RNA-seq of Ewing sarcoma tumors to calculate transcriptome-wide PSI values. In addition, we downloaded 10,549 splicing profiles of normal tissues from GTEx database (https://www.gtexportal.org/; version: V7; June 2018). Comparison of ADD3 exon 14 PSI value highlighted a tenfold increase in Ewing sarcoma tumors as compared to normal tissues (p < 10E-15) with a median of 0.223 and 0.0201 respectively **(Figure 33A)**. Strikingly, comparison of epithelial and mesenchymal-origin tissues from GTEx database highlighted a clear difference (p < 10E-15) of ADD3 exon 14 splicing **(Figure 33B)**. Mesenchymal cells appeared to specifically spliced-out exon 14. This was further confirmed using RT-PCR on a subset of cell lines from the lab showing a clear separation between mesenchymal and fibroblast cells, which completely spliced-out ADD3 exon 14, and epithelial cell lines **(Figure 33C)**. We also tested if ADD3 exon 14 splicing was modulated in neuroblastoma cell lines. Recently, it has been described that neuroblastoma cell lines are separated in three subgroups depending on their differentiation state: (i) the mesenchymal identity, (ii) the noradrenergic identity and (iii) an intermediate identity with mixed cell population (Boeva et al., 2017; van Groningen et al., 2017). Astonishingly, we have demonstrated that ADD3 exon 14 splicing is clearly dependent on the cell identity of neuroblastoma cell lines **(Figure 33D)**. The SKNSH neuroblastoma cell line arbor a mixed phenotype with both cell identities. Using fluorescence activated cell sorting (FACS), we were able to sort cells according to the mesenchymal marker CD44 and performed RNA sequencing. Sashimi plot shown in **Figure 33E** nicely supported our data and confirmed that ADD3 exon 14 splicing pattern is highly dependent from the cell phenotype.
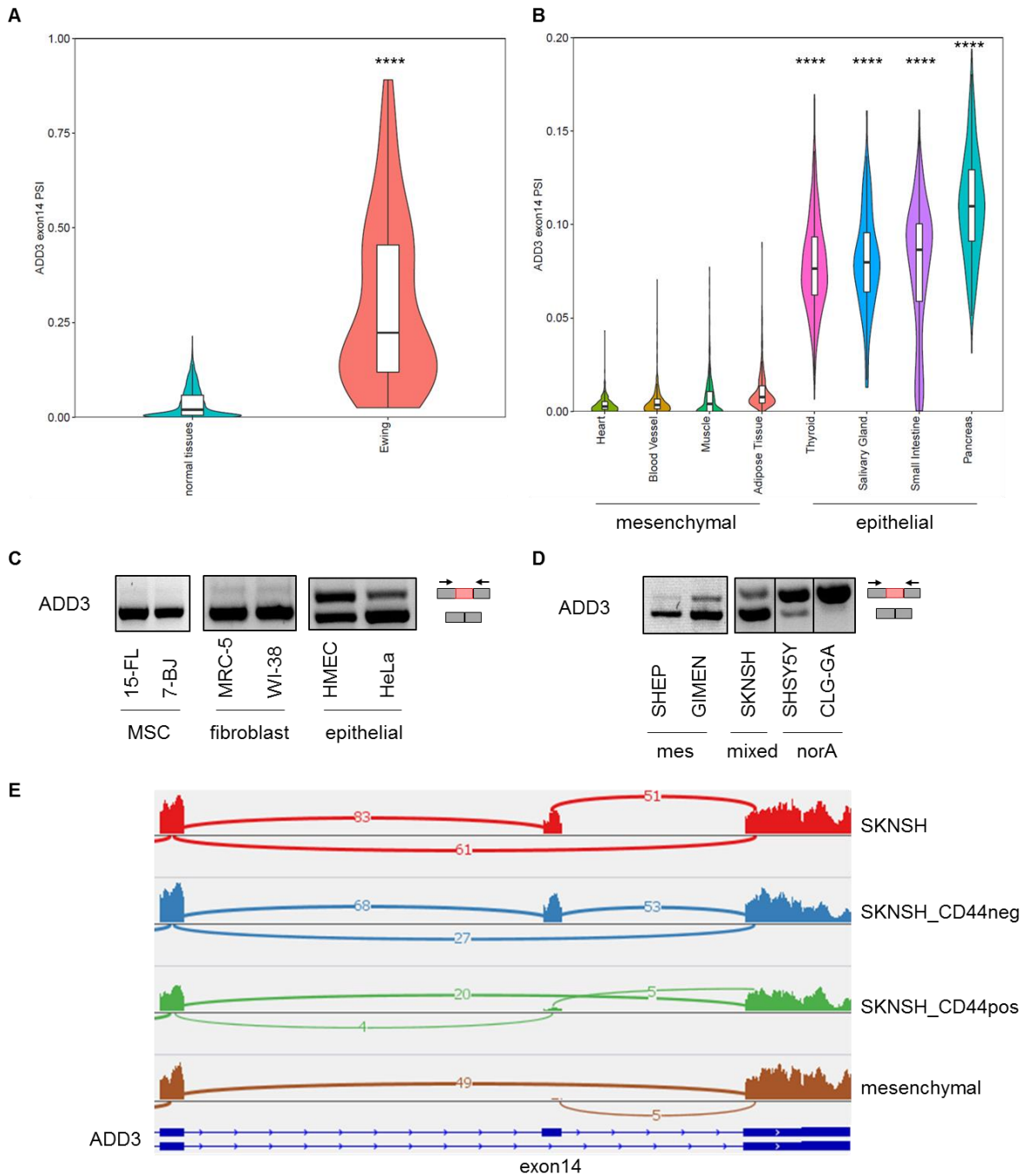
Figure 33: ADD3 exon 14 transcript is expressed in Ewing tumors and is absent in mesenchymal cells. **(A)** Violin plot of ADD3 exon 14 PSI value in Ewing sarcoma tumors and normal tissues extracted from GTEx database. **(B)** Violin plot of ADD3 exon 14 PSI value from a subset of epithelial and mesenchymal origin tissues. **(C)** RT-PCR showing an absence of ADD3 exon 14 transcript in mesenchymal cell lines as compared to epithelial cell lines. **(D)** RT-PCR in neuroblastoma cell lines showing a differential splicing pattern according to the subgroup-identity. **(E)** Sashimi plots extracted from RNA-seq dataset on parental SKNSH neuroblastoma cell line and after cell sorting according the mesenchymal CD44 marker. Mesenchymal cell line is displayed as control. Abbreviations: mesenchymal stem cells (MSC), mesenchymal identity (mes), noradrenergic identity (norA).

113

Genome editing of ADD3 exon 14 using CRISPR-Cas9 technology

To further study the functional impact of ADD3 mis-splicing in Ewing sarcoma, we performed inhibition of ADD3 exon 14 transcript using two techniques: (i) specific siRNAs transfection and (ii) CRISPR-Cas9 technology. We used three independent siRNAs that target specifically exon 14 of ADD3 and a combination of guide RNA (gRNA) to delete specifically the exon 14 region of ADD3. Both technologies were fully concordant but I will describe here the results mainly obtained by CRISPR-Cas9 technology. The experimental design is presented in **Figure 34A**. Single cut (ex14) and double cut were performed using combination of left (L1 and L2) and right (R1 and R2) gRNAs to entirely delete exon 14. Clone transfected with L1/R2 gRNAs with genomic homozygous deletion was confirmed using PCR **(Figure 34B)** and Sanger sequencing (data not shown). In addition, we confirmed using RT-PCR that our selected clone completely abolished ADD3 exon 14 mRNA expression **(Figure 34C)**.
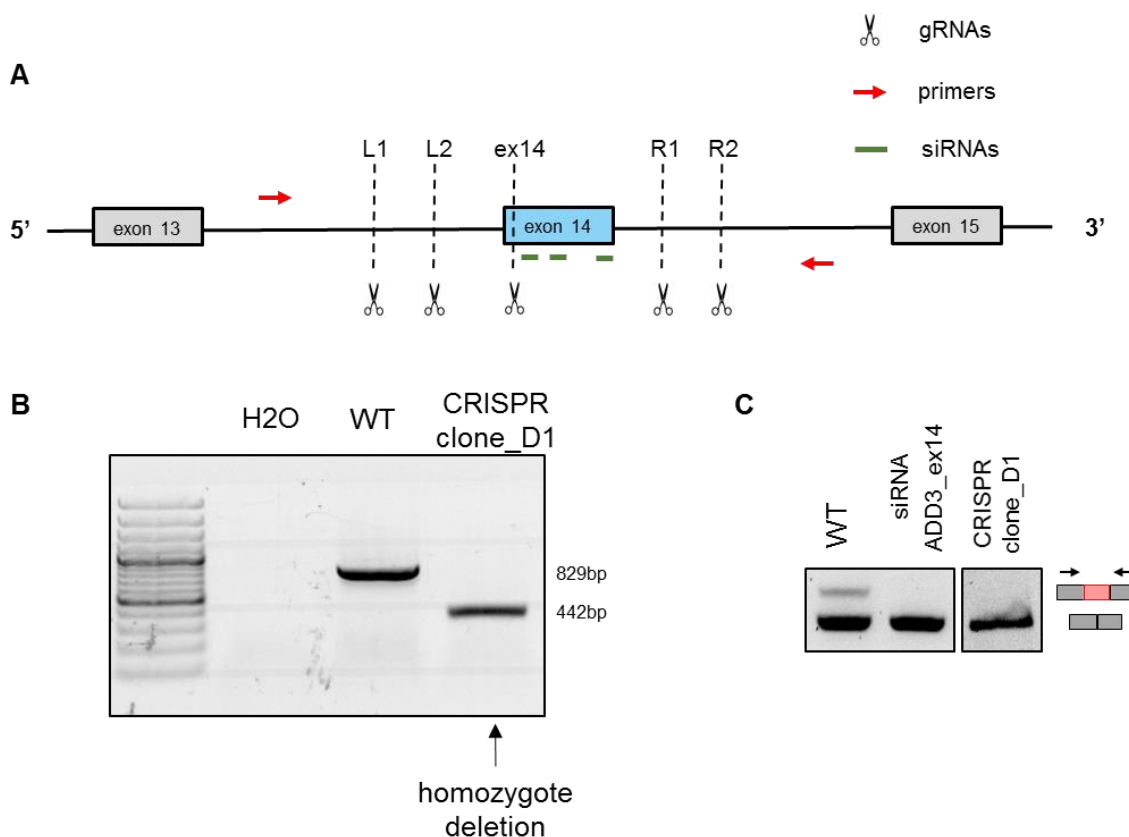
Figure 34: Deletion of ADD3 exon 14 using specific siRNAs and CRISPR-Cas9 technology. **(A)** Experimental design of ADD3 exon 14 inhibition using three independent siRNAs targeting specifically exon 14 of ADD3. Five gRNAs have been designed to perform single and double cut. **(B)** PCR validation of ADD3 exon 14 genomic deletion after clone selection. **(C)** RT-PCR analysis showing complete loss of ADD3 exon 14 mRNA expression by both techniques.

114

Cells depleted for ADD3 exon 14-containing transcript have a mesenchymal-migrating phenotype

We next investigated the functional consequences of ADD3 exon 14 transcript deletion. In concordance with what has been previously described **(Manuscript Figure 6)**, depletion of ADD3 exon 14 transcript induces a drastic modification of cell properties such as cytoarchitecture, invasiveness, proliferation and chemosensitivity. Ewing sarcoma cells depleted for ADD3 exon 14 transcript highly adhere to plastic flask and have a flattened phenotype under the microscope **(Figure 35A)**. There is a massive modulation of the cell organization structure with apparition of numerous stress fibers **(Figure 35B)** using F-actin staining (phalloidin). This phenotype correlates with pro-invasive property as observed in 3D collagen type-I spheroid **(Figure 35C)**. Cell growth is also impacted **(Figure 35D)** and cells are more resistant than wild-type cells to standard chemotherapy **(Figure 35E)**. This last observation is coherent with the growth speed of cells because chemotherapies mainly target highly proliferative cells. Altogether, these results highlighted that the splicing of ADD3 exon 14 induced by EWS-FLI1 plays a major role on cell structure organization and may influence Ewing sarcoma chemosensitivity.
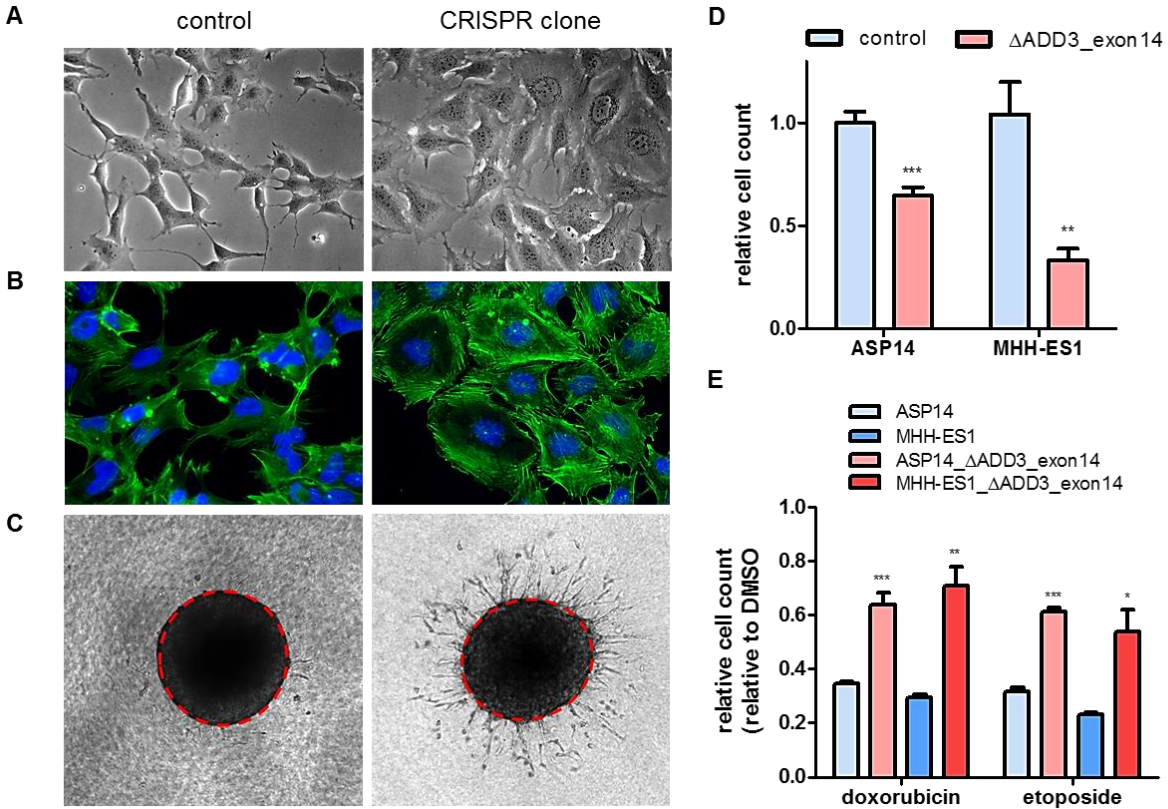
Figure 35: ADD3 exon 14 transcript deletion has major impact on cell phenotype. **(A)** Brightfield imaging of Ewing sarcoma cells. **(B)** Immunofluorescence using phalloidin staining. **(C)** Three dimentional collagen type-I spheroid showing high invasive property of ADD3 exon 14-depleted cells. **(D)** Cell count 48 hours post-transfection by siCTRL or siRNA targeting ADD3 exon 14 transcript. **(E)** Cell count 48 hours post treatment using DMSO, doxorubicin (150nM) or etoposide (2.5μM). Cells were treated one day by siCTRL or siRNA targeting ADD3 exon 14 transcript prior the chemotherapy treatment.

