UNIVERSITAT DE BARCELONA

Unveiling the role of the Mediator complex in Ewing sarcoma tumorigenesis

Irene Cuervas Oliveras

Universitat de Barcelona

Doctoral Programme in Biomedicine – Facultat de Medicina

Unveiling the role of the Mediator complex in Ewing sarcoma tumorigenesis

Irene Cuervas Oliveras

Developmental Tumor Biology Laboratory

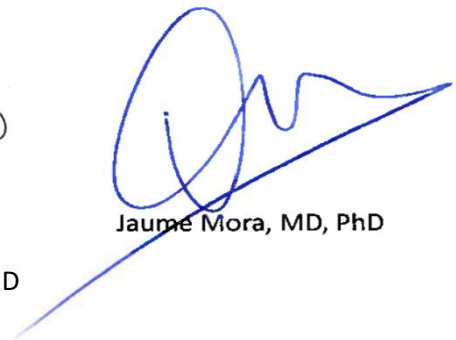
Institut de Recerca Sant Joan de Déu (IRSJD) – Pediatric Cancer Center Barcelona (PCCB)

PhD candidate

Irene Cuervas Oliveras

Thesis directors

María Inmaculada Hernández Muñoz, PhD



Thesis tutor

Neus Agell Jané, PhD

Que la vida que ens hem perdut simplement no existeix.

La bola de cristall - Manel

A en Martí.

ACKNOWLEDGMENTS

Si penso en com he arribat fins aquí, el primer que recordo és una classe de Biologia a 3r d'ESO on em van explicar com van clonar a l'ovella Dolly. Vaig pensar "uau! Jo vull poder fer alguna cosa així". Recordo també que la professora de Biologia de 1r de batxillerat em va dir que se'm donava bé això de la ciència, que continués per aquí. Però en aquell moment mai vaig pensar que acabaria fent una tesis.

Por eso quiero empezar agradeciendo a Inma, mi mentora en todo este trayecto. Muchas gracias por guiarme siempre y confiar en mi desde el primer día. Por empujarme a seguir adelante y a tener un pensamiento crítico. Y también por darme siempre toda la libertad para hacer las cosas a mi manera. També moltes gràcies a en Jaume, per la confiança, el suport i la dedicació al meu treball. I per transmetrem la importància i la passió per la nostra feina.

Gracias a Ramón, por ser mi primer jefe en este curioso mundo de la investigación. Por ver algo especial en mí y ayudarme a conseguir esta oportunidad.

A la primera persona que vaig conèixer quan vaig entrar al PRBB, l'Evelyn. Gràcies per tot el que m'has ensenyat i per tot el que hem passat. Si el primer any no hagués estat tan bé gràcies a tu, segurament la decisió de quedar-me aquí hagués sigut una mica diferent.

A tot el grup del laboratori d'Oncologia de Sant Joan de Déu. Gràcies per acollir-me sempre tan bé tot i venir tan poc. I sobretot gràcies a la Carla, la Maria i el Pablo, per fer que els divendres fossin més fàcils i fer-me la vida molt millor.

Als meus amics de la Uni. Perquè els millors anys a la UdG ens van fer els científics que som ara, i m'encanta saber que després de tots aquests anys encara seguim sent els mateixos.

Gràcies a tothom que ha fet que Barcelona sigui casa meva. A Sara, por ser ya mi familia. A la Nura, per ser el meu +1 en tot cada dia dins i fora del PRBB. A Jesús, mi más mejor compañero siempre. A Pablo, por escucharnos y hacernos la vida mucho mejor. A Anna, porque algún día llegaremos a correr de verdad. Y a la familia Revoltosa, por ser siempre el sitio al que ir cuando estoy perdidísima.

A les meves amigues de Girona, perquè són i sempre seran les millors. Per ser el millor suport i confiar sempre en mi. I perquè des de Barcelona, Girona, Lisboa, Amsterdam o Madrid sempre trobarem un moment per nosaltres. Us estimo molt! (Especial menció a la Marta Tapia per fer-me la portada més preciosa jeje).

A la meva família, sobretot al Papa i la Mama. Per haver-me donat tot el que tinc i haver-me deixat arribar fins aquí sense qüestionar-me mai res. Per estimar-me i cuidar-me sempre. I a l'avi, perquè com sempre deia: "Vaig arribar a Catalunya només amb una maleta, i mira tot el que tinc ara!". Quina pena que no pugui veure tot això.

I al més important de tots, en Martí. Per ser el millor regal que m'ha donat la vida. Per haver-me acompanyat sempre i perquè com tu sempre dius: som més que amics, som GERMANS. Perquè tot i jo ser la gran, el meu referent ets tu. Per ser molt més intel·ligent, sensat i fort que jo. Per saber que sempre hi haurà un camí encara que no sigui l'establert. T'estimo molt!.

ABSTRACT

Ewing sarcoma is an aggressive tumor arising in bones and soft tissues that affects children and young adults. This disease is determined by an aberrant chromosomal translocation that forms an oncogenic fusion protein, being EWSR1-FLI1 the most common. The oncogene is responsible for tumorigenesis through epigenetic deregulation in a particular cell context. It generates *de novo* super-enhancers that regulate crucial genes for cellular maintenance and tumorigenesis, and which are largely enriched in cofactors associated to RNA-pol II such as the Mediator complex.

In order to characterize the specific functions of super-enhancer-regulated genes that enable tumorigenesis, we have identified *in silico* the Mediator complex subunit MED13L and RERE as genes regulated by EWSR1-FLI1-bound super-enhancers. We found that EWSR1-FLI1 regulates the expression of MED13L and RERE, which are differentially expressed in Ewing sarcoma. Both proteins bind to overlapping regions with EWSR1-FLI1 on CpG islands corresponding to genes involved in protein translation and alternative splicing. In addition, MED13L and RERE regulate the expression of spliceosome components by altering the splicing pattern of important genes for Ewing sarcoma tumorigenesis. We observed that Ewing sarcoma cell lines are highly sensitive to the splicing inhibitor Pladienolide B, and that cells respond differently depending on MED13L and RERE levels. Therefore, we propose that the synergistic effect due to the combination of Pladienolide B and the CDK8 Mediator kinase inhibitor BRD6989 might be further studied for consideration as a new therapeutic approach in Ewing sarcoma.

INDEX

LIST OF FIGURES

LIST OF TABLES

ABBREIATIONS

ALL: Acute lymphocytic leukemia

AML: Acute myeloid leukemia

APS: Ammonium persulfate

AS: Alternative Splicing

ATA: Amino terminal transcriptional activation

BPS: Branch point sequence

BSA: Bovine serum albumin

ChIP: Chromatin immunoprecipitation

CTD: C-terminal domain

DAPI: 4',6-diamidino-2-phenylindole (DAPI)

DBD: DNA binding domain

DEG: Differential expressed gene

DNMT: DNA methyltransferase

DOC: Deoxycholic acid sodium salt

DSRCT: Desmoplastic small round-cell tumor

EDTA: Ethylenediaminetetraacetic acid

EGTA: Ethylene glycol tetraacetic acid

ES: Ewing sarcoma

ESC: Embryonic stem cells

esiRNA: Endoribonuclease-prepared siRNA

FDG: Fluorodeoxyglucose

FRET: Förster resonance energy transfer

GO: Gene ontology

GSEA: Gene Set Enrichment Analysis

GTF: General transcription factor

HDAC: Histone deacetylases (HDAC)

hMSC: Human mesenchymal stem cells

ICP: Infancy-childhood-puberty

IF: Immunofluorescence

IHC: Immunohistochemistry

IP: Immunoprecipitation

LB: Luria Bertani

LDH: Lactate dehydrogenase

lncRNA: Long non-coding RNAs

miRNA: MicroRNAs

MSC: Mesenchymal stem cells

ncRNA: Non-coding RNA

NGS: Next-generation sequencing

PBS: Phosphate Buffered Saline

PCR: Polymerase Chain Reaction

PET: Positron emission tomography

PIC: Pre-initiation complex

PPT: Polypyrimidine tract

PSI: Percent spliced-in

PTM: Post-translational modification

qPCR: Quantitative PCR

RISC: RNA-induced silencing complex

RMS: Rhabdomyosarcoma

RNA pol: RNA polymerase

rRNA: Ribosomal RNA

RT: Reverse transcription

SDS: Sodium dodecyl sulfate

SDS-PAGE: Sodium dodecyl sulfate polyacrylamide gel

SE: Super-enhancer

shRNA: Short hairpin RNA

siRNA: Small interfering RNA

snRNA: Small nuclear RNA

snRNP: Small nuclear ribonucleoprotein

SS: Splice site

TBE: Tris-Acetate-EDTA

TBS-T: Tris-buffered saline-Tween

tRNA: Transfer RNA

TSS: Transcription start site

UCSC: University of California Santa Cruz

UTR: Untranslated region

WB: Western Blot

WHO: World Health Organization

INTRODUCTION

1. Pediatric cancer: a disease different from adult cancer

1.1. Definition

Cancer is a large group of diseases that can affect almost any organ or tissue consisting in the uncontrollable expansion of abnormal cells [1, 2]. The features of the transformed cells and the environment in which they arise will define the biology of the tumor [2]. Pediatric cancer is the diverse group of solid and hematological tumors that is developed in children and adolescents from 0 to 19 years old [3]. Although it is considered a rare disease, 400000 cases emerge per year and is the main cause of death among children and adolescents in high-income countries [3-6]. Survival rates are mainly dependent on the socio-economic health care status of each country. In high-income countries, more than 80% of children with cancer survive. However, in low- and middle-income countries, survival rates remain below 30% [3].

1.2. Origin of pediatric cancer

Pediatrics is the branch of medicine that takes care of infants, children and adolescents. During this period the human body grows and develops transforming pluripotent undifferentiated cells to thirteen billion of well-differentiated cells that ensure the organization and function of the adult human body [2]. The post-natal process was described by Kalberg by a mathematical model of infancy-childhood-puberty (ICP) [7]. The three

components of the ICP model include the different hormonal phases that take place in the growth process following a sigmoid curve (**Figure 1**). The infancy period comprises the second half of pregnancy to 2-3 years after birth and is characterized by being non-dependent on hormones. It comprises a peak velocity of growth that deaccelerates during the first two years of life. Childhood begins after the first 2-3 years, when growth hormone starts to take effect. During this period, body growth remains slow. Finally, puberty is the period where growth hormone and sexual steroids cooperate until initiation of adulthood, when height and full reproductive maturity is achieved (at 21-25 years of age) [2, 7].