Cell phenotype is partially rescued by ADD3 exon 14 overexpression

Next, we performed rescue experiment by expressing the full-length ADD3 transcript containing exon 14 in mesenchymal cell line and ASP14 EWS-FLI1-depleted cell line (both cell lines do not express ADD3 exon 14 transcript). We transfected cells with empty pcDNA3.1(+) expression vector and with ADD3_ex14_inclusion_pcDNA3.1(+) vector (Genscript, Hong Kong). We observed 48 hours post-transfection a modification of cell morphology from a mesenchymal-like phenotype towards cells exhibiting neurite outgrowth **(Figure 36A)**. The rescue is not complete, cells do not revert back an Ewing sarcoma phenotype (even if we do observe loss of stress fibers). This can be explained by the fact that overexpression of ADD3 exon 14 transcript in an expression vector leads to a massive amount of mRNAs molecules that is completely unphysiological. As seen by RT-PCR, we cannot detect any band representing the initial exon 14 skipping transcript, which is still present **(Figure 36B)**. This is due to the limitation of the RT-PCR that is not a sensitize detection method for extreme differences of mRNAs molecules. Using RT-qPCR, we found that exon 14 transcript is expressed more than 600 fold as compared to Δexon 14 transcript (data not shown). It would be interesting to redo the experiment by modifying the vector promoter or by adding a limiting amont of siRNAs targeting ADD3 exon 14 transcript to control the amount of mRNA molecules of this transcript and to reproduce physiological levels of ADD3 exon 14 mRNA.
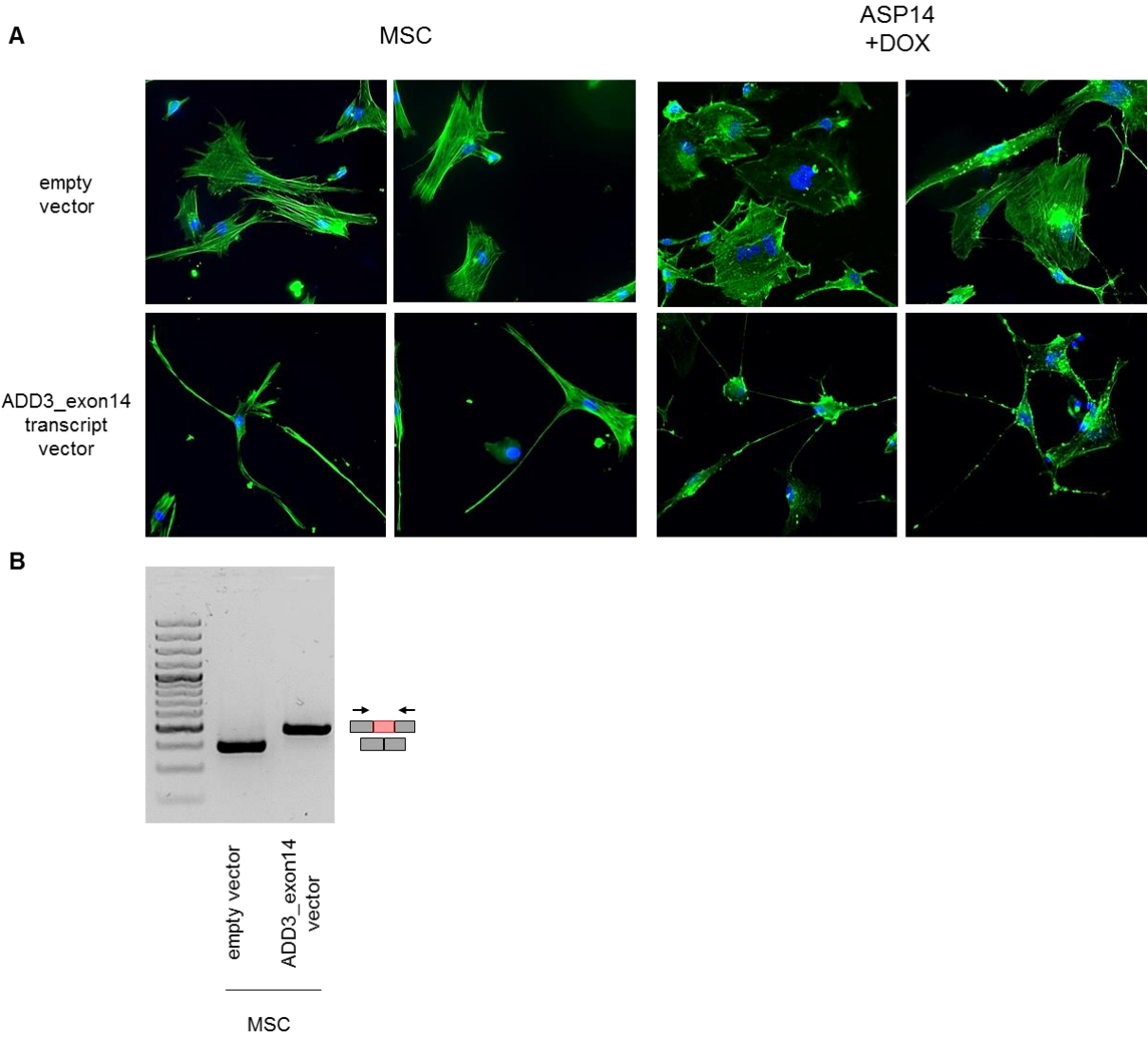
Figure 36: ADD3 exon 14 rescue expression leads to phenotypical changes. **(A)** Immunofluorescence staining of F-actin highlights loss of stress fibers and cell structure upon ADD3 exon 14 overexpression in mesenchymal stem cells (MSC) and Ewing sarcoma cells depleted for EWS-FLI1. (ASP14 +DOX). **(B)** RT-PCR validation of vector transfection in MSC.

DISCUSSION

I.   Deciphering the role of ERG and EWS-ETS proteins on splicing

So far, few transcription factors have been linked to post-transcriptional processes such as splicing. However, recent studies suggest that transcription regulators have a broader role on mRNA processes including alternative splicing. For instance, transcription factors can affect the splicing landscape of a cell by multiple mechanisms including the regulation of mRNA expression levels of splicing factor or the regulation of the RNA POL2 elongation rate (Das et al., 2012; Sanchez et al., 2008a). In addition, it has been shown that chromatin-bound proteins, such as transcription factors, lacking canonincal RNA binding domains, can bind RNA (G Hendrickson et al., 2016). To date, only Spi-1/PU.1 among the ETS transcription factor family, has been identified to control alternative splicing (Hallier et al., 1996, 1998). Collectively, this suggest a complex multilayered control of alternative splicing process. Thanks to next generation sequencing, public available databases, such as protein-protein interaction databases, and computational approaches, it becomes, now, "easier" to uncover unannotated splicing regulators such as transcription factors or chromatin binding proteins (Han et al., 2017).

Here, by taking advantage of the BioGRID (https://thebiogrid.org) and STRING (https://string-db.org) databases, we curated list of ERG-interactors to keep only physical interactions. We performed gene set enrichment analysis using ToppGene suite (https://toppgene.cchmc.org) and, to our surprise, found a highly significant enrichment for proteins that bind RNA **(Manuscript Figure 1A)**. This analysis may be biaised because known interactors are described in the literature depending on the research area context of a given protein and expression of RNA-binding proteins are quite abundant in human cells. To test this, we performed the same analysis on two very well characterized transcription factors: JUN and MYOD1. Gene set enrichment analysis revealed a highly significant enrichment of proteins related to transcription activity and chromatin binding, as expected **(Figure 38).**
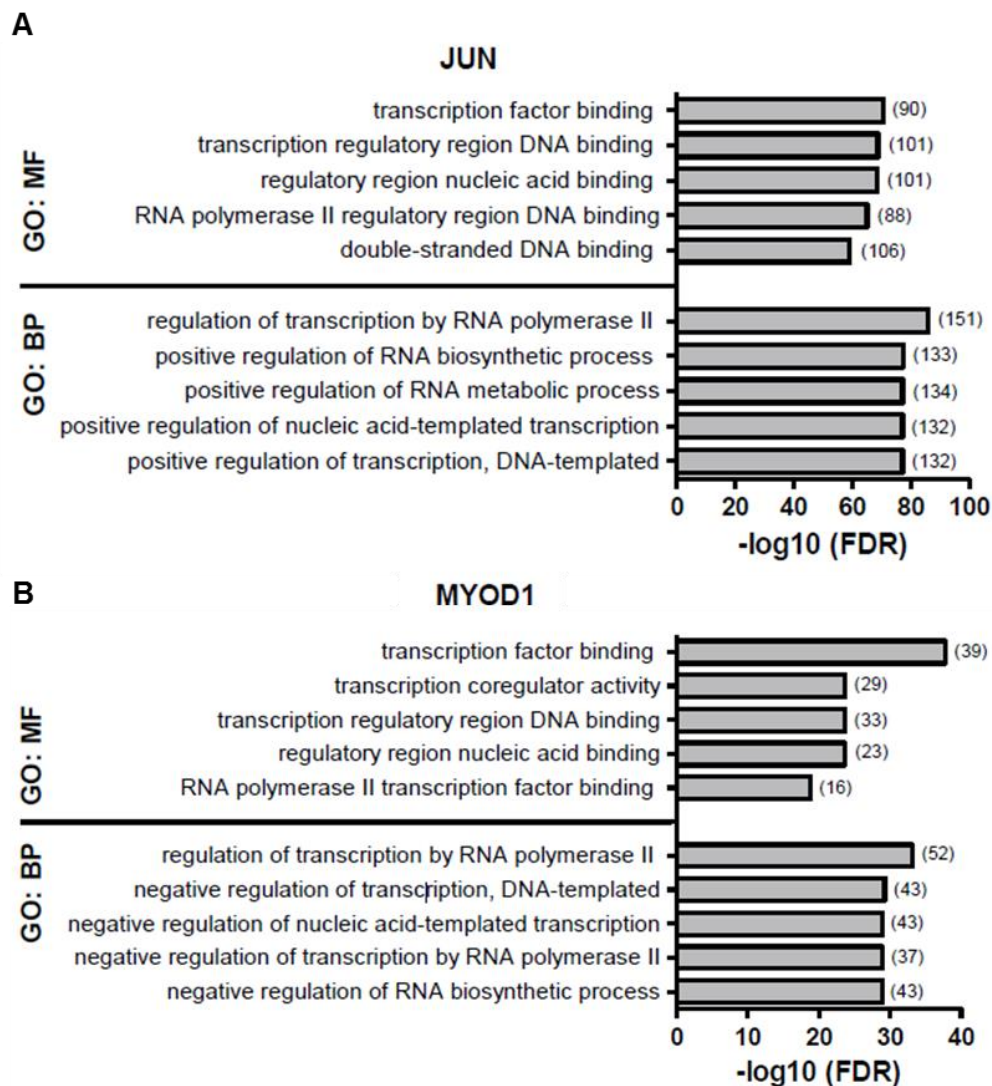
Figure 38: Distribution of top significant GO molecular function and biological process terms from curated lists of (A) JUN protein interaction and (B) MYOD1 protein interaction.

This observation confirmed the true enrichment of proteins related to RNA-binding in ERG partners and led us to investigate the potential role of ERG in splicing process. Using co-immunoprecipitation experiments, we confirmed the previously described interactions with spliceosome-associated proteins and we also demonstrated that ERG was able to bind core components of the spliceosome machinery, such as U1-70K and U2AF65. Altogether, these data revealed a potential implication of ERG in alternative splicing regulation.

Using a minigene splicing reporter assay and further confirmed by RNA-seq analysis, we demonstrated that ERG was indeed able to alter alternative splicing landscape of HeLa cells. Interestingly, we performed differential gene expression analysis between control and ERG-depleted cells and we found no significant overlap between differentially expressed genes and differentially spliced genes. These observations suggested that functions of ERG as transcription factor and splicing regulator are independent and may involve different domains or co-factors. Indeed, we demonstrated that the C-terminal domain of ERG was required for its splicing function. This observation confirmed that ERG functions in transcription and splicing regulation are independent because the C-terminal domain does not overlap with the ETS DNA binding, which is required for ERG transcriptional activity (Siddique et al., 1993).

ERG is frequently overexpressed in prostate cancer (> 50%), mainly due to gene fusion with the TMPRSS2 partner (Tomlins Science 2005). Our results suggest that ERG-rearranged proteins, which include the C-terminal domain, modify the splicing landscape in prostate cancer. It would be worth looking into splicing differences of ERG-rearranged and ERG WT prostate cancer. These observations might provide new insights on ERG oncogenic function in prostate cancer.

To investigate the specificity and mechanisms of ERG-induced splicing regulation, we performed motif enrichment analysis on exons sequences that were alternately spliced by ERG using a collection of more than one hundred RNA-binding proteins (Anderson et al., 2012; Ray et al., 2013). We observed a significant enrichment of RBFOX motif upstream of ERG spliced-out exons **(Manuscript Figure 2A)**. Comparison between the known RNA-map of RBFOX2 **(Figure 39)** and our RBFOX RNA-map generated from ERG alternative splicing exons highlighted a similar enrichment of the RBFOX motif upstream of spliced-out exons. We hypothetized that both proteins may collaborate to control splicing. To investigate this, we performed RNA-seq experiments following RBFOX2 depletion and confirmed a highly significant overlap between RBFOX2- and ERG-splicing targets. Furthermore, comparison between common splicing targets highlighted that both proteins similarly regulated splicing. These observations led us to investigate the functional interaction between ERG and RBFOX2.

Concordant with this hypothesis, we demonstrated that ERG was able to directly interact with RBFOX2 *via* its C-terminal domain, the same domain that is required for ERG splicing activity. Conversely, we identified the C-terminal domain of RBFOX2 as required for ERG interaction. This observation is consistent with previous observations that showed this domain as required for most RBFOX interactions and for its splicing activity (Ying et al., 2017).

Because the C-terminal domain of ERG remains in oncogenic Ewing sarcoma fusions, we next assessed if EWS-FLI1 fusion was also able to bind RBFOX2 and to induce an RBFOX2-dependent splicing pattern. Immunoprecipitation experiments and RNA-sequencing confirmed that EWS-FLI1 fusion protein binds RBFOX2 and induces splicing alteration of a large set of exons. As expected, we found that spliced exons regulated by EWS-FLI1 were enriched for RBFOX binding motif and overlap between EWS-FLI1- and RBFOX2-dependent splicing targets is highly significant. Suprisingly, we observed that RBFOX binding motif was enriched upstream of EWS-FLI1 spliced-in exons, while RBFOX motif was enriched upstream of ERG spliced-out exons.

This fascinating observation led us to hypothetized that EWS-FLI1 and RBFOX2 might have an opposite role on splicing despite the functional interaction between these two proteins. Very strikingly, we demonstrated by RNA-seq analysis that, indeed, more than half of the common splicing targets between EWS-FLI1 and RBFOX2 are oppositely regulated. We have shown that the binding of RBFOX2 on its pre-mRNA target is increased following EWS-FLI1 depletion suggesting that EWS-FLI1 antagonizes RBFOX2 splicing function by repressing its pre-mRNA binding.

Using exon array, Selanathan and colleagues have shown that EWS-FLI1 modulated the exon expression pattern of Ewing sarcoma cells (Selvanathan et al., 2015). Consistent with their findings, we found that the most common splicing event altered by EWS-FLI1 was cassette exons. They indentified 120 genes with differential exon expression, 49 of them (40%) were also identified in our analysis ($p < 0.05$). Authors claimed that EWS-FLI1 is able to bind RNA, especially at exon-intron boundaries, using CLIP-seq experiments. However, we performed CLIP experiments following RNA labeling by radioactivity and found no RNA molecules bound by EWS-FLI1 protein.

II.   RBFOX2 plays a key role on ERG- and EWS-ETS-mediated splicing

RBFOX proteins are known as master splicing regulators in a tissue-specific manner but also play a major role in the epithelial-to-mesenchymal transition of cancer cells (Jin et al., 2003; Nakahata and Kawamoto, 2005; Warzecha et al., 2010). The splicing regulatory mechanism of RBFOX proteins is well described: RBFOX binding motif is enriched upstream of spliced-out exons and downstream of spliced-in exons (Jangi et al., 2014; Yeo et al., 2009; Zhang et al., 2008) as illustrated in **Figure 39**.

Figure 39: RBFOX2 RNA-map representation. RNA map illustrating the correlation between the RBP binding position and its impact on alternative splicing. Peaks represent the intensity of the RBP binding. Blue and yellow peaks represent the pre-mRNA associated with spliced-out and spliced-in, respectively. From B. Chabot et al., 2015.