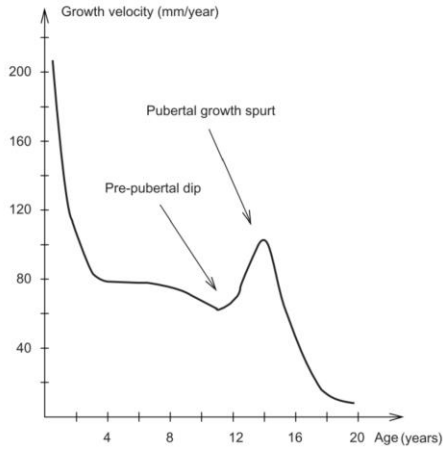


Figure 1. **Velocity curve of human growth.** Human growth according to the Karlberg ICP mathematical model [7] defined by the different hormonal phases. Adapted from Mora (2012) [2].

According to the ICP model, pediatric tumors can be classified as infant, childhood or puberty tumors as it is represented in **Table 1** [2, 7]. Tumors that develop during infancy arise from cellular populations that have not completed differentiation. Since each tissue develops chronologically differently, childhood tumors develop in cellular populations of tissues or organs that have not completed differentiation in this stage. For example, childhood leukemias occur while the immune system continues its

development during childhood [2, 8]. Finally, puberty tumors are generated from cellular populations that are undifferentiated and arise when the pubertal growth spurt occurs due to growth factors and sexual steroids are secreted [2, 7].

Infancy tumors	Childhood tumors	Puberty tumors
Hemangioma	Childhood leukemia (ALL and AML)	Hodgkin's lymphoma
Infant neuroblastoma	Childhood astrocytoma	Gonadal germ cell tumors
Congenital germ cell tumors	Nephroblastoma (Wilms tumor)	Ewing sarcoma
Infant leukemia (ALL and AML)	Childhood neuroblastoma	Osteosarcoma
Atypical teratoid/rhabdoid tumor	Embryonal rhabdomyosarcoma	Desmoplastic small round-cell tumor (DSRCT)
Central PNET	Langerhans cell histiocytosis	Synovial sarcoma
Ependymoblastoma	Non-Hodgkin's lymphoma	Alveolar rhabdomyosarcoma
Hepatoblastoma	Fibromatosis	Neuroblastoma of adolescents and young adults
Retinoblastoma	Group A posterior fossa ependymoma	Medulloblastoma (SHH type of young adults)
Mesoblastic nephroma		Group B posterior fossa ependymoma
Congenital fibrosarcoma		
Inflammatory myofibroblastic tumor		
Pulmonary pleuroblastoma		

Table 1. **Classification of pediatric tumors.** ICP model of the human growth and development allows classification of pediatric tumors into infancy, childhood and puberty tumors. ALL, acute lymphoblastic leukemia; AML, acute myeloblastic leukemia; DSRCT, desmoplastic small round-cell tumor; PNET, primitive neuroectodermal tumor; SHH, sonic hedgehog homolog. Adapted from Mora (2012) [2].

Tumorigenesis in pediatric cancers is associated to organogenesis, growth and maturation considering that each tumor arises chronologically when the corresponding tissue arrives at maximum development [9]. Besides, the genotypic and phenotypic characteristics of the pediatric cancer cells resemble the biology and environment of the undifferentiated cells from the developing tissues in which they arise [2, 6]. In agreement with this, advances in next-generation sequencing have allowed to compare pediatric tumors with developing normal tissues to prove the transcriptional relationship of cancer cells to undifferentiated cells of each organ. For example, medulloblastoma with cerebellum precursors, Wilms tumors with metanephric development of the kidney; and neuroblastoma with sympatho-adrenal precursors of the neural crest [10-12].

These data, together with genetic and epigenetic analyses of transcription, enhancer states and DNA methylation have shown that pediatric cancer cells exhibit defects in normal differentiation mechanisms that lead to blocked or impinged maturation [2, 6, 9, 13]. Epigenetic deregulation has a key role, as has been reported to be in a locked-in state [14]. Altogether, these data reveal a unique biology for pediatric tumors as the cell of origin of each tumor allows tumorigenesis in a particular development state. Genetic alterations occurring in cells with a high proliferation rate during development are the first genetic hits that will result in an accumulation of undifferentiated cells blocked in their maturation stage [2, 14].

1.3. Unique characteristics of pediatric cancers and differences from adult cancers

Pediatric cancers have unique characteristics that distinguish them from adult cancers. 85% of adult cancers arise in the epithelia of different organs giving rise to carcinomas. In the case of pediatric cancers, they predominantly derive from mesodermal and ectodermal germ layers and 50% are found in the hematopoietic and central nervous system, while only 9% have an epithelial origin [2, 6, 9]. Other important features that separate pediatric cancers from adult cancers are the type of genetic alterations and the genes affected by those alterations. Adult cancers usually arise due to linear accumulation of mutations because of aging or an extended exposure to mutagens on account of environmental and lifestyle factors. Indeed, progenitor cells in adult tissues have accumulated many mutations during multiple cell cycles over the years. The genetic alterations frequently found in adult tumors are mutations including base substitutions, insertions and deletions (indels), genome rearrangements and changes in chromosome copy-number. In contrast, chromosomal fusions are often the only genetic alterations found in pediatric cancer, usually involving genes related to embryonic development are frequent in pediatric cancer, whereas the presence of translocations is much more restricted in adult tumors.

The cellular processes frequently involved in adult cancers include a broad spectrum of signaling pathways and cell cycle regulation, such as MAPK, PI3K or WNT/β-catenin signaling [15, 16]. After cell division, adult cells tend to rapidly differentiate or undergo apoptosis. Thus, multi-hit tumorigenesis is required to overcome cellular controls, such as that imposed by TP53, which prevents the accumulation of DNA damage [2, 6, 9, 15, 16]. TP53 is the

tumor suppressor gene which is most often mutated, especially in breast, lung, ovarian and colon cancers. The kinase PIK3CA is the second most frequent mutation found above all in breast and endometrial carcinomas. The protooncogene KRAS is usually mutated in pancreas, lung and colorectal cancer. EGFR aberrant alterations are also found in lung adenocarcinoma and in adult glioblastoma. Other commonly mutated tumor suppressor genes are CSMD3 and LRP1B. Protein kinases BRAF and ATM, phosphatases PTPRT and PTEN, and the receptor MUC16 are also among the most mutated genes [15, 19].

In contrast, many of the genetic alterations in pediatric cancer affect the epigenetic machinery and lead to an epigenetic deregulation that results in tumorigenesis when occurs in the appropriate cell environment [14]. For example, diffuse intrinsic pontine glioma (DIPG), which originates in the brainstem, is genetically defined by substitution of methionine for lysine at site 27 of histone 3 (H3K27M) [20]. Otherwise, histone modifying enzymes like KDM6A and chromatin remodeling complexes as the BRG1/BRM associated factor (BAF) complex and Polycomb repressive complex 2 (PRC2) are also affected in some pediatric cancers like leukemias [14, 18]. In addition, CpG-islands methylation state confers another category of epigenetic regulation that can be altered, such as CpG-island hypermethylation in posterior fossa group A ependymomas [14, 21]. The mutational burden is low in pediatric cancers [6, 18] and it has been described that ~8% of cases have a predisposing germline mutation. The most common germline variants are found in TP53, APC, BRCA2, NF1, PMS2, RB1 and RUNX1 [18, 22].

1.4. Pediatric sarcomas

Sarcomas are a heterogeneous group of malignancies with mesenchymal origin that develop in bones and soft tissues such as cartilages, muscles, peripheral nerves, fibrous, fat, joints, deep skin tissues or blood vessels [23-25]. This type of cancer includes more than 50 different histologic subtypes that can arise in any part of the body at any age. They represent 1% of all adult solid cancers, and 21% of all solid malignancies in pediatrics [25]. Pediatric sarcomas are the fourth most prevalent tumors after leukemia-lymphoma, brain tumors and neuroblastoma [26].

Each sarcoma type is characterized by specific molecular alterations that can be classified according to the presence of chromosomal translocations or activating mutations [27]. For instance, translocation-driven sarcomas include Ewing sarcoma, alveolar rhabdomyosarcoma (RMS) or synovial sarcoma. Translocations give rise to fusion proteins that trigger tumorigenesis by transcriptional rewiring, epigenetic dysregulation or initiation of oncogenic signaling pathways. Most of the fusions result in the expression of aberrant transcription factors such as in Ewing sarcoma, while others codify proteins with tyrosine kinase activity like in congenital fibrosarcoma, or autocrine growth factors such as in dermatofibrosarcoma. [23, 27, 28]. It has been described that these specific chromosomal rearrangements may be the consequence of increased gene size and intron length of the fusion genes, higher access to genes due to open chromatin conformation, proximity of the genes after three dimensional rearrangements, or synchronized gene expression within shared transcriptional hubs [23]. Epigenetic dysregulation is also observed in sarcomas that do not carry translocations, such as embryonal RMS or in rhabdoid tumors, the latter of which presents mutations or deletions

affecting SMARCB1 or other members of the SWI/SNF complex. As an exception to the low burden of genetic alterations characteristic of pediatric tumors, osteosarcoma is distinguished by chromothripsis and genomic complexity [26].

2. Ewing sarcoma: a unique chromosomal translocation in a suitable epigenetic environment

2.1. Epidemiology

Ewing sarcoma is an aggressive malignancy that arises in bones and soft tissues. Worldwide, it is the second most common malignant bone cancer in children and young adults, and the first in Spain. Peak incidence is 15 years of age [25]. This peak is consistent with an increased production of growth hormone during adolescent growth spurt that facilitates fast growth of bone and associated tissues [2, 7].

Ewing sarcoma represents 2% of all pediatric cancers affecting 1.5 cases per million people each year [29]. In Spain, the incidence rate is 5.2 cases per million people per year between the ages of 0 and 14 [30]. Incidence is significantly higher in European ancestry population compared to Asians (0.8 cases per million) or African (0.2 cases per million). Americans with African ancestry have also a lower incidence when compared to Americans with European ancestry. Environmental or lifestyle factors seem not

determinant for Ewing sarcoma. Instead, genetic variants of European ancestry influence disease risk. Prevalence is also slightly predominant in males, with a 3:2 ratio. [31-32].