We propose that the opposite effect observed between EWS-FLI1 and RBFOX2 might be due to RBFOX2 sequestration outside of its targets because of the low complexity region of EWS, which remains in EWS-ETS fusions. To test this hypothesis, it would be interesting to overexpress fluorescent-tagged RBFOX2 and EWS-FLI1 proteins to identify a putative co-localization and foci formation (Chong et al., 2018). In addition, we carry out large-scale analysis using CLIP sequencing data in control and EWS-FLI1-depleted cells to decipher the specificity of RBFOX2 inhibition. Indeed, we also observed that half of common targets between EWS-FLI1 and RBFOX2 are regulated similarly, suggesting specific features such as splicing sites or motif co-occurences.

Interestingly, in addition to its role on post-transcriptional regulation, a recent study has demonstrated that RBFOX2 is also implicated in transcription-related processes through a protein interaction with the polycomp repressive complex 2 (PRC2) (Wei et al., 2016). It demonstrated that RBFOX2 is required to coordinate PRC2 targeting thus inducing transcription repression. Because RBFOX2 is involved in gene expression regulation *via* the PRC2 complex, it would be very interesting to explore the role of RBFOX2 in the ERG- and EWS-FLI1-dependent transcription regulation.

Most studies focusing on splicing regulation aimed to decipher the splicing code of RNA binding proteins to predict their splicing functions. However, Damianov and colleagues have shown that RBFOX2 is part of a large complex called LASR for large assembly of splicing regulators (Damianov et al., 2016). This macromolecular complex is composed of several splicing factors such as hnRNPM, hnRNPH and DDX5. It has been shown that LASR complex extends the splicing repertoires by interconnexions between splicing-related proteins and tunes previously described splicing function. Together this study unravels a potential new mechanism of splicing regulation and new perspectives on the cascade of consequences if a component of this complex is functionally inactivated. We speculate that the interaction between EWS-FLI1 and RBFOX2 might partially disrupt the LASR complex, thus affecting LASR-dependent splicing outcomes. Additionally, RBFOX2 also interacts with several RBPs that are not included in the LASR complex, such as QKI. In this context, we plan to test the ability of RBFOX2 to bind proteins of the LASR complex and other RBFOX2 partners in the presence of EWS-FLI1 fusion. Furthermore, Selvanathan and colleagues have shown that EWS-FLI1 directly interacted with DDX5, a spliceosome-associated protein that is found in the LASR complex (Selvanathan et al., 2015). It would be worth testing the interaction between EWS-FLI1 and proteins of the LASR complex.

Intriguingly, we also demonstrated that QKI motif was enriched in EWS-FLI1-mediated exons and its function is antagonized by EWS-FLI1. QKI is also known to regulate the mesenchymal phenotype, in particular by inducing a mesenchymal-specific splicing program but also by promoting exon circularization and circular RNA structures (Conn et al., 2015). Very nicely, we found that QKI was highly and similarly enriched in RBFOX2-mediated exons, using motif enrichment analysis **(Figure 40)**. Both RBPs share a similar splicing mechanism and program. This observation confirms that RBFOX2 and QKI act together to regulate a common splicing program. We could wonder if the moiety of exons that are similarly regulated by EWS-FLI1 and RBFOX2 have a co-occurrence of the QKI motif. The repressive function of EWS-FLI1 on QKI splicing function might be also mediated by the formation of circular RNAs.

Figure 40: False Discovery Proportion (FDP) of motif enrichment analysis on RBFOX2-regulated exons. The threshold of 25% of FDP was choosen to control robust and reliable enrichment of RBPs. The same threshold is used in enrichement analysis such as Gene Set Enrichement Analysis (GSEA). RBFOX2 and QKI motif are framed in red.

"Handshakes and Fights: The Regulatory Interplay of RNA-Binding Proteins" (Dassi et al., 2017)
We have described a collaborative and an antagonist effect of ERG and EWS-FLI1, respectively, on RBFOX2 splicing function. It is well established that regulation of alternative splicing is mediated by an important regulatory network, which is prone to involve proteins with distinct functions. Three main models of splicing regulation by RBPs have been proposed: they can be either synergistic, competitive or mutual **(Figure 41)** (Dassi, 2017). For instance, RBFOX2 and MBNL1 cooperate on a large set of splicing targets to induce similar splicing program, in particular in stem cell differentiation (Venables et al., 2013a). In our context, we think that the splicing regulation induced by EWS-FLI1 is process that involves a complex network hub and the underlying mechanisms remain unclear.
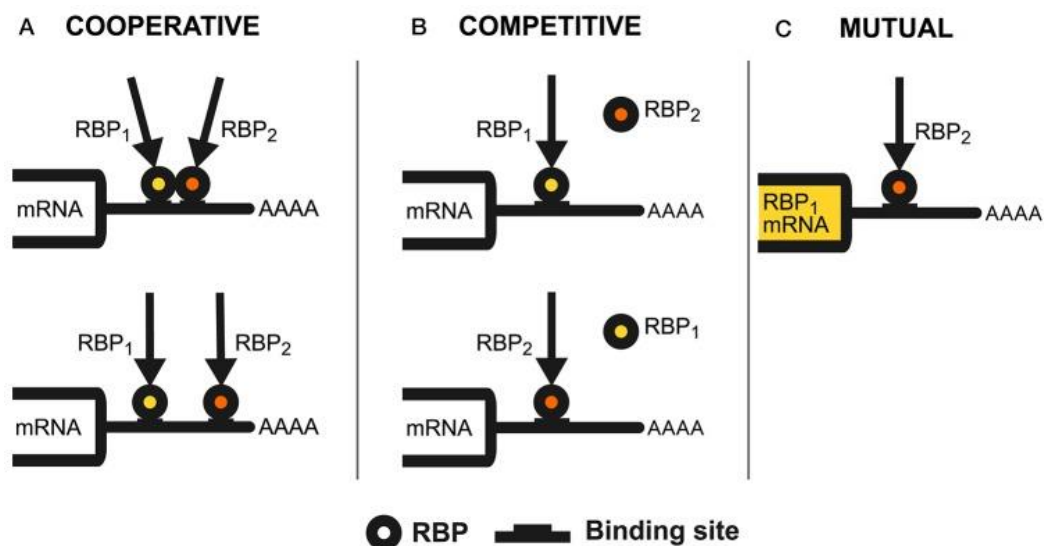
Figure 41: Interplay of RNA binding proteins to control splicing outcomes. Adapted from Dassi et al., 2017.

III. Splicing alterations participate in the cellular plasticity of Ewing sarcoma

Because Ewing sarcoma is not an epithelial cancer (carcinoma), the term EMT *stricto sensu* is not correct, we could refer to it as plasticity. Ewing sarcoma harbors a metastable phenotype, acquiring properties from both epithelial and mesenchymal cells. Plasticity is a key feature that is involved in Ewing sarcoma biology (Chaturvedi et al., 2012, 2014; Franzetti et al., 2017; Katschnig et al., 2017; Wiles et al., 2013). Several studies have shown that EWS-FLI1 depletion in Ewing cells leds to major phenotypic changes towards a mesenchymal-migrating cell phenotype. These observations suggest that EWS-FLI1 is required to respress the mesenchymal phenotype of the cell of origin and to induce cell growth and tumorigenesis.

Interestingly, the lab has recently revealed the existence of cell-cell heterogeneity of EWS-FLI1 expression. Putative stochastic expression of EWS-FLI1 allows Ewing cells to harbor migration properties as well as proliferative capacities. Indeed, cells expressing low levels of EWS-FLI1 reduce their proliferation rate in order to migrate from the tumor site, hence potentiate metastasis. On the contrary, cells expressing high levels of EWS-FLI1 increase their proliferation rate to expand the global tumor mass. Overall, the variation of EWS-FLI1 expression contributes to the plasticity observed in Ewing sarcoma and is responbile for the aggressive phenotype of these tumors by driving invasion at distal sites and expanding tumor volume (Franzetti et al., 2017). Interestingly, from our recent lab observations, we also found that cells that express very high level of EWS-FLI1 adapt their cellular processes and are not associated to a highly proliferative state. Together, we speculate that there is the existence of a dose of EWS-FLI1 expression that cells can tolerate, neither too low, nor too high.

Further experiments need to be performed to characterize mechanisms governing EWS-FLI1 expression variability and Ewing sarcoma plasticity. These results are necessary to better understand mechanisms of resistance in Ewing sarcoma and to develop adapted therapies based on these cellular features.

Beyond transcription, we next wondered if splicing alteration might be implicated in Ewing plasticity, hence underlying an important role on Ewing sarcoma biology. To assess the functional relevance of these findings, we focused on the *ADD3* gene, which harbors two transcripts differing on whether exon 14 is included or not. We chose this splicing event for multiple reasons including: (i) its regulation by EWS-FLI1, (ii) its role on cytoskeleton remodeling, a well known process involved in Ewing sarcoma biology, (iii) its previously described splicing pattern overexpressed in cancer cells (Eswaran et al., 2013; Langer et al., 2010) and associated to metastasis (Dutertre et al., 2010) and (iv) a potential additional evidence of the Ewing sarcoma cell of origin.

Indeed, we have shown that mesenchymal cells do not express exon 14 of *ADD3* and EWS-FLI1-depleted cells preferentially spliced-out exon 14 as mesenchymal cells. Therefore, we speculated that functional experiments on ADD3 isoforms would also provide insights on the cell of origin.

We designed three independent siRNAs targeting exon 14 of *ADD3*, as well as genome editing using CRISPR technology, to specifically inhibit this transcript. Both technologies abolished exon 14-contaning transcript expression. Immufluorescence experiments and spheroid invasion assay clearly demonstrated that cells depleted for exon 14-containing transcript of *ADD3* increased F-actin stress fibers formation and invasive capacities. These observations were completely independent of EWS-FLI1 expression **(Manuscript Figure S6B)** and HeLa cells have been used as a heterologous cell model to confirm these results (data not shown).

Altogether these data indicate that isoform of ADD3 contaning exon 14 is essential to repress the mesenchymal phenotype of Ewing sarcoma and to induce cell growth. *In vivo* experiments would, definetly add a great value to the work by combining xenograft injection of cells upon spliced-out of exon 14. This would provide new insights in the mechanisms of resistance to chemotherapy and might give perspectives to target Ewing sarcoma plasticity. It would be very exciting to investigate stem cell properties of these cells because the putative cell of origin of Ewing sarcoma is MSC and RBFOX2 has been previously linked to survival of embryonic stem cells (Yeo et al., 2009).

IV.   PacBio sequencing as a promising technology for splicing analysis

The current gold standard to study transcriptome-wide splicing alterations is to do RNA-seq experiments on an Illumina plateform. Several algorithms have been developed to identify differentially spliced events across multiple RNA-seq datasets. Tools are categorized into two main categories, the reconstruction-based tools and the local event-based tools. Cufflinks, developed by Cole Trapnell at the Broad Institute, is one of the reconstruction-based tools, which assembles full-length transcripts and estimates their relative abundance. In my opinion, the reconstruction-based tools are not well suited for short reads sequencing data because it is a challenging task to reconstruct a full-length transcript (median size of 2kb) with 100bp reads. Most of the splicing analyses now consider local-event-based tools such as rMATS to robustly find differentially expressed events, such as cassette exons. By using RNA-seq, we were able to identify transcriptome-wide alternative splicing targets regulated by ERG and EWS-FLI1. We validated a subset of differentially spliced events using RT-PCR and found a high correlation between both technologies, indicating that our pipeline analysis was robust and reliable for our purpose.

PacBio sequencing is a promising option to study splicing alteration changes. Long-reads sequencing experiments provide a great opportunity to map full-length transcripts on annotated transcript databases but also to discover unannotated transcripts. In this context, we performed 4 SMRT cells (equivalent of Illumina flow cells) per condition on control cells and EWS-FLI1-depleted cells. Interestingly, **Figure 42 & 43** illustrate the processivity of the PacBio polymerase as compared to Illumina. The region highlighted is a homopolymeric region of poly-G repeats, in which we observed a considerable drop of the depth of coverage in Illumina sequencing data.
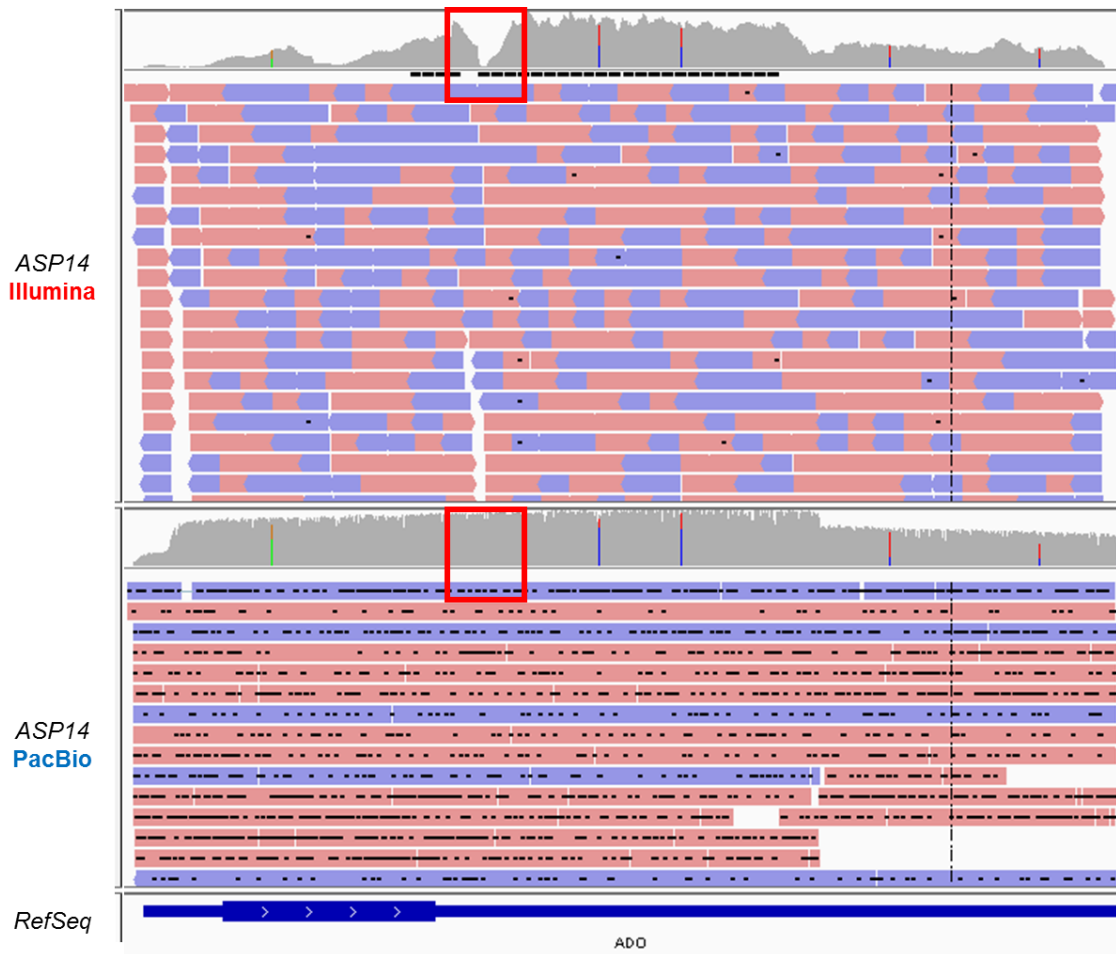
Figure 42: IGV screenshot of ADO gene. Red and blue boxes are individual reads. We can nicely see that PacBio sequencing has full-length transcript reads as compared to short-reads for Illumina. Represented region is around 3.5kb. Red framed region is zoomed in Figure 43.
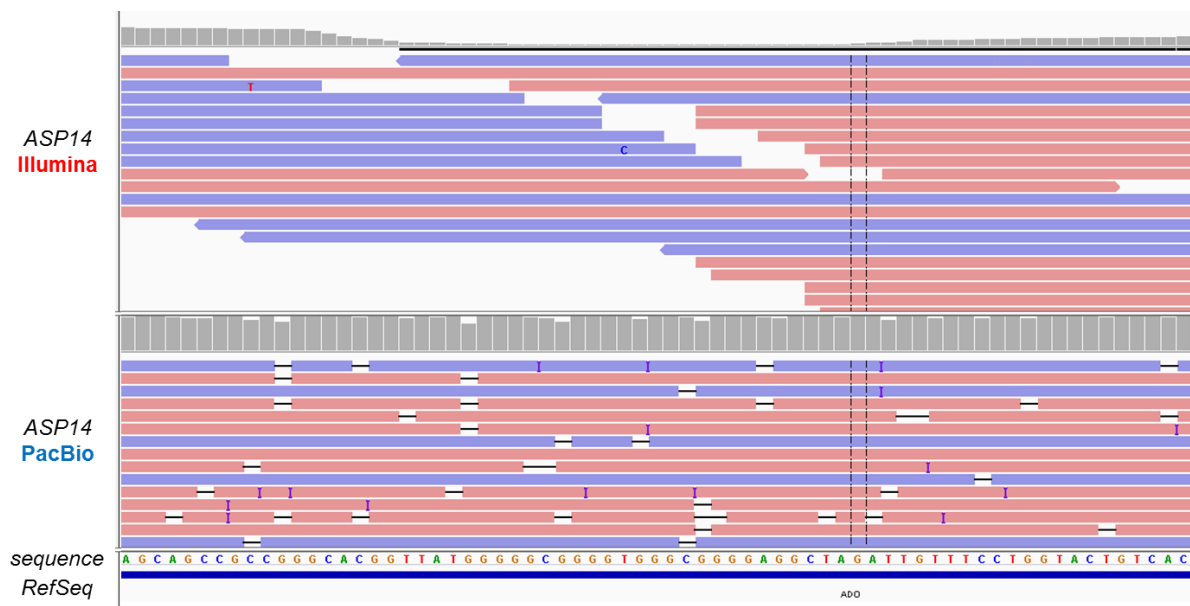
Figure 43: IGV screenshot of the red framed region in Figure 42. Genomic sequence is displayed at the bottom of the image showing an homopolymeric region of poly-G repeats.

In addition, we demonstrated that, as the Illumina technology, the PacBio technology detects alternatively spliced exons accurately **(Figure 44)** and provides a great alternative to study alternative splicing alterations in cancer.
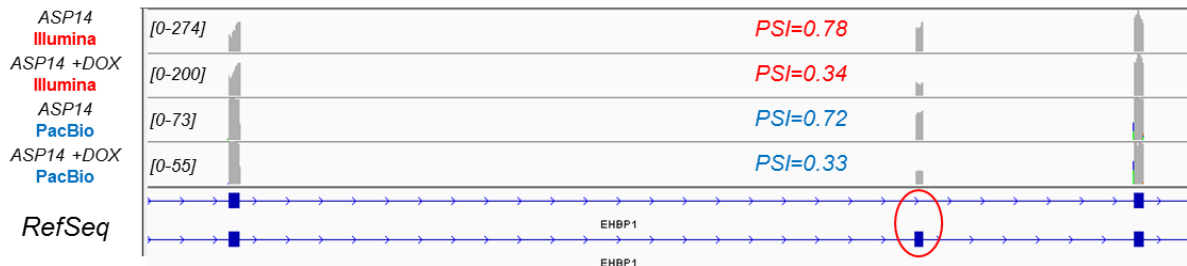
Figure 44: IGV screenshot of EHBP1 gene showing an alternatively spliced exon upon EWS-FLI1 depletion. Alternatively spliced exon is framed in red.

V.   **Significance and model**

Collectively, we have shown that, in addition to their role as transcription factors, ERG proteins (ERG, FLI1 and FEV) control alternative splicing programs through their interactions with the master splicing regulator RBFOX2. In Ewing sarcoma, EWS-FLI1 oncoprotein interfers with RBFOX2 to antagonize its splicing function. Our study highlights how oncogenic transcription factors hijack splicing regulators to drive aberrant splicing programs. This work provides evidences that splicing alterations induced by EWS-FLI1 participate in the cellular plasticity, thus revealing a novel mechanism involved in Ewing sarcoma biology.

Altogether, this new identification of ERG splicing function expands the spectrum of splicing abnormalities in cancer and provide evidences that transcription factors affect splicing outcomes. We propose a putative model of alternative splicing regulation by ERG proteins and the EWS-FLI1 oncoprotein **(Figure 45)**. Our study gives new relevant perspectives on ERG-rearranged protein functions in cancer and how these cancers might exhibit splicing alterations.
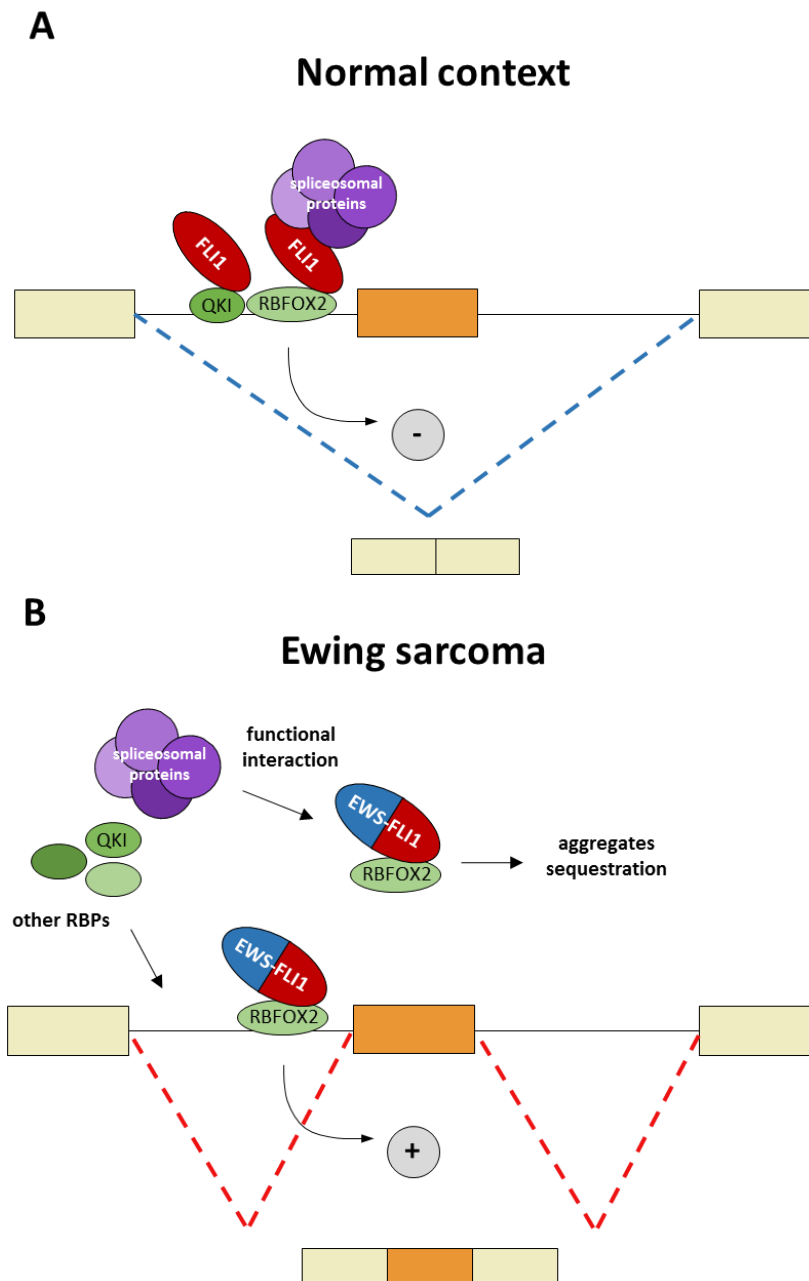
Figure 45: Schematic model of alternative splicing regulation by ERG proteins (ERG, FLI1 and FEV) and EWS-FLI1 oncoprotein. **(A)** ERG, and therefore FLI1, modify the splicing landscape of a cell. FLI1 interacts with RNA binding proteins RBFOX2 and QKI to regulate common splicing programs. FLI1 regulates splicing in a similar manner as RBFOX2 and QKI. It has been shown that RBFOX2 and QKI physically interact. We hypothetized that FLI1 is part of a large complex including RBFOX2 and QKI, as well as known RBFOX2 interactors such as LASR complex and spliceosomal proteins to favor splicing. **(B)** EWS-FLI1 interacts with RBFOX2 but no longer with QKI. We speculated that the RBFOX2/QKI complex is disrupted by the interaction between EWS-FLI1 and RBFOX2, thus explaining the opposite effect seen on splicing by EWS-FLI1 and RBFOX2 (and QKI). It would be interesting to test if RBFOX2 is sequestrated by EWS-FLI1 into protein aggregates but also if spliceosomal proteins, such as U1-70K and U2F65, are still able to bind EWS-FLI1.

REFERENCES

- Acloque, H., Adams, M.S., Fishwick, K., Bronner-Fraser, M., and Nieto, M.A. (2009). Epithelial-mesenchymal transitions: the importance of changing cell state in development and disease. J. Clin. Invest. *119*, 1438–1449.

- Ahlstrom, J.D., and Erickson, C.A. (2009). The neural crest epithelial-mesenchymal transition in 4D: a `tail' of multiple non-obligatory cellular mechanisms. Dev. Camb. Engl. *136*, 1801–1812.

- de Alava, E., Lessnick, S.L., and Sorensen, P.H. (2013). WHO Classification of Tumours of Soft Tissue and Bone. 305–309.

- Alberti, S., Halfmann, R., King, O., Kapila, A., and Lindquist, S. (2009). A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. Cell *137*, 146–158.

- Allemand, E., Myers, M.P., Garcia-Bernardo, J., Harel-Bellan, A., Krainer, A.R., and Muchardt, C. (2016). A Broad Set of Chromatin Factors Influences Splicing. PLoS Genet. *12*.

- Alsafadi, S., Houy, A., Battistella, A., Popova, T., Wassef, M., Henry, E., Tirode, F., Constantinou, A., Piperno-Neumann, S., Roman-Roman, S., et al. (2016). Cancer-associated SF3B1 mutations affect alternative splicing by promoting alternative branchpoint usage. Nat. Commun. *7*, 10615.

- 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.

- Anczuków, O., and Krainer, A.R. (2016). Splicing-factor alterations in cancers. RNA *22*, 1285–1301.

- Anczuków, O., Rosenberg, A.Z., Akerman, M., Das, S., Zhan, L., Karni, R., Muthuswamy, S.K., and Krainer, A.R. (2012). THE SPLICING FACTOR SRSF1 REGULATES APOPTOSIS AND PROLIFERATION TO PROMOTE MAMMARY EPITHELIAL CELL TRANSFORMATION. Nat. Struct. Mol. Biol. *19*, 220–228.

- Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinforma. Oxf. Engl. *31*, 166–169.

- Anderson, E.S., Lin, C.-H., Xiao, X., Stoilov, P., Burge, C.B., and Black, D.L. (2012). The cardiotonic steroid digitoxin regulates alternative splicing through depletion of the splicing factors SRSF3 and TRA2B. RNA *18*, 1041–1049.

- Anderson, N.D., de Borja, R., Young, M.D., Fuligni, F., Rosic, A., Roberts, N.D., Hajjar, S., Layeghifard, M., Novokmet, A., Kowalski, P.E., et al. (2018). Rearrangement bursts generate canonical gene fusions in bone and soft tissue tumors. Science *361*, eaam8419.

- Arumugam, T., Ramachandran, V., Fournier, K.F., Wang, H., Marquis, L., Abbruzzese, J.L., Gallick, G.E., Logsdon, C.D., McConkey, D.J., and Choi, W. (2009). Epithelial to mesenchymal transition contributes to drug resistance in pancreatic cancer. Cancer Res. *69*, 5820–5828.

- Auboeuf, D., Hönig, A., Berget, S.M., and O'Malley, B.W. (2002). Coordinate regulation of transcription and splicing by steroid receptor coregulators. Science *298*, 416–419.

- Auboeuf, D., Batsché, E., Dutertre, M., Muchardt, C., and O'Malley, B.W. (2007). Coregulators: transducing signal from transcription to alternative splicing. Trends Endocrinol. Metab. TEM *18*, 122–129.

- Aurias, A. (1983). Chromosomal translocations in Ewing's sarcoma. N. Engl. J. Med. *309*, 496–498.

- Bainbridge, M.N., Warren, R.L., Hirst, M., Romanuik, T., Zeng, T., Go, A., Delaney, A., Griffith, M., Hickenbotham, M., Magrini, V., et al. (2006). Analysis of the prostate cancer cell line LNCaP transcriptome using a sequencing-by-synthesis approach. BMC Genomics *7*, 246.
- 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.
- Baliko, F., Bright, T., Poon, R., Cohen, B., Egan, S.E., and Alman, B.A. (2007). Inhibition of notch signaling induces neural differentiation in Ewing sarcoma. Am. J. Pathol. *170*, 1686–1694.
- Barbosa-Morais, N.L., Irimia, M., Pan, Q., Xiong, H.Y., Gueroussov, S., Lee, L.J., Slobodeniuc, V., Kutter, C., Watt, S., Colak, R., et al. (2012). The evolutionary landscape of alternative splicing in vertebrate species. Science *338*, 1587–1593.
- Batsché, E., Yaniv, M., and Muchardt, C. (2006). The human SWI/SNF subunit Brm is a regulator of alternative splicing. Nat. Struct. Mol. Biol. *13*, 22–29.
- Bebee, T.W., Park, J.W., Sheridan, K.I., Warzecha, C.C., Cieply, B.W., Rohacek, A.M., Xing, Y., and Carstens, R.P. (2015). The splicing regulators Esrp1 and Esrp2 direct an epithelial splicing program essential for mammalian development. ELife *4*.
- Bennani-Baiti, I.M., Machado, I., Llombart-Bosch, A., and Kovar, H. (2012). Lysine-specific demethylase 1 (LSD1/KDM1A/AOF2/BHC110) is expressed and is an epigenetic drug target in chondrosarcoma, Ewing's sarcoma, osteosarcoma, and rhabdomyosarcoma. Hum. Pathol. *43*, 1300–1307.
- Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., Hall, K.P., Evers, D.J., Barnes, C.L., Bignell, H.R., et al. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. Nature *456*, 53–59.
- Berget, S.M., Moore, C., and Sharp, P.A. (1977). Spliced segments at the 5' terminus of adenovirus 2 late mRNA. Proc. Natl. Acad. Sci. U. S. A. *74*, 3171–3175.
- Bertolotti, A., Melot, T., Acker, J., Vigneron, M., Delattre, O., and Tora, L. (1998). EWS, but not EWS-FLI-1, is associated with both TFIID and RNA polymerase II: interactions between two members of the TET family, EWS and hTAFII68, and subunits of TFIID and RNA polymerase II complexes. Mol. Cell. Biol. *18*, 1489–1497.
- Beyer, A.L., and Osheim, Y.N. (1988). Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. Genes Dev. *2*, 754–765.
- Bielli, P., Bordi, M., Di Biasio, V., and Sette, C. (2014). Regulation of BCL-X splicing reveals a role for the polypyrimidine tract binding protein (PTBP1/hnRNP I) in alternative 5' splice site selection. Nucleic Acids Res. *42*, 12070–12081.
- Blencowe, B.J. (2006). Alternative splicing: new insights from global analyses. Cell *126*, 37–47.
- Boeva, V., Surdez, D., Guillon, N., Tirode, F., Fejes, A.P., Delattre, O., and Barillot, E. (2010). De novo motif identification improves the accuracy of predicting transcription factor binding sites in ChIP-Seq data analysis. Nucleic Acids Res. *38*, e126.
- Boeva, V., Louis-Brennetot, C., Peltier, A., Durand, S., Pierre-Eugène, C., Raynal, V., Etchevers, H.C., Thomas, S., Lermine, A., Daudigeos-Dubus, E., et al. (2017). Heterogeneity of neuroblastoma cell identity defined by transcriptional circuitries. Nat. Genet. *49*, 1408–1413.
- Boise, L.H., González-García, M., Postema, C.E., Ding, L., Lindsten, T., Turka, L.A., Mao, X., Nuñez, G., and Thompson, C.B. (1993). bcl-x, a bcl-2-related gene that functions as a dominant regulator of apoptotic cell death. Cell *74*, 597–608.