Most pediatric and adult sarcomas are soft tissue sarcomas, only 10% are bone sarcomas [25]. However, 80% of patients with Ewing sarcoma have tumors in the bones [33]. Primary sites affecting bones include lower extremities (41%), pelvis (26%), chest wall (16%), upper extremities (9%), spine (6%), hand and foot (3%) and skull (2%). Extraosseous primary Ewing tumors arise predominantly in paravertebral and thoracic soft tissues near the esophagus, pancreas, ileum, kidney, bladder, vagina, uterus, penis, adrenal gland, lung, breast, spinal cord, orbit, and intracranial tissue. 32% in the trunk, 26% in the extremities, 18% in head and neck, 16% in the retroperitoneum and 9% in other sites [29, 34].

In recent years, the 5-year survival rate has increased to 75-80% for children younger than 15 years and to 65% for patients 15 to 19 years [34]. However, the prognosis for patients with Ewing sarcoma over 18 years of age remains dismal, and it has been described that age older than 18 years is an independent poor prognostic factor more relevant than metastasis [30]. Overall survival for patients with high levels of serum lactate dehydrogenase (LDH), large primary neoplasms, pelvic tumors and incomplete tumor regression after treatment is also lower. Ewing sarcoma has a predisposition to early metastasize to the lungs, bone and bone marrow. Metastasis in lymph nodes, liver and brain are rare. About 20-25% of patients present with metastasis at diagnosis and have a 5-year overall survival of <30%. Patients with metastasis exclusively in the lungs have a slightly better prognosis with around 50% overall survival [24, 29, 31].

2.2. Clinical features and diagnosis

Ewing sarcoma patients usually present with intermittent pain and swelling. The pain increases at night or after exercise. In 10 to 15% of the cases patients develop a bone fracture. Some patients present with a palpable and painless mass. The diagnosis is more difficult for patients with pelvic, chest wall or femoral primary tumors. Cases with advanced disease can suffer from nonspecific constitutive symptoms such as fever, night sweats, fatigue, and weight loss [29, 31]. The lack of specific symptomatology explains a median time of 2 to 5 months from initial symptoms to final diagnosis [34]. Nevertheless, time to diagnosis is not related to metastasis or prognosis of the disease [31, 34].

The diagnosis starts with physical examination and medical history since Ewing sarcoma may associate with rare germline mutations [31]. A complete blood test should be run including blood count, blood serum chemistry, erythrocyte sedimentation rate and coagulation tests. These tests can reflect the levels of inflammation and bone remodeling markers, like alkaline phosphatase. In addition, high levels of LDH associate to tumor burden and are a marker of poor prognosis [29, 31, 34]. Diagnostic imaging includes radiological evaluation; computed tomography of the lungs; and bone scintigraphy. Radiographs usually show lytic bone lesions that appear as a destructive mass or “moth eaten” pattern (**Figure 2A**). Images also might show a multilayered appearance of the diaphyseal–metaphyseal bone known as “onion peel” that indicates the displaced periosteum, and a new subperiosteal bone growing known as the Codman triangle (**Figure 2A**) [29, 31]. Magnetic resonance imaging is used mainly to evaluate the soft tissue compartment. Functional imaging includes the 18-fluorodeoxyglucose positron emission tomography (FDG-PET) which

evaluates the consumption of FDG by the tumor. 18-FDG-PET scan allows to evaluate the presence of bone marrow metastasis making bone marrow aspiration not necessary anymore [31, 34].

Final diagnosis always comes from biopsies of the tumor mass. Material is used for histology, immunohistochemistry (IHC) and molecular pathology.

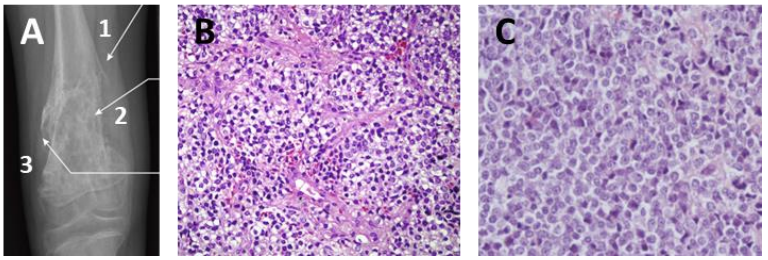


Figure 2. **Histological characteristics of Ewing sarcoma tumors. A)** X-ray images showing the Codman triangle formed as a new subperiosteal bone (1), the "moth eaten" pattern due to osteolytic lesions (2) and the "onion peel" aspect of periosteal layers (3) of an Ewing sarcoma tumor. **B-C)** Hematoxylin and eosin (H&E) staining of Ewing sarcoma tumors presenting undifferentiated small round blue cells with higher nucleus to cytoplasm ratio**.** Adapted from Grünewald *et al.,* (2018) [31] and Riggi *et al.,* (2021) [29].

2.3. Histopathology

Ewing sarcoma morphologically presents as small, round, undifferentiated blue cells, so it is included in the differential diagnosis of the small round blue cell tumors of childhood (**Figure 2B-C**) [34]. They are classified as blue cells given the higher proportion of nucleus over cytoplasm, so the eosin staining is weaker [24]. Cells include a round-to-oval nuclei with dispersed chromatin without nucleoli. Hematoxylin and eosin staining are used to detect these cells during diagnosis. Cytoplasm of Ewing sarcoma tumor cells normally presents large deposits of glycogen making them positive for periodic acid-Schiff staining. Cells grow firmly packed and following a

diffuse pattern since they do not hold a structural organization. This makes reticulin staining negative because of lack of matrix. [24, 31, 34].

In addition to Ewing sarcoma, the group of small round blue cell tumors includes neuroblastoma, DSRCT, alveolar rhabdomyosarcoma and poorly differentiated synovial sarcoma, among others [29]. These tumors share several morphological and clinicopathological features. it is therefore necessary to identify specific histological and genetic markers to correctly diagnose Ewing sarcoma. CD99 is a surface membrane glycoprotein expressed in approximately 95% of Ewing sarcoma tumors. However, CD99 is also expressed in other tumors such as lymphoblastic lymphoma and leukemia, synovial sarcoma, alveolar RMS and gastrointestinal stromal tumors. In contrast, lymphoblastic lymphoma expresses CD45, while alveolar rhabdomyosarcoma expresses Desmin, Myogenin and MyoD1, whereas Ewing sarcoma does not [24, 31, 34]. In some cases, depending on the level of neuroectodermal differentiation, Ewing sarcoma cells can display a rosette-like arrangement and express neural cell markers like neural-specific enolase, S-100, Synaptophysin and CD56 and CD57. In addition, some of these tumors can also express neurofilaments. Ewing sarcoma cells also express Vimentin and 20% of the cases Cytokeratin. Neuroblastoma cells express S-100 as well as Ewing sarcoma cells, but are negative for Vimentin and positive for neurofilament staining [24, 29, 31].

The lack of specificity in the expression of these markers prevents a final diagnosis of Ewing sarcoma, which makes necessary the use of molecular biology techniques to detect the fusion translocations and rearrangements that are conclusively diagnostic of Ewing sarcoma. Such techniques include reverse transcription polymerase chain reaction (RT-PCR), next-generation sequencing (NGS) and fluorescence in situ hybridization (FISH) [29, 31].

Regarding FISH, the commercial break-apart probes used in clinical diagnostic labs detects EWSR1 rearrangements, but not the EWSR1-ETS fusions. Since EWSR1 rearrangements can also be found in other tumors like desmoplastic small round cell tumors, EWSR1-NFATc2-translocated sarcomas or malignant rhabdoid tumors, RNA-driven techniques are used to detect the fusions [31].

2.4. Treatment

The treatment of Ewing sarcoma is multimodal and includes a combination of non-selective cytotoxic chemotherapeutics (doxorubicin, etoposide, cyclophosphamide, vincristine or ifosfamide); and local control measures including surgery and radiation therapy. Chemotherapy is used before local treatment to reduce tumor size and micrometastatic disease and after local control to remove residual cells [29, 31, 33-35]. The Children's Oncology Group (COG) published in 2012 the results of a randomized clinical trial demonstrating the effectiveness of dose intensification by interval compression. They showed that patients receiving fourteen cycles of standard doses of vincristine/doxorubicin/cyclophosphamide (VDC) alternated every two weeks with ifosfamide/etoposide (IE) cycles (VDC/IE), had an improved event-free survival (EFS). These results have established this multi-agent protocol as the standard first-line treatment for Ewing sarcoma patients [31, 33-35].

For local therapy, surgery is the recommended modality if the tumor is amenable. Otherwise, radiation treatment is required when surgery cannot ensure normal function or physical appearance [34]. Combination of

radiotherapy and surgery is used when tumors have been removed but residual microscopic disease or margins are positive [31, 34].

Although survival rates have increased during the recent years, prognosis for patients with disseminated disease, with relapse or with metastasis remains dismal. Patients with disseminated disease are treated following the same principles as for localized disease patients [31, 35]. However, they usually develop chemo and radiotherapy resistance [24, 33]. Treatment for patients with recurrent disease is individualized depending on the time from diagnosis to first relapse, region of relapse, therapies received and toxicities presented [34, 35]. The most common treatment for relapsed patients is the combination of vincristine/irinotecan/temozolomide (VIT) [35]. High-risk patients include those with large tumor volumes, poor response to chemotherapy, pelvic primary tumors, metastasis or older than 18 years [30, 35]. In these cases, high-dose myeloablative chemotherapy followed by autologous hematopoietic stem cell transplant has been tested with no clear success [31, 34]. The Spanish Sarcoma Group demonstrated the efficacy of two cycles of gemcitabine and docetaxel (G/D) in the multicenter trial GEIS-21 [30].

Despite the cure rates for patients with localized disease, the burden of side effects is notable. Patients can develop long term side effects such as reduced fertility, premature menopause, nephrotoxicity and cardiotoxicity [35]. In addition, they can develop secondary cancers associated to radiotherapy and chemotherapy such as myeloid dysplastic syndromes or leukemia [29]. Therefore, it is imperative to design new and better tumor-tailored treatments. The lack of enzymatic activity and the disordered structure of the fusion translocations that characterize Ewing Sarcoma

makes it difficult to specifically target Ewing sarcoma cells. Current research is looking for alternative mechanism-based therapies [29, 35].

2.5. Genetic alterations

Chromosomal translocations

Ewing sarcoma is characterized by a very low mutation rate (0.15 mutations/megabase) [29]. In many cases, the only genetic alteration detected is a chromosomal translocation between genes from the *FET* family and genes from the *ETS* family of transcription factors. In around 85% of the patients, the resulting chromosomal translocation is t(11;22)(q24;q12), giving rise to the *EWS-FLI1* fusion oncogene (**Figure 3A-B**). The *EWS* gene is also known as Ewing's sarcoma breakpoint region 1 (*EWSR1*). 10% of Ewings show an alternative translocation t(21;12)(q22;q12), generating the *EWS-ERG* fusion oncogene. *EWS* can also partner with the ETS transcription factors *E1AF*, *ETV1* and *FEV*, but with a frequency of less than 1%. On rare cases the *TLS/FUS* gene, another member of the FET family, fuses with ETS transcription factors (**Figure 3B**). Also, *EWS* can undergo a translocation with genes that are not from the ETS family [24, 29, 34]. The identification of these translocations by molecular techniques constitutes the definitive diagnosis of Ewing Sarcoma [24].

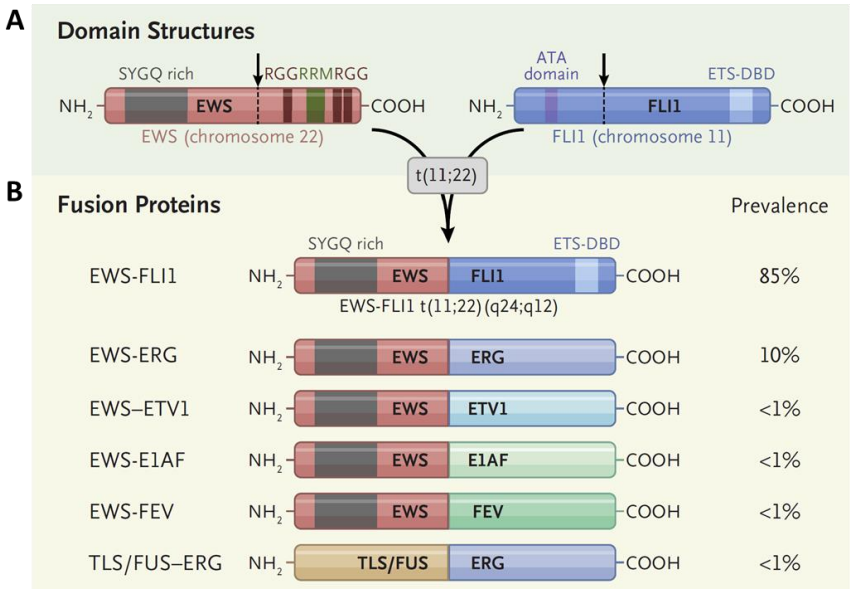


Figure 3. **Chromosomal translocations in Ewing Sarcoma tumors. A)** *EWS* and *FLI1* domain structures indicating with dashed lines the breakpoint of each gene in the chromosomal translocation. **B)** Fusion proteins formed due to a chromosomal translocation between *FET* and *ETS* genes that arise Ewing Sarcoma, and frequency of each fusion. ATA, amino terminal transcriptional activation; DBD, DNA binding domain. Adapted from Riggi *et al.,* (2021) [29].

FET proteins, including EWS, are expressed in mammalian cells ubiquitously [24]. They have the capacity to bind RNA due to an RNA recognition motif formed by 87 amino acids. Their structure is constituted by a disordered, low-complexity, prion-like N-terminal transactivation domain enriched in serine, tyrosine, glycine, and glutamine residues (SYGQ). In addition, they have three arginine-and-glycine-rich (RGG) repeats. Each RGG is different in length. RGG1 is separated from RGG2 by the RNA recognition motif, and between RGG2 and RGG3 there is a zinc-finger domain (**Figure 3A**) [29]. FET proteins are RNA-binding proteins implicated in RNA metabolism. They also regulate transcription by binding to the transcription factor TFIID, to the RNA polymerase II (RNA pol II) through the prion-like domain, and to transcription repressors and

activators such as p300-CBP [24, 29]. In addition, FET proteins interact with the splicing machinery, regulating mRNA processing [24].

FLI1, the gene that is translocated to *EWSR1* in most cases, is expressed in mammals during development in hematopoietic and endothelial cells, and in neural crest-derived mesenchymal cells [24]. FLI1 has two ETS-binding domains with a structure of helix-loop-helix, that are separated by a FLI1 specific domain (FLS). The 5' ETS and FLS domains form the amino terminal transcriptional activation (ATA) domain. The 3' ETS domain has a DNA binding domain (DBD) and forms a carboxyterminal transactivation domain (CTA) (**Figure 3A**) [24, 29].

When the translocation occurs, the N-terminal region of EWSR1 fuses to the C-terminal region of FLI1. The fusion protein that is formed usually contains exon 7 of EWSR1 and exon 6 of FLI1 in 60% of cases, or exon 5 of FLI1 in 25% of cases. Thus, different EWSR1-FLI1 transcripts can be generated depending on the breakpoint in each gene [31, 33, 34]. The ATA domain from FLI1, which is not present in the oncoprotein, is less active than the N-terminal transactivation domain from EWSR1, non-active in the wild-type protein due to inhibition of the RGG repeats. Thus, in the fusion protein the ATA domain of FLI1 is substituted by the EWSR1 domain, which is highly active after the replacement of the C-terminal part [29].

As it was mentioned before, *EWSR1* can partner with other genes of the ETS transcription factor family and with other genes forming fusions related to other sarcomas or leukemias [24, 29]. This suggests that the transactivation domain of EWSR1 is important for cell transformation, while the partner gene in the translocation would define the target genes in the cell of origin determining the tumor phenotype [29].

Non-ETS fusion proteins with EWSR1 leading to undifferentiated round-cell tumors include EWSR1-NFATC2, EWSR1-POUF1, EWSR1-PATZ1, EWSR1-SMARCA, and EWSR1-SP3 [29]. However, non-FET and non-ETS genes can translocate and originate proteins that form tumors of bone and soft tissues like Ewing sarcoma [31, 34]. In 2020, the WHO classification of undifferentiated small round-cell sarcomas included Ewing sarcoma, defined by FET-ETS gene fusions, and three main categories formed by EWSR1 with non-ETS gene fusions, CIC-rearranged sarcomas, and BCOR-rearranged sarcomas. These three categories, denominated Ewing-like sarcomas, are biologically different and are less frequent than Ewing sarcoma [31, 34, 36]. Pathological and clinical characteristics are also different, and morphology of the cells can be determinant [36]. Nevertheless, patients developing these tumors receive the same treatment protocol as for Ewing sarcoma. Although BCOR-sarcoma patients behave similarly to Ewing sarcoma, CIC-sarcoma patients have a poorer response to therapy and harbor a worse prognosis [31, 34, 36].

Additional genetic alterations

Despite the majority of Ewing sarcomas tumors only have the EWSR1-ETS translocation, which is pathognomonic for the disease, some recurrent mutations may also be present in Ewing sarcomas. It has been described that mutations leading to STAG2 loss are present in 15 to 20% of patients [31, 34, 37]. STAG2 is a protein of the Cohesin chromatin complex, which is involved in sister chromatid embracement and three-dimensional structure of chromatin [31]. Alterations in the expression of this protein can be due to point mutations, rearrangements and non-genetic mechanisms[37]. In addition, mutations in the tumor suppressor gene *TP53* can be found in 6 to 7% of the Ewing Sarcoma cases, and in 12 to 22% there

are deletions of the cell cycle regulator *CDKN2A* [31, 34]. Even less frequently mutations are found in *EZH2*, *ZMYM3* and *BCOR* genes [31]. Although none of these mutations are necessary for tumor initiation or maintenance, they are associated to poor prognosis, dissemination and may accelerate the disease progression in advanced stages [29, 37].

Copy number variations

Different studies have described recurrent chromosomal alterations in Ewing sarcoma. Sometimes these aberrations include complete chromosome 2, 8 and 20 gains. Chromosome arm variations include chromosome 1q gain and 16q loss [31]. In addition, about 25% of Ewing tumors display a gain of chromosome 1q that can occur concurrently with a deletion of chromosome 16q. This simultaneity is the result of an unbalanced translocation and is associated with poor prognosis, even as a single gain of chromosome 1q [31, 34]. Chromosome 8 aberrations involve the gain of the entire chromosome, leading to trisomy that is present in about 50% of the tumors [31]. Finally, deletion of 9p21 is also frequent and involves *CDKN2A* [31, 37].

Genetic predisposition

Germline predisposition to develop Ewing sarcoma was studied by different sequencing studies where rare polymorphisms or mutations in cancer susceptibility genes related to DNA damage repair such as *BRCA1* and *FANCC* were described [31, 33, 34]. However, the effect of this pathogenicity and whether Ewing sarcoma patients' families would require genetic counselling is unclear [31]. Genome-wide association studies were also performed demonstrating susceptibility loci with common variants affecting genes that cooperate with the *EWSR1-FLI1* oncogene. These

genes include *TARDBP*, which has a similar structure to *EWSR1*, target genes of the EWSR1-FLI1 fusion as *EGR2* and *NKX2-2*, genes of the EWSR1-FLI1 core regulatory circuitry; and genes with a role in centrosome stabilization and apoptosis. Besides, it has been described that these risk alleles show different frequencies among populations. This might help explain in part the divergent incidence of Ewing sarcoma in Caucasians, Africans or Asians [31].

2.6. Cell of origin

To identify the precursor cell of Ewing sarcoma tumors is one of the major focuses of research in the field. The undifferentiated phenotype and the fact that tumors arise in both the periosteum of bones and soft tissues makes it really hard to elucidate the cell of origin [31]. Ectopic expression of EWSR1-FLI1 in non-transformed differentiated primary cells led to cell cycle arrest and apoptosis. Only primary pluripotent stem cells confer the permissive environment required for EWSR1-FLI1 expression and maintenance [29]. Ewing tumor cells express neural stem-cell markers such as CD133 and SOX2, which suggest the histogenesis in neural crest stem cells [29, 31, 38]. However, Ewing sarcoma tumors share transcriptional signatures with mesenchymal stem cells (MSC) and the transcriptome of EWSR1-FLI1 knockdown cells resembles the one of MSCs [31, 39, 40]. Besides, sarcomas mainly arise from MSCs. This cell lineage originates from the mesoderm, the embryonic layer that will generate the skeletal and connective tissues, among others [24, 29]. Moreover, it has been described that mesenchymal cells can also derive from neural crest cells forming the mesenchyme of ectodermal origin [41]. Thus, given the histological and

transcriptional features of Ewing sarcoma tumors, it was hypothesized an origin from neural crest-derived MSCs blocked in differentiation by EWSR1-FLI1 [24, 42].