- Bolender, D.L., and Markwald, R.R. (1979). Epithelial-mesenchymal transformation in chick atrioventricular cushion morphogenesis. Scan. Electron Microsc. 313–321.
- Bonomi, S., di Matteo, A., Buratti, E., Cabianca, D.S., Baralle, F.E., Ghigna, C., and Biamonti, G. (2013). HnRNP A1 controls a splicing regulatory circuit promoting mesenchymal-to-epithelial transition. Nucleic Acids Res. *41*, 8665–8679.
- 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.
- Boulay, G., Volorio, A., Iyer, S., Broye, L.C., Stamenkovic, I., Riggi, N., and Rivera, M.N. (2018). Epigenome editing of microsatellite repeats defines tumor-specific enhancer functions and dependencies. Genes Dev.
- Boyer, B., and Thiery, J.P. (1993). Epithelium-mesenchyme interconversion as example of epithelial plasticity. APMIS Acta Pathol. Microbiol. Immunol. Scand. *101*, 257–268.
- Braeutigam, C., Rago, L., Rolke, A., Waldmeier, L., Christofori, G., and Winter, J. (2014). The RNA-binding protein Rbfox2: an essential regulator of EMT-driven alternative splicing and a mediator of cellular invasion. Oncogene *33*, 1082–1092.
- Breathnach, R., and Chambon, P. (1981). Organization and expression of eucaryotic split genes coding for proteins. Annu. Rev. Biochem. *50*, 349–383.
- Breathnach, R., Benoist, C., O'Hare, K., Gannon, F., and Chambon, P. (1978). Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. Proc. Natl. Acad. Sci. U. S. A. *75*, 4853–4857.
- 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.
- Brosseau, J.-P., Lucier, J.-F., Nwilati, H., Thibault, P., Garneau, D., Gendron, D., Durand, M., Couture, S., Lapointe, E., Prinos, P., et al. (2014). Tumor microenvironment-associated modifications of alternative splicing. RNA N. Y. N *20*, 189–201.
- Buljan, M., Chalancon, G., Eustermann, S., Wagner, G.P., Fuxreiter, M., Bateman, A., and Babu, M.M. (2012). Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks. Mol. Cell *46*, 871–883.
- Burd, C.G., and Dreyfuss, G. (1994). Conserved structures and diversity of functions of RNA-binding proteins. Science *265*, 615–621.
- Cáceres, J.F., and Kornblihtt, A.R. (2002). Alternative splicing: multiple control mechanisms and involvement in human disease. Trends Genet. TIG *18*, 186–193.
- Cáceres, J.F., Screaton, G.R., and Krainer, A.R. (1998). A specific subset of SR proteins shuttles continuously between the nucleus and the cytoplasm. Genes Dev. *12*, 55–66.
- Cano, A., Pérez-Moreno, M.A., Rodrigo, I., Locascio, A., Blanco, M.J., del Barrio, M.G., Portillo, F., and Nieto, M.A. (2000). The transcription factor snail controls epithelial-mesenchymal transitions by repressing E-cadherin expression. Nat. Cell Biol. *2*, 76–83.
- Caplan, A.I. (1991). Mesenchymal stem cells. J. Orthop. Res. Off. Publ. Orthop. Res. Soc. *9*, 641–650.
- 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.

• Castillero-Trejo, Y., Eliazer, S., Xiang, L., Richardson, J.A., and Ilaria, R.L. (2005). Expression of the EWS/FLI-1 oncogene in murine primary bone-derived cells Results in EWS/FLI-1-dependent, ewing sarcoma-like tumors. Cancer Res. *65*, 8698–8705.

• Chan, C.K.F., Gulati, G.S., Sinha, R., Tompkins, J.V., Lopez, M., Carter, A.C., Ransom, R.C., Reinisch, A., Wearda, T., Murphy, M., et al. (2018). Identification of the Human Skeletal Stem Cell. Cell *175*, 43-56.e21.

• 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.

• Chaturvedi, A., Hoffman, L.M., Jensen, C.C., Lin, Y.-C., Grossmann, A.H., Randall, R.L., Lessnick, S.L., Welm, A.L., and Beckerle, M.C. (2014). Molecular dissection of the mechanism by which EWS/FLI expression compromises actin cytoskeletal integrity and cell adhesion in Ewing sarcoma. Mol. Biol. Cell *25*, 2695–2709.

• Chaudhury, A., Cheema, S., Fachini, J.M., Kongchan, N., Lu, G., Simon, L.M., Wang, T., Mao, S., Rosen, D.G., Ittmann, M.M., et al. (2016). CELF1 is a central node in post-transcriptional regulatory programmes underlying EMT. Nat. Commun. *7*, 13362.

• Chen, L., Bush, S.J., Tovar-Corona, J.M., Castillo-Morales, A., and Urrutia, A.O. (2014). Correcting for differential transcript coverage reveals a strong relationship between alternative splicing and organism complexity. Mol. Biol. Evol. *31*, 1402–1413.

• Chen, P.B., Chen, H.V., Acharya, D., Rando, O.J., and Fazzio, T.G. (2015). R loops regulate promoter-proximal chromatin architecture and cellular differentiation. Nat. Struct. Mol. Biol. *22*, 999–1007.

• Chong, S., Dugast-Darzacq, C., Liu, Z., Dong, P., Dailey, G.M., Cattoglio, C., Heckert, A., Banala, S., Lavis, L., Darzacq, X., et al. (2018). Imaging dynamic and selective low-complexity domain interactions that control gene transcription. Science *361*, eaar2555.

• Chow, L.T., Gelinas, R.E., Broker, T.R., and Roberts, R.J. (1977). An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. Cell *12*, 1–8.

• Cidre-Aranaz, F., and Alonso, J. (2015). EWS/FLI1 Target Genes and Therapeutic Opportunities in Ewing Sarcoma. Front. Oncol. *5*, 162.

• Climente-González, H., Porta-Pardo, E., Godzik, A., and Eyras, E. (2017). The Functional Impact of Alternative Splicing in Cancer. Cell Rep. *20*, 2215–2226.

• Collesi, C., Santoro, M.M., Gaudino, G., and Comoglio, P.M. (1996). A splicing variant of the RON transcript induces constitutive tyrosine kinase activity and an invasive phenotype. Mol. Cell. Biol. *16*, 5518–5526.

• Comijn, J., Berx, G., Vermassen, P., Verschueren, K., van Grunsven, L., Bruyneel, E., Mareel, M., Huylebroeck, D., and van Roy, F. (2001). The Two-Handed E Box Binding Zinc Finger Protein SIP1 Downregulates E-Cadherin and Induces Invasion. Mol. Cell *7*, 1267–1278.

• Conn, S.J., Pillman, K.A., Toubia, J., Conn, V.M., Salmanidis, M., Phillips, C.A., Roslan, S., Schreiber, A.W., Gregory, P.A., and Goodall, G.J. (2015). The RNA Binding Protein Quaking Regulates Formation of circRNAs. Cell *160*, 1125–1134.

• Couthouis, J., Hart, M.P., Shorter, J., DeJesus-Hernandez, M., Erion, R., Oristano, R., Liu, A.X., Ramos, D., Jethava, N., Hosangadi, D., et al. (2011). A yeast functional screen predicts new candidate ALS disease genes. Proc. Natl. Acad. Sci. U. S. A. *108*, 20881–20890.

- Cramer, P., Cáceres, J.F., Cazalla, D., Kadener, S., Muro, A.F., Baralle, F.E., and Kornblihtt, A.R. (1999). Coupling of Transcription with Alternative Splicing: RNA Pol II Promoters Modulate SF2/ASF and 9G8 Effects on an Exonic Splicing Enhancer. Mol. Cell *4*, 251–258.

- Cramer, P., Srebrow, A., Kadener, S., Werbajh, S., de la Mata, M., Melen, G., Nogués, G., and Kornblihtt, A.R. (2001). Coordination between transcription and pre-mRNA processing. FEBS Lett. *498*, 179–182.

- 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.

- Da Cruz, S., and Cleveland, D.W. (2011). Understanding the role of TDP-43 and FUS/TLS in ALS and beyond. Curr. Opin. Neurobiol. *21*, 904–919.

- Daguenet, E., Dujardin, G., and Valcárcel, J. (2015). The pathogenicity of splicing defects: mechanistic insights into pre-mRNA processing inform novel therapeutic approaches. EMBO Rep. *16*, 1640–1655.

- Damianov, A., Ying, Y., Lin, C.-H., Lee, J.-A., Tran, D., Vashisht, A.A., Bahrami-Samani, E., Xing, Y., Martin, K.C., Wohlschlegel, J.A., et al. (2016). Rbfox Proteins Regulate Splicing as Part of a Large Multiprotein Complex LASR. Cell *165*, 606–619.

- Darman, R.B., Seiler, M., Agrawal, A.A., Lim, K.H., Peng, S., Aird, D., Bailey, S.L., Bhavsar, E.B., Chan, B., Colla, S., et al. (2015). Cancer-Associated SF3B1 Hotspot Mutations Induce Cryptic 3' Splice Site Selection through Use of a Different Branch Point. Cell Rep. *13*, 1033–1045.

- Das, S., Anczuków, O., Akerman, M., and Krainer, A.R. (2012). Oncogenic Splicing Factor SRSF1 Is a Critical Transcriptional Target of MYC. Cell Rep. *1*, 110–117.

- Dassi, E. (2017). Handshakes and Fights: The Regulatory Interplay of RNA-Binding Proteins. Front. Mol. Biosci. *4*.

- David, C.J., and Manley, J.L. (2010). Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. Genes Dev. *24*, 2343–2364.

- DeBoever, C., Ghia, E.M., Shepard, P.J., Rassenti, L., Barrett, C.L., Jepsen, K., Jamieson, C.H.M., Carson, D., Kipps, T.J., and Frazer, K.A. (2015). Transcriptome sequencing reveals potential mechanism of cryptic 3' splice site selection in SF3B1-mutated cancers. PLoS Comput. Biol. *11*, e1004105.

- 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.

- Dittmar, K.A., Jiang, P., Park, J.W., Amirikian, K., Wan, J., Shen, S., Xing, Y., and Carstens, R.P. (2012). Genome-wide determination of a broad ESRP-regulated posttranscriptional network by high-throughput sequencing. Mol. Cell. Biol. *32*, 1468–1482.

- Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., et al. (2012). Landscape of transcription in human cells. Nature *489*, 101–108.

- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

- Doyle, L.A. (2014). Sarcoma classification: An update based on the 2013 World Health Organization Classification of Tumors of Soft Tissue and Bone. Cancer *120*, 1763–1774.

- Dujardin, G., Lafaille, C., de la Mata, M., Marasco, L.E., Muñoz, M.J., Le Jossic-Corcos, C., Corcos, L., and Kornblihtt, A.R. (2014). How Slow RNA Polymerase II Elongation Favors Alternative Exon Skipping. Mol. Cell *54*, 683–690.

- Dutertre, M., Lacroix-Triki, M., Driouch, K., de la Grange, P., Gratadou, L., Beck, S., Millevoi, S., Tazi, J., Lidereau, R., Vagner, S., et al. (2010). Exon-Based Clustering of Murine Breast Tumor Transcriptomes Reveals Alternative Exons Whose Expression Is Associated with Metastasis. Cancer Res. *70*, 896–905.

- Early, P., Huang, H., Davis, M., Calame, K., and Hood, L. (1980). An immunoglobulin heavy chain variable region gene is generated from three segments of DNA: VH, D and JH. Cell *19*, 981–992.

- Erkizan, H.V., Schneider, J.A., Sajwan, K., Graham, G.T., Griffin, B., Chasovskikh, S., Youbi, S.E., Kallarakal, A., Chruszcz, M., Padmanabhan, R., et al. (2015). RNA helicase A activity is inhibited by oncogenic transcription factor EWS-FLI1. Nucleic Acids Res. *43*, 1069–1080.

- Eswaran, J., Horvath, A., Godbole, S., Reddy, S.D., Mudvari, P., Ohshiro, K., Cyanam, D., Nair, S., Fuqua, S.A.W., Polyak, K., et al. (2013). RNA sequencing of cancer reveals novel splicing alterations. Sci. Rep. *3*.

- Ewing, J. (1921). Diffuse endothelioma of bone. Proc. N. Y. Pathol. Soc. 17–24.

- Ezkurdia, I., Rodriguez, J.M., Carrillo-de Santa Pau, E., Vázquez, J., Valencia, A., and Tress, M.L. (2015). Most highly expressed protein-coding genes have a single dominant isoform. J. Proteome Res. *14*, 1880–1887.

- Fairbrother, W.G., Yeh, R.-F., Sharp, P.A., and Burge, C.B. (2002). Predictive identification of exonic splicing enhancers in human genes. Science *297*, 1007–1013.

- Fici, P., Gallerani, G., Morel, A.-P., Mercatali, L., Ibrahim, T., Scarpi, E., Amadori, D., Puisieux, A., Rigaud, M., and Fabbri, F. (2017). Splicing factor ratio as an index of epithelial-mesenchymal transition and tumor aggressiveness in breast cancer. Oncotarget *8*, 2423–2436.

- Foulon, S., Brennan, B., Gaspar, N., Dirksen, U., Jeys, L., Cassoni, A., Claude, L., Seddon, B., Marec-Berard, P., Whelan, J., et al. (2016). Can postoperative radiotherapy be omitted in localised standard-risk Ewing sarcoma? An observational study of the Euro-E.W.I.N.G group. Eur. J. Cancer Oxf. Engl. 1990 *61*, 128–136.

- Franchi, A., Pasquinelli, G., Cenacchi, G., Della Rocca, C., Gambini, C., Bisceglia, M., Martinelli, G.N., and Santucci, M. (2001). Immunohistochemical and ultrastructural investigation of neural differentiation in Ewing sarcoma/PNET of bone and soft tissues. Ultrastruct. Pathol. *25*, 219–225.

- 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.

- Friedenstein, A.J., Chailakhjan, R.K., and Lalykina, K.S. (1970). The development of fibroblast colonies in monolayer cultures of guinea-pig bone marrow and spleen cells. Cell Tissue Kinet. *3*, 393–403.

- G Hendrickson, D., Kelley, D.R., Tenen, D., Bernstein, B., and Rinn, J.L. (2016). Widespread RNA binding by chromatin-associated proteins. Genome Biol. *17*, 28.

- Gang, E.J., Jeong, J.A., Hong, S.H., Hwang, S.H., Kim, S.W., Yang, I.H., Ahn, C., Han, H., and Kim, H. (2004). Skeletal myogenic differentiation of mesenchymal stem cells isolated from human umbilical cord blood. Stem Cells Dayt. Ohio *22*, 617–624.

- 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.

- Garalde, D.R., Snell, E.A., Jachimowicz, D., Sipos, B., Lloyd, J.H., Bruce, M., Pantic, N., Admassu, T., James, P., Warland, A., et al. (2018). Highly parallel direct RNA sequencing on an array of nanopores. Nat. Methods *15*, 201–206.

- Garg, M. (2013). Epithelial-mesenchymal transition - activating transcription factors - multifunctional regulators in cancer. World J. Stem Cells *5*, 188–195.

- Garneau, D., Revil, T., Fisette, J.-F., and Chabot, B. (2005). Heterogeneous nuclear ribonucleoprotein F/H proteins modulate the alternative splicing of the apoptotic mediator Bcl-x. J. Biol. Chem. *280*, 22641–22650.

- Garnett, M.J., Edelman, E.J., Heidorn, S.J., Greenman, C.D., Dastur, A., Lau, K.W., Greninger, P., Thompson, I.R., Luo, X., Soares, J., et al. (2012). Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature *483*, 570–575.

- 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.

- Ghigna, C., Moroni, M., Porta, C., Riva, S., and Biamonti, G. (1998). Altered expression of heterogenous nuclear ribonucleoproteins and SR factors in human colon adenocarcinomas. Cancer Res. *58*, 5818–5824.

- Ghigna, C., Giordano, S., Shen, H., Benvenuto, F., Castiglioni, F., Comoglio, P.M., Green, M.R., Riva, S., and Biamonti, G. (2005). Cell Motility Is Controlled by SF2/ASF through Alternative Splicing of the Ron Protooncogene. Mol. Cell *20*, 881–890.

- Gilbert, W. (1978). Why genes in pieces? Nature *271*, 501.

- Gilbert, L.A., Horlbeck, M.A., Adamson, B., Villalta, J.E., Chen, Y., Whitehead, E.H., Guimaraes, C., Panning, B., Ploegh, H.L., Bassik, M.C., et al. (2014). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell *159*, 647–661.

- Girard, C., Will, C.L., Peng, J., Makarov, E.M., Kastner, B., Lemm, I., Urlaub, H., Hartmuth, K., and Lührmann, R. (2012). Post-transcriptional spliceosomes are retained in nuclear speckles until splicing completion. Nat. Commun. *3*, 994.

- Gonzalez, I., Munita, R., Agirre, E., Dittmer, T.A., Gysling, K., Misteli, T., and Luco, R.F. (2015). A lncRNA regulates alternative splicing via establishment of a splicing-specific chromatin signature. Nat. Struct. Mol. Biol. *22*, 370–376.

- Gonzàlez-Porta, M., Frankish, A., Rung, J., Harrow, J., and Brazma, A. (2013). Transcriptome analysis of human tissues and cell lines reveals one dominant transcript per gene. Genome Biol. *14*, R70.

- Goodwin, S., McPherson, J.D., and McCombie, W.R. (2016). Coming of age: ten years of next-generation sequencing technologies. Nat. Rev. Genet. *17*, 333–351.

- 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.

- Gout, S., Brambilla, E., Boudria, A., Drissi, R., Lantuejoul, S., Gazzeri, S., and Eymin, B. (2012). Abnormal expression of the pre-mRNA splicing regulators SRSF1, SRSF2, SRPK1 and SRPK2 in non small cell lung carcinoma. PloS One *7*, e46539.

- van Groningen, T., Koster, J., Valentijn, L.J., Zwijnenburg, D.A., Akogul, N., Hasselt, N.E., Broekmans, M., Haneveld, F., Nowakowska, N.E., Bras, J., et al. (2017). Neuroblastoma is

composed of two super-enhancer-associated differentiation states. Nat. Genet. *49*, 1261–1266.

• 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.

• Gui, J.F., Lane, W.S., and Fu, X.D. (1994). A serine kinase regulates intracellular localization of splicing factors in the cell cycle. Nature *369*, 678–682.

• 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.

• Guillouf, C., Gallais, I., and Moreau-Gachelin, F. (2006). Spi-1/PU.1 oncoprotein affects splicing decisions in a promoter binding-dependent manner. J. Biol. Chem. *281*, 19145–19155.

• Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jungkamp, A.-C., Munschauer, M., et al. (2010). PAR-CliP--a method to identify transcriptome-wide the binding sites of RNA binding proteins. J. Vis. Exp. JoVE.

• Hall, M.P., Nagel, R.J., Fagg, W.S., Shiue, L., Cline, M.S., Perriman, R.J., Donohue, J.P., and Ares, M. (2013). Quaking and PTB control overlapping splicing regulatory networks during muscle cell differentiation. RNA N. Y. N *19*, 627–638.

• Hallier, M., Tavitian, A., and Moreau-Gachelin, F. (1996). The Transcription Factor Spi-1/PU.1 Binds RNA and Interferes with the RNA-binding Protein p54. J. Biol. Chem. *271*, 11177–11181.

• Hallier, M., Lerga, A., Barnache, S., Tavitian, A., and Moreau-Gachelin, F. (1998). The transcription factor Spi-1/PU.1 interacts with the potential splicing factor TLS. J. Biol. Chem. *273*, 4838–4842.

• Han, H., Braunschweig, U., Gonatopoulos-Pournatzis, T., Weatheritt, R.J., Hirsch, C.L., Ha, K.C.H., Radovani, E., Nabeel-Shah, S., Sterne-Weiler, T., Wang, J., et al. (2017). Multilayered Control of Alternative Splicing Regulatory Networks by Transcription Factors. Mol. Cell *65*, 539-553.e7.

• Han, T.W., Kato, M., Xie, S., Wu, L.C., Mirzaei, H., Pei, J., Chen, M., Xie, Y., Allen, J., Xiao, G., et al. (2012). Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies. Cell *149*, 768–779.

• Harrison, A.F., and Shorter, J. (2017). RNA-binding proteins with prion-like domains in health and disease. Biochem. J. *474*, 1417–1438.

• Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., et al. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res. *22*, 1760–1774.

• Hay, E. (1968). Organization and fine structure of epithelium and mesenchyme in the developing chick embryo. Baltim. Williams Wilkins 31–55.

• Hay, E.D. (1995). An overview of epithelio-mesenchymal transformation. Acta Anat. (Basel) *154*, 8–20.

• Hegele, A., Kamburov, A., Grossmann, A., Sourlis, C., Wowro, S., Weimann, M., Will, C.L., Pena, V., Lührmann, R., and Stelzl, U. (2012). Dynamic protein-protein interaction wiring of the human spliceosome. Mol. Cell *45*, 567–580.

REFERENCES

- Hirose, Y., Tacke, R., and Manley, J.L. (1999). Phosphorylated RNA polymerase II stimulates pre-mRNA splicing. Genes Dev. *13*, 1234–1239.
- Hoell, J.I., Larsson, E., Runge, S., Nusbaum, J.D., Duggimpudi, S., Farazi, T.A., Hafner, M., Borkhardt, A., Sander, C., and Tuschl, T. (2011). RNA targets of wild-type and mutant FET family proteins. Nat. Struct. Mol. Biol. *18*, 1428–1431.
- Hu-Lieskovan, S., Zhang, J., Wu, L., Shimada, H., Schofield, D.E., and Triche, T.J. (2005). EWS-FLI1 fusion protein up-regulates critical genes in neural crest development and is responsible for the observed phenotype of Ewing's family of tumors. Cancer Res. *65*, 4633–4644.
- Huttlin, E.L., Bruckner, R.J., Paulo, J.A., Cannon, J.R., Ting, L., Baltier, K., Colby, G., Gebreab, F., Gygi, M.P., Parzen, H., et al. (2017). Architecture of the human interactome defines protein communities and disease networks. Nature *545*, 505–509.
- Im, Y.H., Kim, H.T., Lee, C., Poulin, D., Welford, S., Sorensen, P.H., Denny, C.T., and Kim, S.J. (2000). EWS-FLI1, EWS-ERG, and EWS-ETV1 oncoproteins of Ewing tumor family all suppress transcription of transforming growth factor beta type II receptor gene. Cancer Res. *60*, 1536–1540.
- International Human Genome Sequencing Consortium (2004). Finishing the euchromatic sequence of the human genome. Nature *431*, 931–945.
- Jangi, M., Boutz, P.L., Paul, P., and Sharp, P.A. (2014). Rbfox2 controls autoregulation in RNA-binding protein networks. Genes Dev. *28*, 637–651.
- 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.
- Jeon, I.S., Davis, J.N., Braun, B.S., Sublett, J.E., Roussel, M.F., Denny, C.T., and Shapiro, D.N. (1995). A variant Ewing's sarcoma translocation (7;22) fuses the EWS gene to the ETS gene ETV1. Oncogene *10*, 1229–1234.
- Jin, Y., Suzuki, H., Maegawa, S., Endo, H., Sugano, S., Hashimoto, K., Yasuda, K., and Inoue, K. (2003). A vertebrate RNA-binding protein Fox-1 regulates tissue-specific splicing via the pentanucleotide GCAUG. EMBO J. *22*, 905–912.
- Johnson, K.M., Mahler, N.R., Saund, R.S., Theisen, E.R., Taslim, C., Callender, N.W., Crow, J.C., Miller, K.R., and Lessnick, S.L. (2017). Role for the EWS domain of EWS/FLI in binding GGAA-microsatellites required for Ewing sarcoma anchorage independent growth. Proc. Natl. Acad. Sci. U. S. A. *114*, 9870–9875.
- Jolly, M.K., Boareto, M., Huang, B., Jia, D., Lu, M., Ben-Jacob, E., Onuchic, J.N., and Levine, H. (2015). Implications of the Hybrid Epithelial/Mesenchymal Phenotype in Metastasis. Front. Oncol. *5*, 155.
- Juergens, C., Weston, C., Lewis, I., Whelan, J., Paulussen, M., Oberlin, O., Michon, J., Zoubek, A., Juergens, H., and Craft, A. (2006). Safety assessment of intensive induction with vincristine, ifosfamide, doxorubicin, and etoposide (VIDE) in the treatment of Ewing tumors in the EURO-E.W.I.N.G. 99 clinical trial. Pediatr. Blood Cancer *47*, 22–29.
- Jurica, M.S., and Moore, M.J. (2003). Pre-mRNA splicing: awash in a sea of proteins. Mol. Cell *12*, 5–14.
- Kadener, S., Cramer, P., Nogués, G., Cazalla, D., de la Mata, M., Fededa, J.P., Werbajh, S.E., Srebrow, A., and Kornblihtt, A.R. (2001). Antagonistic effects of T-Ag and VP16 reveal a role for RNA pol II elongation on alternative splicing. EMBO J. *20*, 5759–5768.
- Kahles, A., Lehmann, K.-V., Toussaint, N.C., Hüser, M., Stark, S.G., Sachsenberg, T., Stegle, O., Kohlbacher, O., Sander, C., Cancer Genome Atlas Research Network, et al. (2018).

Comprehensive Analysis of Alternative Splicing Across Tumors from 8,705 Patients. Cancer Cell *34*, 211-224.e6.

- Kalluri, R., and Weinberg, R.A. (2009). The basics of epithelial-mesenchymal transition. J. Clin. Invest. *119*, 1420–1428.

- Kalsotra, A., and Cooper, T.A. (2011). Functional consequences of developmentally regulated alternative splicing. Nat. Rev. Genet. *12*, 715–729.

- Kaneko, Y., Yoshida, K., Handa, M., Toyoda, Y., Nishihira, H., Tanaka, Y., Sasaki, Y., Ishida, S., Higashino, F., and Fujinaga, K. (1996). Fusion of an ETS-family gene, EIAF, to EWS by t(17;22)(q12;q12) chromosome translocation in an undifferentiated sarcoma of infancy. Genes. Chromosomes Cancer *15*, 115–121.

- Karim, F.D., Urness, L.D., Thummel, C.S., Klemsz, M.J., McKercher, S.R., Celada, A., Van Beveren, C., Maki, R.A., Gunther, C.V., and Nye, J.A. (1990). The ETS-domain: a new DNA-binding motif that recognizes a purine-rich core DNA sequence. Genes Dev. *4*, 1451–1453.

- Karni, R., de Stanchina, E., Lowe, S.W., Sinha, R., Mu, D., and Krainer, A.R. (2007). The gene encoding the splicing factor SF2/ASF is a proto-oncogene. Nat. Struct. Mol. Biol. *14*, 185–193.

- Kato, M., Han, T.W., Xie, S., Shi, K., Du, X., Wu, L.C., Mirzaei, H., Goldsmith, E.J., Longgood, J., Pei, J., et al. (2012). Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. Cell *149*, 753–767.

- 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.

- Katz, Y., Wang, E.T., Silterra, J., Schwartz, S., Wong, B., Thorvaldsdóttir, H., Robinson, J.T., Mesirov, J.P., Airoldi, E.M., and Burge, C.B. (2015). Quantitative visualization of alternative exon expression from RNA-seq data. Bioinformatics *31*, 2400–2402.

- Kiang, K.M.-Y., and Leung, G.K.-K. (2018). A Review on Adducin from Functional to Pathological Mechanisms: Future Direction in Cancer. BioMed Res. Int. *2018*, 1–14.

- Kim, A.-Y., Kwak, J.-H., Je, N.K., Lee, Y., and Jung, Y.-S. (2015). Epithelial-mesenchymal Transition is Associated with Acquired Resistance to 5-Fluorocuracil in HT-29 Colon Cancer Cells. Toxicol. Res. *31*, 151–156.

- Kim, S., Denny, C.T., and Wisdom, R. (2006). Cooperative DNA binding with AP-1 proteins is required for transformation by EWS-Ets fusion proteins. Mol. Cell. Biol. *26*, 2467–2478.

- Knoop, L.L., and Baker, S.J. (2000). The splicing factor U1C represses EWS/FLI-mediated transactivation. J. Biol. Chem. *275*, 24865–24871.

- Knoop, L.L., and Baker, S.J. (2001). EWS/FLI alters 5'-splice site selection. J. Biol. Chem. *276*, 22317–22322.

- Kornblihtt, A.R. (2005). Promoter usage and alternative splicing. Curr. Opin. Cell Biol. *17*, 262–268.

- Kornblihtt, A.R. (2007). Coupling transcription and alternative splicing. Adv. Exp. Med. Biol. *623*, 175–189.

- Kuroyanagi, H. (2009). Fox-1 family of RNA-binding proteins. Cell. Mol. Life Sci. CMLS *66*, 3895–3907.

- Kwak, E.L., Bang, Y.-J., Camidge, D.R., Shaw, A.T., Solomon, B., Maki, R.G., Ou, S.-H.I., Dezube, B.J., Jänne, P.A., Costa, D.B., et al. (2010). Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. N. Engl. J. Med. *363*, 1693–1703.

- Kwiatkowski, T.J., Bosco, D.A., Leclerc, A.L., Tamrazian, E., Vanderburg, C.R., Russ, C., Davis, A., Gilchrist, J., Kasarskis, E.J., Munsat, T., et al. (2009). Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. Science *323*, 1205–1208.

- Ladenstein, R., Pötschger, U., Le Deley, M.C., Whelan, J., Paulussen, M., Oberlin, O., van den Berg, H., Dirksen, U., Hjorth, L., Michon, J., et al. (2010). Primary disseminated multifocal Ewing sarcoma: results of the Euro-EWING 99 trial. J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol. *28*, 3284–3291.

- Ladomery, M. (2013). Aberrant alternative splicing is another hallmark of cancer. Int. J. Cell Biol. *2013*, 463786.

- Langer, W., Sohler, F., Leder, G., Beckmann, G., Seidel, H., Grone, J., Hummel, M., and Sommer, A. (2010). Exon Array Analysis using re-defined probe sets results in reliable identification of alternatively spliced genes in non-small cell lung cancer. BMC Genomics *11*, 676.

- Lawrence, M.S., Stojanov, P., Polak, P., Kryukov, G.V., Cibulskis, K., Sivachenko, A., Carter, S.L., Stewart, C., Mermel, C.H., Roberts, S.A., et al. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature *499*, 214–218.

- Licatalosi, D.D., Mele, A., Fak, J.J., Ule, J., Kayikci, M., Chi, S.W., Clark, T.A., Schweitzer, A.C., Blume, J.E., Wang, X., et al. (2008). HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature *456*, 464–469.

- Lim, J., Hao, T., Shaw, C., Patel, A.J., Szabó, G., Rual, J.-F., Fisk, C.J., Li, N., Smolyar, A., Hill, D.E., et al. (2006). A protein-protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration. Cell *125*, 801–814.

- Liu, X., Li, H., Rajurkar, M., Li, Q., Cotton, J.L., Ou, J., Zhu, L.J., Goel, H.L., Mercurio, A.M., Park, J.-S., et al. (2016). Tead and AP1 Coordinate Transcription and Motility. Cell Rep. *14*, 1169–1180.

- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

- Luco, R.F., Pan, Q., Tominaga, K., Blencowe, B.J., Pereira-Smith, O.M., and Misteli, T. (2010). Regulation of Alternative Splicing by Histone Modifications. Science *327*, 996–1000.

- Luco, R.F., Allo, M., Schor, I.E., Kornblihtt, A.R., and Misteli, T. (2011). Epigenetics in alternative pre-mRNA splicing. Cell *144*, 16–26.

- Machado, I., López-Guerrero, J.A., Scotlandi, K., Picci, P., and Llombart-Bosch, A. (2018). Immunohistochemical analysis and prognostic significance of PD-L1, PD-1, and CD8+ tumor-infiltrating lymphocytes in Ewing's sarcoma family of tumors (ESFT). Virchows Arch. Int. J. Pathol. *472*, 815–824.

- 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.