2.7. Disease models

Disease models are useful tools in research, since they allow testing new therapies and understanding the biological mechanisms of diseases. In Ewing sarcoma, several attempts have been carried out to obtain a transgenic mouse model by expressing the oncogene in different cell lineages throughout development and differentiation in a variety of genetic backgrounds. However, most of them end up with developmental defects or embryonic lethality [43]. The zebrafish model achieved the formation of tumors with similar histology to human tumors, but a p53 mutant background was required to generate these tumors. In this experimental model, p53-deficient zebrafish embryos were injected with EWSR1-FLI1 using the Tol2 transposon and 10 out of 150 embryos spontaneously formed small-round-blue-cell tumors. Microarray analysis of the tumor transcriptomes indicated enrichment of the human Ewing sarcoma signature [44]. More recently, a spontaneous mutant variant of the oncogene EWSR1-FLI1 has been shown to form tumors in Drosophila. This model recapitulates the oncogenic characteristics of Ewing sarcoma and might be suitable for further *in vivo* investigation of tumorigenesis and drug screening approaches [45].

Other studies have attempted to generate experimental Ewing sarcomas by ectopic expression of EWSR1-FLI1 in different pluripotent stem cells.

One of these approaches used murine primary bone-derived cells from long bones, the rib cage and vertebrae to stably express the oncogene: while early passage cells were immortalized but could not form tumors, late-passage cells were able to develop tumors in immunocompetent syngeneic mice subcutaneously, intraperitoneal and intravenous. Tumors lacked markers for epithelial, neural, muscle, macrophage and dendritic lineages and presented a sarcomatous morphology with the nuclear compartment larger than cytoplasm [46]. EWSR1-FLI1 expression in adult human mesenchymal stem cells (hMSC) and in pediatric hMSC could reproduce a transcriptome profile very similar to Ewing sarcoma. Nevertheless, these cells were not tumorigenic when injected into mice [47, 48]. Osteochondrogenic progenitor cells derived from the superficial part of long bones of late murine embryos were induced with EWSR1-FLI1 and Ewing sarcoma-like small tumors were obtained. However, gene expression similarities were not statistically significant, and tumors were similar to neuroblastoma and synovial sarcoma [49]. Human embryonic stem cells (hESC) expressing the oncogene and a short hairpin RNA targeting the tumor suppressor gene *TP53* were established to give rise to embryoid bodies. Ewing sarcoma gene expression signature and transformation features were observed *in vitro*, but when embryoid bodies were injected subcutaneously into immunocompromised mice tumors were not formed[50].

3. Epigenetics: molecular mechanisms underlying chromatin organization

The overall mechanisms that control cellular differentiation rely mainly upon epigenetic regulation. Epigenetic mechanisms will enable chromatin to change its configuration according to each development stage. Therefore, the epigenetic state will contribute to the phenotype of developmental tumors, especially Ewing sarcoma [14, 51, 52].

3.1. Chromatin structure and organization

Chromatin is a highly organized structure in the nucleus of eukaryotic cells that packs the DNA. It is a dynamic complex that includes the DNA and small basic proteins that enable to condensate the approximately 2 meters of DNA, in the case of human cells, into a length up to ~2 × 10^{-5}-times smaller. Most of the proteins forming the chromatin are histones which are highly conserved in eukaryotes. Histones have a high proportion of basic amino acids arginine and lysine to facilitate the binding of the negatively charged DNA. H1, H2A, H2B, H3 and H4 are the five major groups of histone proteins. Histones H2A, H2B, H3 and H4 are considered the core histones and are formed by a folding domain that includes three α-helices connected by two loops to promote interactions among them. In addition, histones have a positively charged N-terminal tail that interacts with the DNA. Two copies of histones H2A, H2B, H3 and H4 are congregated forming an octamer with 146-147 base pairs of wrapped DNA forming a nucleosome, which is the basic unit of chromatin (**Figure 4**). Nucleosomes are associated to the linker histone H1 forming a chromatosome and are

joined by a linker DNA of about 80 base pairs associated to form 10 nm beads-on-a-string fibers with a beaded appearance. A higher order of chromatin organization is achieved with the formation of 30 nm fibers by chromatosome arrays packed in a two-start helical model due to the interactions between core histones [54, 55].

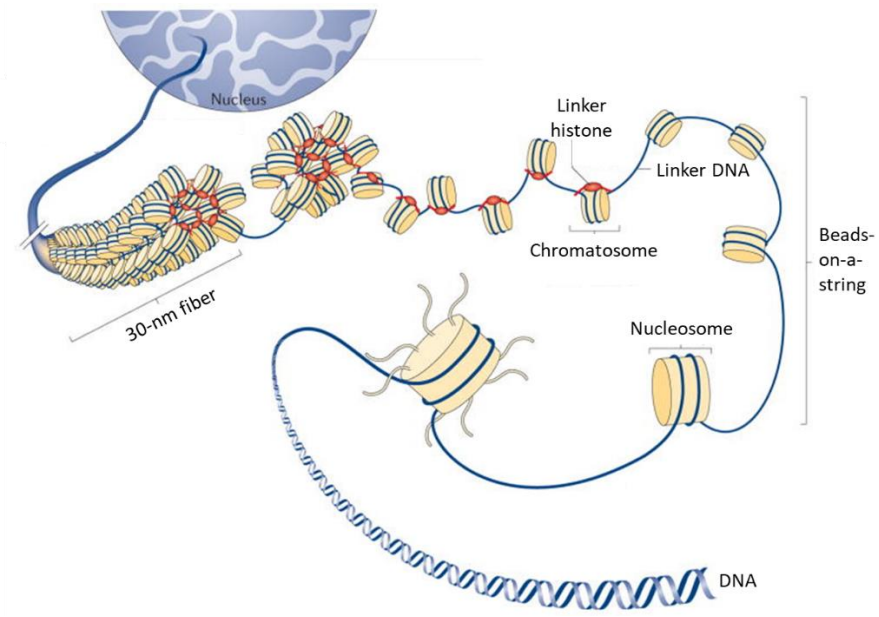


Figure 4. **Chromatin organization in eukaryotic cells.** DNA and core histone proteins form nucleosome units which in turn form bead-on-a-string dispositions, chromatosome particles with the help of H1 linker histone and 30-nm fibers which will condensate to form chromosome structures in the cell nucleus. Adapted from Fyodorov *et al*., (2018) [56].

Apart from histone proteins, there are non-histone chromatin proteins that contribute to organize the chromatin. This is an heterogeneous group of proteins including transcription factors that bind to DNA in the linker or nucleosome to regulate gene expression, proteins that limit the access of transcription factors to the DNA such as HP1 or the Polycomb group of proteins; and structural proteins that alter the chromatin fibers and the chromatin dynamics like HMG, CTCF and cohesins [54, 55, 57].

The highest level of chromatin condensation is the chromosome with the degree of condensation changing along the cell cycle. During interphase, most of the chromatin is relaxed and decondensed allowing DNA replication and gene expression. In these conditions, chromatin is named euchromatin and is organized as 30-nm fiber forming large loops or 10-nm fibers. Part of the chromatin is still tightly condensed and compacted forming heterochromatin, which favors a transcriptionally inactive state in these regions. Heterochromatin coincides with the condensed state that is present during cell mitosis when the X-shaped metaphase chromosomes are formed to be accurately separated to each daughter cell. Gene transcription is blocked at that stage. Therefore, chromatin structure is associated to gene expression regulation in eukaryotes [54].

3.2. Definition of epigenetics

Conrad Waddington coined the term epigenetics. He described it as "the branch of the biology which studies the casual interactions between genes and their products, which bring the phenotype into being" [57, 58]. All the cells in an organism are genotypically equal. However, cells with different differentiation states, morphogenesis, functions, variability and adaptability to environmental changes exist. Therefore, cell type specialization needs a precise regulation of gene expression. Epigenetics consists of heritable molecular changes that alter gene expression at the transcriptional and post-transcriptional level, and also at the translation and post-translation level without altering the DNA sequence. These changes include DNA modifications, DNA-binding proteins, histone modifications and RNA-mediated processes. Nowadays the definition of

epigenetics has become more precise as it is understood as the bridge between the genotype and the phenotype of a specific cell type by molecular information that is not encoded in the DNA sequence [57-60]. This information can be inherited and transferred to the progeny during mitosis and meiosis, but it can also be erased when histone modifications are removed [57, 59, 60].

Epigenetic information can be classified as *cis* mechanisms, which are distinctive molecular marks associated with chromatin and inherited during DNA replication. *Cis* mechanisms comprehend covalent modifications in the DNA, like DNA methylation, or modifications in histones. Non-histone proteins, higher-order chromatin structure and nuclear localization can also encode *cis* epigenetic information. The other type of signals corresponds to *trans* mechanisms, when self-propagating transcriptional states are regulated by transcription factors and feedback loops. Transcription factors are inherited in each cell division even when the developing stimulus is withdrawn and then restart their function in the regulatory regions of the DNA sequence [59].

3.3. Regulation of chromatin organization

Chromatin structure and organization determines the accessibility of DNA to undergo gene transcription and DNA replication and repair. The dynamic processes to transform heterochromatin regions into euchromatin or vice versa are performed by epigenetic mechanisms. Therefore, the modifications that the chromatin suffers are essential regulators of gene transcriptions and, hence, of cell identity. These modifications include DNA

methylation, histone modifications and non-coding RNAs, among others (**Figure 5**) [57, 60- 62].

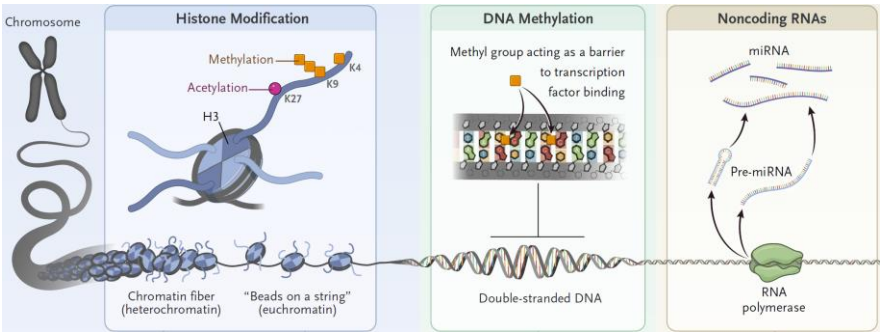


Figure 5. **Chromatin modifications contributing to DNA accessibility.** Epigenetic mechanisms that regulate chromatin structure and accessibility for gene transcription or repression are histone modifications such as methylation or acetylation of histone tails, DNA methylation and non-coding RNAs like microRNAs (miRNAs). Adapted from Riggi *et al.*, (2021) [29].

DNA methylation

Epigenetic regulation by DNA methylation consists in the covalent transfer of a methyl group to the 5-carbon position of the DNA's cytosine ring. The reaction is catalyzed by DNA methyltransferases (DNMTs). Four DNMTs have been described in mammals: DNMT1, which maintains DNA methylation during replication; DNMT2, that is involved is weak DNA methylation; and DNMT3a and DNMT3b, which methylate DNA *de novo* [60]. DNA methylation acts as a repressive mark by preventing binding of transcription factors [63]. 98% of DNA methylation takes place in a CpG dinucleotide environment, also called CpG islands. They are DNA sequences that are guanine and cytosine-rich [63, 64]. When these regions are non-methylated, as in promoters or enhancer regions, they enable the creation of transcriptionally active chromatin states. Instead, when they are methylated gene expression is silenced [64]. In ESCs, part of the methylation also appears in a non-CpG context [63]. Methylated DNA is

recognized by the methyl-binding family of proteins MBD that will recruit other chromatin remodelers or histone modifying enzymes which are also going to contribute to gene transcription regulation [60].

Histone modifications

Alteration of chromatin structure due to epigenetic mechanisms can be determined by histones. These mechanisms include the presence of different histone variants, changes in their position in chromatin and diverse post-translational modifications (PTMs) that will result in two main functions: the formation of chromatin environments and the arrangement of DNA-based biological processes [29, 59, 65]. To establish the appropriate chromatin state, modifications enable the development of euchromatin domains for gene transcription and heterochromatin domains for gene repression. DNA-based functions involve gene transcription, DNA repair and replication and chromosome condensation [65].

Histone PTMs are found mostly in histone tails and comprise acetylation, methylation, phosphorylation, ribosylation, ubiquitylation, sumoylation, deimination and proline isomerization. The most well-defined are the small covalent modifications acetylation, methylation and phosphorylation. Since histones are rich in the basic aminoacids arginine and lysine, the possibilities of PTMs are huge. Lysine residues can accept from one to three methyl residues and arginine can be mono- or dimethylated. Modifications will depend on cell signaling and will have different effects [60, 65]. Gene activation is associated to histone H3 acetylation of lysine 27 (H3K27ac), while methylation can be involved in both activation and repression [29]. For instance, H3 monomethylation of lysine 4 (H3K4me1) characterizes active and poised enhancers, which are

distal DNA *cis*-regulatory elements that regulate gene transcription independently of their orientation, while H3 dimethylation of lysine 4 (H3K4me2) defines active gene body marks, and H3 trimethylation of lysine 4 (H3K4me3) is a mark of active and poised transcription start sites[66]. In contrast, transcriptional repression is associated to H3 trimethylation of lysine 9 (H3K9me3) and lysine 27 (H3K27me3) [29].

Histone PTMs are mediated by enzymes that can contribute to or reverse the modifications. Histone-modifying enzymes that can add and remove acetylation are histone acetyltransferases such as p300-CBP and histone deacetylases (HDACs). Methyl groups are added or removed by histone methyltransferases as the mixed lineage leukemia (MLL) complex and by histone demethylases in lysine residues (i.e. lysine-specific histone demethylase 1, or LSD1), respectively. Arginine methylation can be removed by deiminases. Histone phosphorylation is carried out by histone kinases that phosphorylate residues such as serine, tyrosine and threonine, and phosphatases revert these modifications [29, 59, 60].

Histone modifications will result in their specific effects by disruption of interactions between nucleosomes through recruitment of non-histone proteins that are ATP-dependent chromatin remodeling complexes. These complexes cooperate with the histone-modifying enzymes to allow transcription or to conserve a repressed chromatin state, since modifications will disrupt histone charges resulting in chromatin structural changes. Chromatin remodeling complexes are recruited depending on the modifications found in each histone and their binding to specific domains [60, 65]. These complexes alter histone interactions and histone-DNA communications. They can act remodeling nucleosomes by mobilizing nucleosomes through the DNA to expose a previous closed region

(nucleosome sliding), by nucleosome eviction to give access to the associated DNA, or generating nucleosomes wrapping the DNA and changing the distances between nucleosomes in improper arrays (nucleosome spacing) [67].

Chromatin remodeling complexes are grouped in four main families. The SWI/SNF complex contain a helicase-SANT (HAS) domain, but is more described by having a carboxy-terminal bromodomain that recognizes acetyl groups in lysine residues. The ISWI family has a SANT-SLIDE module, as well as the CHD family. In addition, this last family has chromodomains in tandem that interact with methyl groups. Finally, the INO80 group of proteins possess an insertion between the two ATP domains longer than the other families. They are involved in genome processes such as gene expression regulation, replication and DNA damage repair [67].

Apart from histone PTMs, core histones can be exchanged by other H2A and H3 histone variants with their own modification patterns that are less frequent, such as H3.3 and H2A.Z. Chromatin remodeling complexes can help the replacement of these variants acting as exchanger complexes [58, 60, 67].

Non-coding RNAs

Non-coding RNAs (ncRNA) are RNA transcripts that are not translated into proteins, but can be functional [68]. They are part of the epigenetic regulation and include, among others, transfer RNAs (tRNAs), ribosomal RNAs (rRNAs), small nuclear RNAs (snRNAs) and small nucleolar RNAs (snoRNAs) [59, 60]. They can regulate transcription, translation and splicing by recognition of specific sequences or RNA substrates or by catalysis. They

are classified as small ncRNAs if its length is <200 nucleotides or long when they have >200 nucleotides [60, 68].

Small ncRNAs usually arise from cleavage of RNA precursor molecules by the RNAse III-family of enzymes, such as Drosha and Dicer. MicroRNAs (miRNAs) are single-stranded small ncRNA molecules of about 20-30 nucleotides that originates from defective hairpin structures of lncRNAs precursors or by introns. They are cleaved by DROSHA and DICER and they inhibit translation of target mRNAs by binding them or by degradation through the RNA-induced silencing (RISC) complexes. Another type of small ncRNAs are short interfering RNAs (siRNAs), which are similar in length to miRNAs. They arise from double-stranded RNA precursors cleaved by DICER and degrade their target mRNAs. PIWI-interacting RNAs and repeat-associated RNAs are other classes of small ncRNAs [60].

LncRNAs can be products of transcriptional “noise” or artefacts, or they can act as small ncRNAs precursors and regulate gene expression by inducing or repressing transcription of a downstream target gene. To do so, they can modify RNA Polymerase II recruitment or change chromatin configuration [60].

In addition, ncRNAs can interact with the DNA methylation machinery, chromatin remodeling complexes and histone modifying enzymes to add repressive marks or activating modifications that will generate the epigenetic environment of each cell. For instance, inactivation of X chromosome in females requires a crosstalk between ncRNAs, DNA methylation, histone modifications and their modifying enzymes. In addition, their expression can be regulated by histone post-translational modifications and DNA methylation [58-60, 68].

3.4. Transcriptional regulatory elements in chromatin: promoters, enhancers and super-enhancers

Cell identity is determined by the organized transcription of genes that will display different functions. For transcription to occur the enzyme RNA polymerase (RNA pol) cooperates with initiation factors to recognize promoter regions, which are DNA sequences upstream of a gene. They are in a closed chromatin state until they are recognized by the RNA pol and the pre-initiation complex (PIC) is formed. The RNA pol opens the double DNA strand and starts to synthetize RNA until the termination signal. Three different types of RNA pol exist: RNA pol I, which synthetizes ribosomal RNAs; RNA pol II, to transcribe mRNAs and different non-coding RNAs; and RNA pol III, for transfer RNAs and small ribosomal RNAs [69].

In order to open chromatin and recognize RNA Pol, promoters are activated by transcription factors that can also bind to enhancer regions. Enhancers are cis-regulatory elements of few hundred base pairs (bp) that regulate transcription. When enhancer regions are located continuously within 12.5 Kb of each other, they are called super-enhancers (SE) [70]. SEs are large clusters of enhancers of about 50 Kb characterized for being widely occupied by master transcription factors, which recognize their consensus binding sequences. Large complexes are also recruited to SEs, and many different proteins can be found in these regions, such as RNA pol II, the Mediator complex, cohesin proteins like NIPBL, or the member of the bromodomain protein family BRD4. Histone modifying enzymes such as the acetyltransferase p300-CBP or the LSD1-NuRD complex, proteins from the chromatin remodeling complexes SWI/SNF and CHD7, or methyl-CpG binding proteins as MBD3 are likewise found enriched at SEs (**Figure 6**). Thus, cofactors and chromatin regulators that have a role in gene

activation, DNA looping, histone modification and nucleosome remodeling are enriched in SE regions. As a result, the H3K27ac mark is highly enriched in SEs, and SE-regulated genes are more expressed than typical enhancer regulated genes. Consequently, SEs are more sensitive to perturbations due to reduction in Mediator activity or expression, and display enhanced DNase I sensitivity *in vitro* [71-74].

SEs regulate the expression of lineage-specific genes since they are recognized and bound by key transcription factors important for cell identity, while typical enhancers regulate the expression of both cell type-specific genes and genes expressed in multiple cell types. Therefore, SEs are important to establish and maintain cell state [71-74].