- Mackenzie, I.R., Rademakers, R., and Neumann, M. (2010). TDP-43 and FUS in amyotrophic lateral sclerosis and frontotemporal dementia. Lancet Neurol. *9*, 995–1007.

- Maharana, S., Wang, J., Papadopoulos, D.K., Richter, D., Pozniakovsky, A., Poser, I., Bickle, M., Rizk, S., Guillén-Boixet, J., Franzmann, T.M., et al. (2018). RNA buffers the phase separation behavior of prion-like RNA binding proteins. Science *360*, 918–921.

- Malcovati, L., Papaemmanuil, E., Bowen, D.T., Boultwood, J., Della Porta, M.G., Pascutto, C., Travaglino, E., Groves, M.J., Godfrey, A.L., Ambaglio, I., et al. (2011). Clinical significance of

SF3B1 mutations in myelodysplastic syndromes and myelodysplastic/myeloproliferative neoplasms. Blood *118*, 6239–6246.

- Mallinjoud, P., Villemin, J.-P., Mortada, H., Espinoza, M.P., Desmet, F.-O., Samaan, S., Chautard, E., Tranchevent, L.-C., and Auboeuf, D. (2014). Endothelial, epithelial, and fibroblast cells exhibit specific splicing programs independently of their tissue of origin. Genome Res. *24*, 511–521.

- Manara, M.C., Landuzzi, L., Nanni, P., Nicoletti, G., Zambelli, D., Lollini, P.L., Nanni, C., Hofmann, F., García-Echeverría, C., Picci, P., et al. (2007). Preclinical in vivo study of new insulin-like growth factor-I receptor--specific inhibitor in Ewing's sarcoma. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. *13*, 1322–1330.

- Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.-J., Chen, Z., et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. Nature *437*, 376–380.

- Markus, M.A., Heinrich, B., Raitskin, O., Adams, D.J., Mangs, H., Goy, C., Ladomery, M., Sperling, R., Stamm, S., and Morris, B.J. (2006). WT1 interacts with the splicing protein RBM4 and regulates its ability to modulate alternative splicing in vivo. Exp. Cell Res. *312*, 3379–3388.

- de la Mata, M., Alonso, C.R., Kadener, S., Fededa, J.P., Blaustein, M., Pelisch, F., Cramer, P., Bentley, D., and Kornblihtt, A.R. (2003). A Slow RNA Polymerase II Affects Alternative Splicing In Vivo. Mol. Cell *12*, 525–532.

- Mattick, J.S., and Makunin, I.V. (2006). Non-coding RNA. Hum. Mol. Genet. *15 Spec No 1*, R17-29.

- May, W.A., Lessnick, S.L., Braun, B.S., Klemsz, M., Lewis, B.C., Lunsford, L.B., Hromas, R., and Denny, C.T. (1993). The Ewing's sarcoma EWS/FLI-1 fusion gene encodes a more potent transcriptional activator and is a more powerful transforming gene than FLI-1. Mol. Cell. Biol. *13*, 7393–7398.

- Meissner, M., Lopato, S., Gotzmann, J., Sauermann, G., and Barta, A. (2003). Proto-oncoprotein TLS/FUS is associated to the nuclear matrix and complexed with splicing factors PTB, SRm160, and SR proteins. Exp. Cell Res. *283*, 184–195.

- Misteli, T., and Spector, D.L. (1999). RNA polymerase II targets pre-mRNA splicing factors to transcription sites in vivo. Mol. Cell *3*, 697–705.

- Moore, M.J., and Sharp, P.A. (1993). Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. Nature *365*, 364–368.

- Mugneret, F., Lizard, S., Aurias, A., and Turc-Carel, C. (1988). Chromosomes in Ewing's sarcoma. II. Nonrandom additional changes, trisomy 8 and der(16)t(1;16). Cancer Genet. Cytogenet. *32*, 239–245.

- Nakahata, S., and Kawamoto, S. (2005). Tissue-dependent isoforms of mammalian Fox-1 homologs are associated with tissue-specific splicing activities. Nucleic Acids Res. *33*, 2078–2089.

- Ng, T.L., O'Sullivan, M.J., Pallen, C.J., Hayes, M., Clarkson, P.W., Winstanley, M., Sorensen, P.H.B., Nielsen, T.O., and Horsman, D.E. (2007). Ewing Sarcoma with Novel Translocation t(2;16) Producing an In-Frame Fusion of FUS and FEV. J. Mol. Diagn. JMD *9*, 459–463.

- Nieto, M.A., Huang, R.Y.-J., Jackson, R.A., and Thiery, J.P. (2016). EMT: 2016. Cell *166*, 21–45.

- Nilsen, T.W. (2003). The spliceosome: the most complex macromolecular machine in the cell? BioEssays News Rev. Mol. Cell. Dev. Biol. *25*, 1147–1149.

- Nye, J.A., Petersen, J.M., Gunther, C.V., Jonsen, M.D., and Graves, B.J. (1992). Interaction of murine ets-1 with GGA-binding sites establishes the ETS domain as a new DNA-binding motif. Genes Dev. *6*, 975–990.

- Oikawa, T., and Yamada, T. (2003). Molecular biology of the Ets family of transcription factors. Gene *303*, 11–34.

- Oltean, S., and Bates, D.O. (2014). Hallmarks of alternative splicing in cancer. Oncogene *33*, 5311–5318.

- O'Neill, A., Shah, N., Zitomersky, N., Ladanyi, M., Shukla, N., Uren, A., Loeb, D., and Toretsky, J. (2013). Insulin-like growth factor 1 receptor as a therapeutic target in ewing sarcoma: lack of consistent upregulation or recurrent mutation and a review of the clinical trial literature. Sarcoma *2013*, 450478.

- Pal, S., Gupta, R., Kim, H., Wickramasinghe, P., Baubet, V., Showe, L.C., Dahmane, N., and Davuluri, R.V. (2011). Alternative transcription exceeds alternative splicing in generating the transcriptome diversity of cerebellar development. Genome Res. *21*, 1260–1272.

- Papasaikas, P., Tejedor, J.R., Vigevani, L., and Valcárcel, J. (2015). Functional Splicing Network Reveals Extensive Regulatory Potential of the Core Spliceosomal Machinery. Mol. Cell *57*, 7–22.

- Paronetto, M.P., Achsel, T., Massiello, A., Chalfant, C.E., and Sette, C. (2007). The RNA-binding protein Sam68 modulates the alternative splicing of Bcl-x. J. Cell Biol. *176*, 929–939.

- Paronetto, M.P., Miñana, B., and Valcárcel, J. (2011). The Ewing sarcoma protein regulates DNA damage-induced alternative splicing. Mol. Cell *43*, 353–368.

- Pastushenko, I., Brisebarre, A., Sifrim, A., Fioramonti, M., Revenco, T., Boumahdi, S., Van Keymeulen, A., Brown, D., Moers, V., Lemaire, S., et al. (2018). Identification of the tumour transition states occurring during EMT. Nature *556*, 463–468.

- Patel, A.A., and Steitz, J.A. (2003). Splicing double: insights from the second spliceosome. Nat. Rev. Mol. Cell Biol. *4*, 960–970.

- Patel, A., Lee, H.O., Jawerth, L., Maharana, S., Jahnel, M., Hein, M.Y., Stoynov, S., Mahamid, J., Saha, S., Franzmann, T.M., et al. (2015). A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. Cell *162*, 1066–1077.

- Paulussen, M., Ahrens, S., Craft, A.W., Dunst, J., Fröhlich, B., Jabar, S., Rübe, C., Winkelmann, W., Wissing, S., Zoubek, A., et al. (1998). Ewing's tumors with primary lung metastases: survival analysis of 114 (European Intergroup) Cooperative Ewing's Sarcoma Studies patients. J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol. *16*, 3044–3052.

- Pedersen, E.A., Menon, R., Bailey, K.M., Thomas, D.G., Van Noord, R.A., Tran, J., Wang, H., Qu, P.P., Hoering, A., Fearon, E.R., et al. (2016). Activation of Wnt/beta-catenin in Ewing sarcoma cells antagonizes EWS/ETS function and promotes phenotypic transition to more metastatic cell states. Cancer Res. *76*, 5040–5053.

- Pertea, M. (2012). The Human Transcriptome: An Unfinished Story. Genes *3*, 344–360.

- Peter, M., Couturier, J., Pacquement, H., Michon, J., Thomas, G., Magdelenat, H., and Delattre, O. (1997). A new member of the ETS family fused to EWS in Ewing tumors. Oncogene *14*, 1159–1164.

- Pittenger, M.F., Mackay, A.M., Beck, S.C., Jaiswal, R.K., Douglas, R., Mosca, J.D., Moorman, M.A., Simonetti, D.W., Craig, S., and Marshak, D.R. (1999). Multilineage potential of adult human mesenchymal stem cells. Science *284*, 143–147.

- Postel-Vinay, S., Véron, A.S., Tirode, F., Pierron, G., Reynaud, S., Kovar, H., Oberlin, O., Lapouble, E., Ballet, S., Lucchesi, C., et al. (2012). Common variants near TARDBP and EGR2 are associated with susceptibility to Ewing sarcoma. Nat. Genet. *44*, 323–327.

- Pradella, D., Naro, C., Sette, C., and Ghigna, C. (2017). EMT and stemness: flexible processes tuned by alternative splicing in development and cancer progression. Mol. Cancer *16*, 8.
- Prakash, G., Ganesan, P., Ahuja, A., and Bakhshi, S. (2008). MIC-2 positive granulocytic sarcoma of ulna mimicking Ewing sarcoma. Pediatr. Blood Cancer *51*, 836–837.
- Rabbitts, T.H., Forster, A., Larson, R., and Nathan, P. (1993). Fusion of the dominant negative transcription regulator CHOP with a novel gene FUS by translocation t(12;16) in malignant liposarcoma. Nat. Genet. *4*, 175–180.
- Ramaswami, M., Taylor, J.P., and Parker, R. (2013). Altered "Ribostasis": RNA-protein granule formation or persistence in the development of degenerative disorders. Cell *154*.
- Rambaran, R.N., and Serpell, L.C. (2008). Amyloid fibrils: abnormal protein assembly. Prion *2*, 112–117.
- Rambout, X., Detiffe, C., Bruyr, J., Mariavelle, E., Cherkaoui, M., Brohée, S., Demoitié, P., Lebrun, M., Soin, R., Lesage, B., et al. (2016). The transcription factor ERG recruits CCR4–NOT to control mRNA decay and mitotic progression. Nat. Struct. Mol. Biol. *23*, 663–672.
- Rambout, X., Dequiedt, F., and Maquat, L.E. (2018). Beyond Transcription: Roles of Transcription Factors in Pre-mRNA Splicing. Chem. Rev. *118*, 4339–4364.
- Randall, R.L., Lessnick, S.L., Jones, K.B., Gouw, L.G., Cummings, J.E., Cannon-Albright, L., and Schiffman, J.D. (2010). Is There a Predisposition Gene for Ewing's Sarcoma? J. Oncol. *2010*, 397632.
- Ranieri, D., Rosato, B., Nanni, M., Magenta, A., Belleudi, F., and Torrisi, M.R. (2015). Expression of the FGFR2 mesenchymal splicing variant in epithelial cells drives epithelial-mesenchymal transition. Oncotarget *7*, 5440–5460.
- Rappsilber, J., Ryder, U., Lamond, A.I., and Mann, M. (2002). Large-scale proteomic analysis of the human spliceosome. Genome Res. *12*, 1231–1245.
- Ray, D., Kazan, H., Cook, K.B., Weirauch, M.T., Najafabadi, H.S., Li, X., Gueroussov, S., Albu, M., Zheng, H., Yang, A., et al. (2013). A compendium of RNA-binding motifs for decoding gene regulation. Nature *499*, 172–177.
- Reed, R., and Maniatis, T. (1985). Intron sequences involved in lariat formation during pre-mRNA splicing. Cell *41*, 95–105.
- Remy, I., and Michnick, S.W. (2006). A highly sensitive protein-protein interaction assay based on Gaussia luciferase. Nat. Methods *3*, 977–979.
- Revil, T., Toutant, J., Shkreta, L., Garneau, D., Cloutier, P., and Chabot, B. (2007). Protein kinase C-dependent control of Bcl-x alternative splicing. Mol. Cell. Biol. *27*, 8431–8441.
- Revil, T., Pelletier, J., Toutant, J., Cloutier, A., and Chabot, B. (2009). Heterogeneous nuclear ribonucleoprotein K represses the production of pro-apoptotic Bcl-xS splice isoform. J. Biol. Chem. *284*, 21458–21467.
- Reyes, A., and Huber, W. (2018). Alternative start and termination sites of transcription drive most transcript isoform differences across human tissues. Nucleic Acids Res. *46*, 582–592.
- Riggi, N., Cironi, L., Provero, P., Suvà, M.-L., Kaloulis, K., Garcia-Echeverria, C., Hoffmann, F., Trumpp, A., and Stamenkovic, I. (2005). Development of Ewing's sarcoma from primary bone marrow-derived mesenchymal progenitor cells. Cancer Res. *65*, 11459–11468.
- 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.

REFERENCES

- Rinn, J.L., and Chang, H.Y. (2012). Genome regulation by long noncoding RNAs. Annu. Rev. Biochem. *81*, 145–166.

- Roberts, G.C., Gooding, C., Mak, H.Y., Proudfoot, N.J., and Smith, C.W. (1998). Co-transcriptional commitment to alternative splice site selection. Nucleic Acids Res. *26*, 5568–5572.

- Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative Genomics Viewer. Nat. Biotechnol. *29*, 24–26.

- Rossbach, O., Hung, L.-H., Khrameeva, E., Schreiner, S., König, J., Curk, T., Zupan, B., Ule, J., Gelfand, M.S., and Bindereif, A. (2014). Crosslinking-immunoprecipition (iCLIP) analysis reveals global regulatory roles of hnRNP L. RNA Biol. *11*, 146–155.

- Sakharkar, M.K., Chow, V.T.K., and Kangueane, P. (2004). Distributions of exons and introns in the human genome. In Silico Biol. *4*, 387–393.

- Saldi, T., Cortazar, M.A., Sheridan, R.M., and Bentley, D.L. (2016). Coupling of RNA Polymerase II Transcription Elongation with Pre-mRNA Splicing. J. Mol. Biol. *428*, 2623–2635.

- Sanchez, G., Bittencourt, D., Laud, K., Barbier, J., Delattre, O., Auboeuf, D., and Dutertre, M. (2008a). Alteration of cyclin D1 transcript elongation by a mutated transcription factor up-regulates the oncogenic D1b splice isoform in cancer. Proc. Natl. Acad. Sci. U. S. A. *105*, 6004–6009.

- Sanchez, G., Delattre, O., Auboeuf, D., and Dutertre, M. (2008b). Coupled alteration of transcription and splicing by a single oncogene: boosting the effect on cyclin D1 activity. Cell Cycle Georget. Tex *7*, 2299–2305.

- Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U. S. A. *74*, 5463–5467.

- Sankar, S., Theisen, E.R., Bearss, J., Mulvihill, T., Hoffman, L.M., Sorna, V., Beckerle, M.C., Sharma, S., and Lessnick, S.L. (2014). Reversible LSD1 inhibition interferes with global EWS/ETS transcriptional activity and impedes Ewing sarcoma tumor growth. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. *20*, 4584–4597.

- Sannino, G., Marchetto, A., Kirchner, T., and Grünewald, T.G.P. (2017). Epithelial-to-Mesenchymal and Mesenchymal-to-Epithelial Transition in Mesenchymal Tumors: A Paradox in Sarcomas? Cancer Res. *77*, 4556–4561.

- Sanz, L.A., Hartono, S.R., Lim, Y.W., Steyaert, S., Rajpurkar, A., Ginno, P.A., Xu, X., and Chédin, F. (2016). Prevalent, Dynamic, and Conserved R-Loop Structures Associate with Specific Epigenomic Signatures in Mammals. Mol. Cell *63*, 167–178.

- Schober, M., Rebay, I., and Perrimon, N. (2005). Function of the ETS transcription factor Yan in border cell migration. Dev. Camb. Engl. *132*, 3493–3504.

- Schuetz, A.N., Rubin, B.P., Goldblum, J.R., Shehata, B., Weiss, S.W., Liu, W., Wick, M.R., and Folpe, A.L. (2005). Intercellular junctions in Ewing sarcoma/primitive neuroectodermal tumor: additional evidence of epithelial differentiation. Mod. Pathol. *18*, 1403–1410.

- Schwerk, C., and Schulze-Osthoff, K. (2005). Regulation of Apoptosis by Alternative Pre-mRNA Splicing. Mol. Cell *19*, 1–13.

- Scotlandi, K., Benini, S., Nanni, P., Lollini, P.L., Nicoletti, G., Landuzzi, L., Serra, M., Manara, M.C., Picci, P., and Baldini, N. (1998). Blockage of insulin-like growth factor-I receptor inhibits the growth of Ewing's sarcoma in athymic mice. Cancer Res. *58*, 4127–4131.

- Sebestyén, E., Singh, B., Miñana, B., Pagès, A., Mateo, F., Pujana, M.A., Valcárcel, J., and Eyras, E. (2016). Large-scale analysis of genome and transcriptome alterations in multiple tumors unveils novel cancer-relevant splicing networks. Genome Res. *26*, 732–744.

- Selvanathan, S.P., Graham, G.T., Erkizan, H.V., Dirksen, U., Natarajan, T.G., Dakic, A., Yu, S., Liu, X., Paulsen, M.T., Ljungman, M.E., et al. (2015). Oncogenic fusion protein EWS-FLI1 is a network hub that regulates alternative splicing. Proc. Natl. Acad. Sci. U. S. A. *112*, E1307-1316.

- Serrano-Gomez, S.J., Maziveyi, M., and Alahari, S.K. (2016). Regulation of epithelial-mesenchymal transition through epigenetic and post-translational modifications. Mol. Cancer *15*.

- Shapiro, I.M., Cheng, A.W., Flytzanis, N.C., Balsamo, M., Condeelis, J.S., Oktay, M.H., Burge, C.B., and Gertler, F.B. (2011). An EMT-driven alternative splicing program occurs in human breast cancer and modulates cellular phenotype. PLoS Genet. *7*, e1002218.

- Sharrocks, A.D. (2001). The ETS-domain transcription factor family. Nat. Rev. Mol. Cell Biol. *2*, 827–837.

- 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.

- Shen, S., Park, J.W., Lu, Z., Lin, L., Henry, M.D., Wu, Y.N., Zhou, Q., and Xing, Y. (2014). rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. Proc. Natl. Acad. Sci. U. S. A. *111*, E5593–E5601.

- Shing, D.C., McMullan, D.J., Roberts, P., Smith, K., Chin, S.-F., Nicholson, J., Tillman, R.M., Ramani, P., Cullinane, C., and Coleman, N. (2003). FUS/ERG gene fusions in Ewing's tumors. Cancer Res. *63*, 4568–4576.

- Siddique, H.R., Rao, V.N., Lee, L., and Reddy, E.S. (1993). Characterization of the DNA binding and transcriptional activation domains of the erg protein. Oncogene *8*, 1751–1755.

- Sizemore, G.M., Pitarresi, J.R., Balakrishnan, S., and Ostrowski, M.C. (2017). The ETS family of oncogenic transcription factors in solid tumours. Nat. Rev. Cancer *17*, 337–351.

- Snel, B., Lehmann, G., Bork, P., and Huynen, M.A. (2000). STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. Nucleic Acids Res. *28*, 3442–3444.

- Spraker, H.L., Price, S.L., Chaturvedi, A., Schiffman, J.D., Jones, K.B., Lessnick, S.L., Beckerle, M., and Randall, R.L. (2012). The clone wars - revenge of the metastatic rogue state: the sarcoma paradigm. Front. Oncol. *2*, 2.

- Spurny, C., Kailayangiri, S., Jamitzky, S., Altvater, B., Wardelmann, E., Dirksen, U., Hardes, J., Hartmann, W., and Rossig, C. (2018). Programmed cell death ligand 1 (PD-L1) expression is not a predominant feature in Ewing sarcomas. Pediatr. Blood Cancer *65*.

- Stark, C., Breitkreutz, B.-J., Reguly, T., Boucher, L., Breitkreutz, A., and Tyers, M. (2006). BioGRID: a general repository for interaction datasets. Nucleic Acids Res. *34*, D535-539.

- 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.

- Sun, S., Zhang, Z., Fregoso, O., and Krainer, A.R. (2012). Mechanisms of activation and repression by the alternative splicing factors RBFOX1/2. RNA *18*, 274–283.

- Svetoni, F., Frisone, P., and Paronetto, M.P. (2016). Role of FET proteins in neurodegenerative disorders. RNA Biol. *13*, 1089–1102.

- Tan, A.Y., and Manley, J.L. (2009). The TET family of proteins: functions and roles in disease. J. Mol. Cell Biol. *1*, 82–92.

- Teitell, M.A., Thompson, A.D., Sorensen, P.H., Shimada, H., Triche, T.J., and Denny, C.T. (1999). EWS/ETS fusion genes induce epithelial and neuroectodermal differentiation in NIH 3T3 fibroblasts. Lab. Investig. J. Tech. Methods Pathol. *79*, 1535–1543.

- Tennyson, C.N., Klamut, H.J., and Worton, R.G. (1995). The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced. Nat. Genet. *9*, 184–190.
- Thakore, P.I., D'Ippolito, A.M., Song, L., Safi, A., Shivakumar, N.K., Kabadi, A.M., Reddy, T.E., Crawford, G.E., and Gersbach, C.A. (2015). Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. Nat. Methods *12*, 1143–1149.
- Thiery, J.P. (2002). Epithelial-mesenchymal transitions in tumour progression. Nat. Rev. Cancer *2*, 442–454.
- Thiery, J.P., Acloque, H., Huang, R.Y.J., and Nieto, M.A. (2009). Epithelial-mesenchymal transitions in development and disease. Cell *139*, 871–890.
- 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.
- Tolstorukov, M.Y., Sansam, C.G., Lu, P., Koellhoffer, E.C., Helming, K.C., Alver, B.H., Tillman, E.J., Evans, J.A., Wilson, B.G., Park, P.J., et al. (2013). Swi/Snf chromatin remodeling/tumor suppressor complex establishes nucleosome occupancy at target promoters. Proc. Natl. Acad. Sci. U. S. A. *110*, 10165–10170.
- 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.
- Tomlins, S.A., Rhodes, D.R., Perner, S., Dhanasekaran, S.M., Mehra, R., Sun, X.-W., Varambally, S., Cao, X., Tchinda, J., Kuefer, R., et al. (2005). Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science *310*, 644–648.
- Torchia, E.C., Jaishankar, S., and Baker, S.J. (2003). Ewing tumor fusion proteins block the differentiation of pluripotent marrow stromal cells. Cancer Res. *63*, 3464–3468.
- Toretsky, J.A., Kalebic, T., Blakesley, V., LeRoith, D., and Helman, L.J. (1997). The insulin-like growth factor-I receptor is required for EWS/FLI-1 transformation of fibroblasts. J. Biol. Chem. *272*, 30822–30827.
- Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinforma. Oxf. Engl. *25*, 1105–1111.
- Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. *28*, 511–515.
- Trisciuoglio, D., Tupone, M.G., Desideri, M., Di Martile, M., Gabellini, C., Buglioni, S., Pallocca, M., Alessandrini, G., D'Aguanno, S., and Del Bufalo, D. (2017). BCL-XL overexpression promotes tumor progression-associated properties. Cell Death Dis. *8*, 3216.
- Tsai, J.H., and Yang, J. (2013). Epithelial-mesenchymal plasticity in carcinoma metastasis. Genes Dev. *27*, 2192–2206.
- Turc-Carel, C., Philip, I., Berger, M.P., Philip, T., and Lenoir, G. (1983). [Chromosomal translocation (11; 22) in cell lines of Ewing's sarcoma]. Comptes Rendus Seances Acad. Sci. Ser. III Sci. Vie *296*, 1101–1103.

- Ule, J., Ule, A., Spencer, J., Williams, A., Hu, J.-S., Cline, M., Wang, H., Clark, T., Fraser, C., Ruggiu, M., et al. (2005). Nova regulates brain-specific splicing to shape the synapse. Nat. Genet. *37*, 844–852.

- Vallenius, T. (2013). Actin stress fibre subtypes in mesenchymal-migrating cells. Open Biol. *3*.

- Van Nostrand, E.L., Pratt, G.A., Shishkin, A.A., Gelboin-Burkhart, C., Fang, M.Y., Sundararaman, B., Blue, S.M., Nguyen, T.B., Surka, C., Elkins, K., et al. (2016). Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). Nat. Methods *13*, 508–514.

- Venables, J.P., Lapasset, L., Gadea, G., Fort, P., Klinck, R., Irimia, M., Vignal, E., Thibault, P., Prinos, P., Chabot, B., et al. (2013a). MBNL1 and RBFOX2 cooperate to establish a splicing programme involved in pluripotent stem cell differentiation. Nat. Commun. *4*, 2480.

- Venables, J.P., Brosseau, J.-P., Gadea, G., Klinck, R., Prinos, P., Beaulieu, J.-F., Lapointe, E., Durand, M., Thibault, P., Tremblay, K., et al. (2013b). RBFOX2 is an important regulator of mesenchymal tissue-specific splicing in both normal and cancer tissues. Mol. Cell. Biol. *33*, 396–405.

- Wahl, J., Bogatyreva, L., Boukamp, P., Rojewski, M., van Valen, F., Fiedler, J., Hipp, N., Debatin, K.-M., and Beltinger, C. (2010). Ewing's sarcoma cells with CD57-associated increase of tumorigenicity and with neural crest-like differentiation capacity. Int. J. Cancer *127*, 1295–1307.

- Wahl, M.C., Will, C.L., and Lührmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. Cell *136*, 701–718.

- Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. Nature *456*, 470–476.

- Wang, W.-Y., Pan, L., Su, S.C., Quinn, E.J., Sasaki, M., Jimenez, J.C., Mackenzie, I.R.A., Huang, E.J., and Tsai, L.-H. (2013). Interaction of FUS and HDAC1 regulates DNA damage response and repair in neurons. Nat. Neurosci. *16*, 1383–1391.

- Wang, Z., Rolish, M.E., Yeo, G., Tung, V., Mawson, M., and Burge, C.B. (2004). Systematic identification and analysis of exonic splicing silencers. Cell *119*, 831–845.

- Wang, Z., Li, Y., Kong, D., Banerjee, S., Ahmad, A., Azmi, A.S., Ali, S., Abbruzzese, J.L., Gallick, G.E., and Sarkar, F.H. (2009). Acquisition of epithelial-mesenchymal transition phenotype of gemcitabine-resistant pancreatic cancer cells is linked with activation of the notch signaling pathway. Cancer Res. *69*, 2400–2407.

- Warzecha, C.C., and Carstens, R.P. (2012). Complex changes in alternative pre-mRNA splicing play a central role in the epithelial-to-mesenchymal transition (EMT). Semin. Cancer Biol. *22*, 417–427.

- Warzecha, C.C., Jiang, P., Amirikian, K., Dittmar, K.A., Lu, H., Shen, S., Guo, W., Xing, Y., and Carstens, R.P. (2010). An ESRP-regulated splicing programme is abrogated during the epithelial-mesenchymal transition. EMBO J. *29*, 3286–3300.

- Wei, C., Xiao, R., Chen, L., Cui, H., Zhou, Y., Xue, Y., Hu, J., Zhou, B., Tsutsui, T., Qiu, J., et al. (2016). RBFox2 Binds Nascent RNA to Globally Regulate Polycomb Complex 2 Targeting in Mammalian Genomes. Mol. Cell *62*, 875–889.

- Wiles, E.T., Bell, R., Thomas, D., Beckerle, M., and Lessnick, S.L. (2013). ZEB2 Represses the Epithelial Phenotype and Facilitates Metastasis in Ewing Sarcoma. Genes Cancer *4*, 486–500.

- Will, C.L., and Lührmann, R. (2011). Spliceosome Structure and Function. Cold Spring Harb. Perspect. Biol. *3*.

- Wu, S., and Green, M.R. (1997). Identification of a human protein that recognizes the 3' splice site during the second step of pre-mRNA splicing. EMBO J. *16*, 4421–4432.
- Wu, J.Y., Tang, H., and Havlioglu, N. (2003). Alternative pre-mRNA splicing and regulation of programmed cell death. Prog. Mol. Subcell. Biol. *31*, 153–185.
- Yang, J., Mani, S.A., Donaher, J.L., Ramaswamy, S., Itzykson, R.A., Come, C., Savagner, P., Gitelman, I., Richardson, A., and Weinberg, R.A. (2004). Twist, a Master Regulator of Morphogenesis, Plays an Essential Role in Tumor Metastasis. Cell *117*, 927–939.
- Yang, L., Embree, L.J., Tsai, S., and Hickstein, D.D. (1998). Oncoprotein TLS interacts with serine-arginine proteins involved in RNA splicing. J. Biol. Chem. *273*, 27761–27764.
- Yang, L., Chansky, H.A., and Hickstein, D.D. (2000). EWS·Fli-1 Fusion Protein Interacts with Hyperphosphorylated RNA Polymerase II and Interferes with Serine-Arginine Protein-mediated RNA Splicing. J. Biol. Chem. *275*, 37612–37618.
- Yang, Y., Park, J.W., Bebee, T.W., Warzecha, C.C., Guo, Y., Shang, X., Xing, Y., and Carstens, R.P. (2016). Determination of a Comprehensive Alternative Splicing Regulatory Network and Combinatorial Regulation by Key Factors during the Epithelial-to-Mesenchymal Transition. Mol. Cell. Biol. *36*, 1704–1719.
- Yee, D., Favoni, R.E., Lebovic, G.S., Lombana, F., Powell, D.R., Reynolds, C.P., and Rosen, N. (1990). Insulin-like growth factor I expression by tumors of neuroectodermal origin with the t(11;22) chromosomal translocation. A potential autocrine growth factor. J. Clin. Invest. *86*, 1806–1814.
- Yeo, G.W., Coufal, N.G., Liang, T.Y., Peng, G.E., Fu, X.-D., and Gage, F.H. (2009). An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. Nat. Struct. Mol. Biol. *16*, 130–137.
- Ying, Y., Wang, X.-J., Vuong, C.K., Lin, C.-H., Damianov, A., and Black, D.L. (2017). Splicing Activation by Rbfox Requires Self-Aggregation through Its Tyrosine-Rich Domain. Cell *170*, 312-323.e10.
- Yoshida, K., Sanada, M., Shiraishi, Y., Nowak, D., Nagata, Y., Yamamoto, R., Sato, Y., Sato-Otsubo, A., Kon, A., Nagasaki, M., et al. (2011). Frequent pathway mutations of splicing machinery in myelodysplasia. Nature *478*, 64–69.
- Zanconato, F., Forcato, M., Battilana, G., Azzolin, L., Quaranta, E., Bodega, B., Rosato, A., Bicciato, S., Cordenonsi, M., and Piccolo, S. (2015). Genome-wide association between YAP/TAZ/TEAD and AP-1 at enhancers drives oncogenic growth. Nat. Cell Biol. *17*, 1218–1227.
- Zeisberg, M., and Neilson, E.G. (2009). Biomarkers for epithelial-mesenchymal transitions. J. Clin. Invest. *119*, 1429–1437.
- Zeng, C., and Berget, S.M. (2000). Participation of the C-terminal domain of RNA polymerase II in exon definition during pre-mRNA splicing. Mol. Cell. Biol. *20*, 8290–8301.
- Zhang, C., Zhang, Z., Castle, J., Sun, S., Johnson, J., Krainer, A.R., and Zhang, M.Q. (2008). Defining the regulatory network of the tissue-specific splicing factors Fox-1 and Fox-2. Genes Dev. *22*, 2550–2563.
- Zhang, D., Paley, A.J., and Childs, G. (1998). The transcriptional repressor ZFM1 interacts with and modulates the ability of EWS to activate transcription. J. Biol. Chem. *273*, 18086–18091.
- Zhou, A., Ou, A.C., Cho, A., Benz, E.J., and Huang, S.-C. (2008). Novel splicing factor RBM25 modulates Bcl-x pre-mRNA 5' splice site selection. Mol. Cell. Biol. *28*, 5924–5936.
- 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.