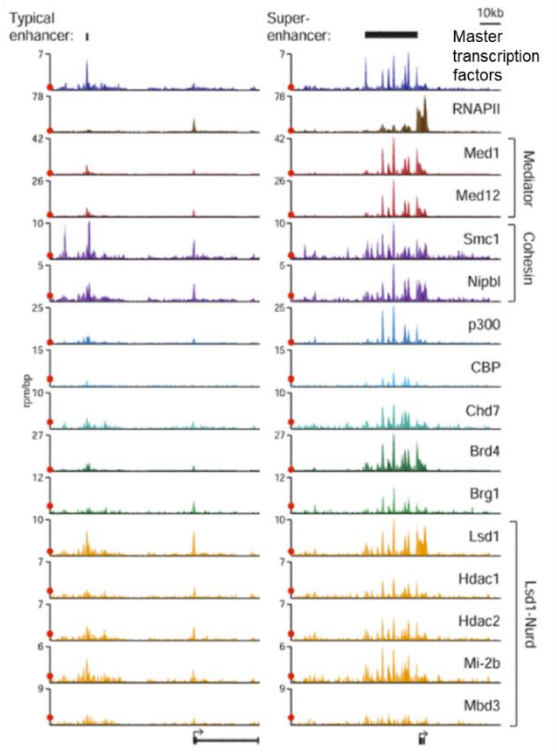


Figure 6. **Super-enhancers are highly enriched in transcription factors, RNA Polymerase II, cofactors and chromatin regulators.** Chromatin immunoprecipitation-sequencing (ChIP-

seq) profiles in a typical enhancer and a superenhancer in embryonic stem cells (ESCs) of master transcription factors, RNA pol II, Mediator subunits MED1 and MED12; cohesin complex subunits SMC1 and NIPBL, p300, CBP, CHD7, BRD4, SWI/SNF complex component BRG1, and LSD1-NuRD complex components LSD1, HDAC1, HDAC2, Mi-2b and MBD3. RNAPII, RNA pol II; rpm/bp, reads per million per base pair. Adapted from Hnisz *et al.*, 2013[74].

4. Mediator: a multitask complex essential for chromatin functions

4.1. Definition and structure

Mediator is a large multiprotein complex with a molecular mass of 1.4 MDa that communicates regulatory signals from DNA-bound transcription factors directly to RNA pol II [71, 72, 75]. It is composed of about 30 subunits that are numbered from MED1 to MED31, and CDK8, CDK19 and Cyclin C. These subunits are distributed in four modules (**Figure 7**): the head, the middle, the tail and the kinase module [71, 72, 75, 76]. The head and the middle form the core part of the Mediator complex [77, 78]. The tail and the kinase module have regulatory functions [77]. The kinase module is the only with enzymatic activity and is formed by CDK8 or its paralogue CDK19, Cyclin C, MED12 or its paralogue MED12L, and MED13 or its paralogue MED13L [71, 72].

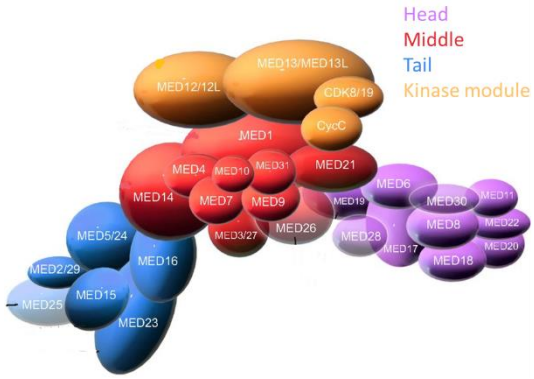


Figure 7. **The Mediator complex.** Mediator complex subunits are distributed in four different modules: the head, the middle, the tail and the kinase module. Adapted from Yin *et al.,* (2014) [71].

Even though Mediator is highly conserved in eukaryotes, subunit composition and sequences have been rearranged during evolution in different organisms [71, 72, 75, 76, 79]. In addition, subunit composition can change depending on the cell type or context, and subunit functions can be gene-specific. The complex is not restricted to a unique isoform and paralogues of some subunits in higher order organisms also confer variability to Mediator [71, 72, 75, 76]. However, two principal variants with or without the kinase module of the complex exist, since the kinase module binds to Mediator reversibly [71, 72, 75-77].

4.2. Functions

Mediator is a key regulator of protein-coding genes transcription due to its binding to RNA pol II. Diverse subunits of the complex interact with the transcription factors and coactivators bound to the DNA regulatory regions to transduce their signals to the transcription machinery in the promoters of the target genes and contribute to both activation or repression. Therefore, Mediator acts as a transcription coordinator [71, 72, 75-78].

Given the high variability and redundancy of Mediator proteins, it has been described that the complex is able to change the composition of its subunits to engage in other functions and ensure specific cellular needs (**Figure 8**) [75, 77, 78]. Besides facilitating transcription initiation, it interacts with other coactivators and corepressors that participate in transcription elongation and termination, mRNA splicing and ncRNAs activation, acting at the chromatin level [71, 75, 78]. In addition, Mediator has a role in cell metabolism and ontogenesis [72]. Thus, Mediator is a master coordinator of development and cell lineage specificity, and its dysregulation has been related to many developmental pathologies and cancers [71, 72].

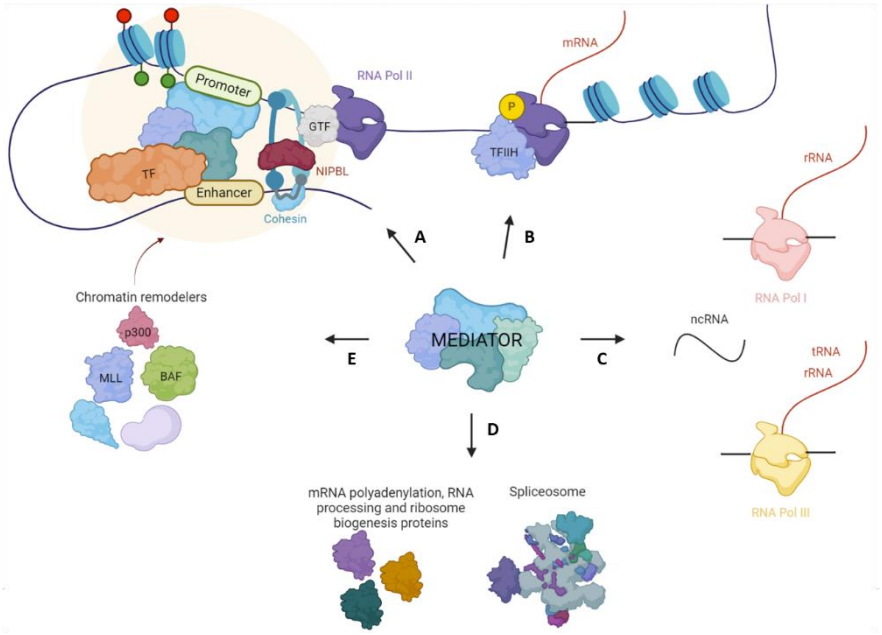


Figure 8. **Functions and interactions of the Mediator complex. A)** Mediator interacts with transcription factors and coactivators bound to enhancers and, with the help of cohesin complex and NIPBL, forms a DNA loop to recruit RNA pol II to promoters of target genes and activate transcription. **B)** Mediator recruits and activates the elongation factor TFIIH, which phosphorylates RNA pol II causing mRNA elongation. **C)** Transcription activation can be mediated by Mediator interacting with ncRNAs, and with RNA pol I and III for non-coding rRNA and tRNA transcription. **D)** Post-transcriptional modifications such as alternative splicing, mRNA polyadenylation, RNA processing and ribosome biogenesis could also be regulated by Mediator. **E)** Interactions of the Mediator complex with chromatin remodelers

contribute to activate transcription (red arrow) and to establish and maintain chromatin conformation. BAF, BRG1/BRM associated factor; GTF, general transcription factors; MLL, mixed lineage leukemia; TF, transcription factor. Created with Biorender.com.

Master transcription coordinator

Transcriptional regulation depends on DNA sequences that are recognized by gene- and tissue-specific transcription factors in response to developmental and environmental signals. These transcription factors recruit co-activators like p300 to transmit the regulatory signals and activate the transcription machinery consisting in general transcription factors (GTFs) TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH, and the RNA pol II (**Figure 8A**). GTFs are regulatory proteins required for RNA pol II to bind specifically to core promoters and form the pre-initiation complex (PIC) to start transcription. Mediator is one of the co-activators that participates in PIC formation [75-77]. The head module of Mediator interacts with RNA pol II at the C-terminal domain (CTD) of the RNA pol II subunit RPB1. Both the head and the middle modules interact with GTFs at promoters. Subunits from the tail module contact with specific transcription factors and activators bound to the regulatory DNA sequences, such as enhancers. Thus, Mediator acts as a scaffold for PIC assembly and stabilization on promoters [72, 75, 76, 78].

Enhancers are located in distal regions from the transcription start sites (TSS) in promoters, and Mediator acts as a bridge that allows transcription factors bound to regulatory regions to recruit RNA pol II to promoters of their target genes (**Figure 8A**) [72, 76, 77]. Consequently, when Mediator is bound to transcription factors and transcriptional activators, it undergoes a conformational change that enables it to interact with the components of the cohesin complex SMC1A, SMC3 and STAG2, and the cohesin-loading factor NIPBL forming a cell-type specific DNA loop [80].

Master regulators are key transcription factors that control lineage-specific transcriptional programs. Many of these master regulators interact with Mediator subunits to govern transcription and have a role in cell fate determination [71]. Therefore, Mediator is considered an integrative hub between regulatory regions and transcription machinery that coordinates cell lineage specificity by controlling master regulators [71, 75].

Elongation and termination

Apart from transcription initiation, Mediator has a role in elongation and termination [71, 72, 76]. When Mediator contributes to PIC formation and stabilization, it also triggers the recruitment and enzymatic activity of the transcription elongation factor TFIIH. Paused RNA pol II CTD is phosphorylated (p-CTD) at the Ser5 residue by TFIIH causing Mediator dissociation due to being unable to bind to p-CTD (**Figure 8B**). This makes RNA pol II to separate from the gene promoter and start the elongation step [71, 72, 77]. In addition, Mediator is also involved in termination by MED18 recruitment of termination factors to the 3' end of genes [71].

ncRNAs

Mediator can interact with ncRNAs to mediate transcriptional activation (**Figure 8C**) [71]. MED12 is recruited to regions where ncRNAs are synthetized to stablish DNA loops and activate target genes [71, 72, 81]. Besides, Mediator interacts with the fourteen subunits of RNA pol I, which is responsible for transcription of the non-coding rRNA forming the large subunit of the ribosome. RNA pol III transcribes rRNA of the smallest ribosome subunit and non-coding tRNA, and also interact with Mediator [72].

RNA processing

Mediator has also functions in post-transcriptional events by regulating alternative splicing (AS), alternative mRNA polyadenylation, RNA processing and ribosome biogenesis. The complex interacts with factors involved in mRNA 3' processing and mRNA degradation (**Figure 8D**) [72]. For instance, MED23 cooperates with HnRNP-L ribonucleoprotein in the regulation of alternative splicing, alternative cleavage and in polyadenylation events [71, 72]. MED31/MED7 complex interacts with TREX2, an mRNA nuclear export factor. In addition, the polyA binding protein PABP and the mRNA degradation protein XRN1 can also interact with Mediator[72].

Epigenetic and chromatin regulator

Cell-fate decisions orchestrated by Mediator take place by regulating master transcription factors and PIC formation, but also by interacting with epigenetic regulators related to chromatin architecture (**Figure 8E**) [71, 72, 75]. For example, Mediator can interact with histone modifying enzymes like p300, the H3K9 methyltransferase G9a, the histone arginine methyltransferase PRMT5, or with WDR77 to recruit the DNA methyltransferase DNMT3A [71]. In addition, Mediator binds to histone H3 and H4 tails and to chromatin remodeling proteins to modify the DNA-histone structure, favoring an euchromatin state with active chromatin markers and allowing the formation of PICs. Epigenetic repression may as well be due to kinase module interactions. In addition, the presence of Mediator at some loci is not associated with transcriptional activity, suggesting a role in chromatin architecture maintenance [72].

Role of the Mediator kinase module

The kinase module reversibly binds to the other Mediator subunits by MED13 interaction [79]. The head and the middle modules interact with the kinase module in the RNA pol II binding site. This results in the Mediator core displaying a closed conformation that sterically prevents RNA pol II binding, leading to gene repression (**Figure 9**) [75, 76, 79]. The kinase module is also present in enhancer regions with the complete Mediator complex and, when it is dissociated from the Mediator core, PIC formation and gene transcription can take place [72, 78, 79]. Thus, repression of RNA pol II binding by the kinase module occurs independently of the kinase activity. However, transcriptional inactivation can also occur by CDK8-dependent phosphorylation of transcriptional activators or initiation GTFs, for instance, by phosphorylating the E2F1 transcription factor [76, 82].

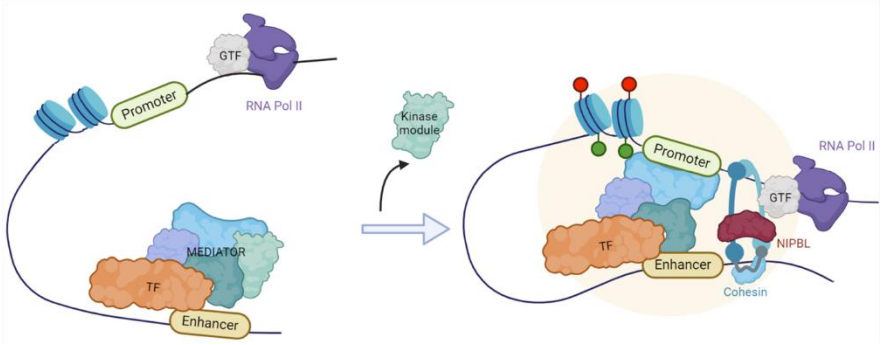


Figure 9. **Transcriptional repression by the kinase module.** The kinase module reversibly blocks the interaction of RNA pol II with the Mediator core, resulting in transcription inactivation. GTF, general transcription factor; TF, transcription factor. Created with Biorender.com.

In addition, the kinase module may favor transcription elongation. When Mediator separates from the promoter region, it binds to the kinase module to enhance the recruitment of CDK7 and CDK9. These cyclin-dependent kinases are necessary for P-TEFb and BRD4 to stimulate the

elongation by RNA pol II [71, 72, 76, 77]. Other functions in transcription of the kinase module depend on the cellular context and the binding of other transcriptional activators and coactivators [72, 75], and many developmental pathways depend on subunits of this module such as MED12 and MED13 independently of the kinase function [75, 76]. For instance, MED12 cooperates with p300 in essential active enhancers of hematopoietic stem cells [83].

5. Epigenetic rewiring triggered by EWSR1-FLI1

5.1. Epigenetic deregulation

The EWSR1-FLI1 oncoprotein is located in the nucleus and binds to DNA due to the DBD of the FLI1 counterpart [24, 29]. Besides, the N-terminal of the fusion protein includes a potent transactivation domain [24]. Due to the translocation, the 3' part of the fusion that corresponds to FLI1 suffers a conformational change that enables the activation of a more extensive repertoire of different genes than the wild-type FLI1 [29]. This allows the oncoprotein to act as an aberrant transcription factor that will deregulate the transcription of several target genes [84]. Different studies have shown that, apart from inducing the expression of different genes, EWSR1-FLI1 also have repressive functions [24, 29].

Considering the low mutation rate of Ewing sarcoma, the fact that the fusion protein is the only genetic alteration required to induce tumorigenesis and that it is only tolerated in a stem cell context, it has been postulated that the oncogene is responsible for tumorigenesis through mechanisms of epigenetic deregulation [29, 47, 85, 86]. Therefore, EWSR1-FLI1,

acting as an aberrant transcription factor, would be able to regulate lineage-specific epigenetic networks that will modify chromatin structure and DNA accessibility to promote Ewing sarcoma tumorigenesis in cells with the permissive cellular context [29, 59].

5.2. Binding to GGAA repeats

Transcription factors of the ETS family have a DBD ETS domain. The specific binding sequences contain a conserved core sequence with high affinity to "GGAA" repeats [87, 88]. It has been described that EWSR1-FLI1 presents similar DNA binding affinity and specificity to GGAA repeats [89]. The sequencing of the chromatin enriched in EWSR1-FLI1 binding revealed the presence of highly repetitive GGAA elements in its target genes [90]. These repetitive elements are microsatellites formed by 10-100 copies of 2-6 bp sequence repeats found in non-coding regions [91]. In some of the genes transcriptionally activated by EWSR1-FLI1, there is an enrichment in GGAA microsatellites and their expression is regulated according to the number of these elements. On the contrary, genes that are down-regulated by the oncogene do not harbor these repetitive motifs [90, 91]. In addition, although other ETS transcription factors can bind to GGAA repeats, only EWSR1-ETS fusions are able to regulate transcription through these elements [92]. Thus, GGAA repetitive elements, which are not conserved in other organisms, are considered as the EWSR1-FLI1 response elements [85, 90-92].

Given the importance of GGAA microsatellites length for DNA binding and gene activation, and the differences in the incidence of Ewing sarcoma in diverse human populations, polymorphisms in GGAA repetitive elements

were studied among Caucasian and African populations. GGAA repeats were significantly larger and with a greater number of repeats in the population of African ancestry than in the Caucasian population. These characteristics may result in a minor transcriptional activation by EWS-FLI1, and could explain the higher frequency of Ewing sarcoma in European ancestry populations [29, 91, 93].

5.3. Rewiring of chromatin marks and binding proteins

In order to enable EWSR1-FLI1 to generate the chromatin structure necessary for gene transcription, different histone modifications are established. The majority of EWSR1-FLI1 binding sites are enriched in the open-chromatin-associated histone mark H3K4me1. In some cases, EWSR1-FLI1 binds to promoter regions enriched in H3K4me3, which characterizes the transcriptional initiation sites. Other EWSR1-FLI1 binding sites correspond to enhancer regions decorated with the H3K27ac histone mark [85, 86]. As mentioned before, these enhancer *cis*-regulatory elements are enriched in GGAA repeats that regulate the transcription of EWSR1-FLI1-target genes [85]. Binding of EWSR1-FLI1 to these microsatellites enables the recruitment of the SWI/SNF complex by the tyrosine residues of EWSR1 that are crucial for phase transition of its prion-like domain of EWSR1 (**Figure 10A**) [94]. The SWI/SNF complex enables chromatin relaxation by blocking the deposition of H327me3 repressive histone mark [29, 94]. Besides, EWSR1-FLI1 interacts with other transcriptional activators as p300, leading to an active enhancer state by histone acetylation of H3K27 [29, 85]. MLL complex, which methylates H3K4 histone residues to activate transcription, can also be recruited by the fusion oncogene [29, 66, 85].

Therefore, multimers of EWSR1-FLI1 bound to GGAA repeats, by recruiting histone modifying enzymes and chromatin remodelers, are able to form accessible euchromatin regions that otherwise would be in a closed conformation. This allows EWSR1-FLI1 to act as a pioneer transcription factor that enables the formation of lineage specific *de novo* enhancers that will interact with their corresponding promoters, leading to the expression of target genes involved in proliferation and survival [29, 31, 85, 86].

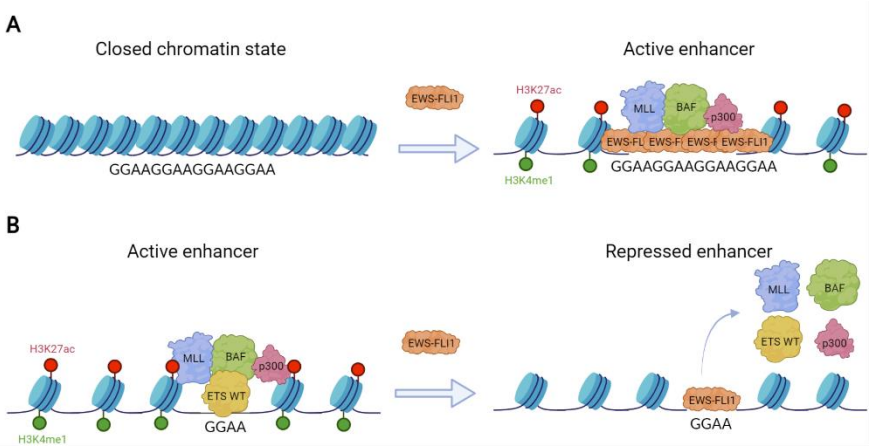


Figure 10. **Mechanisms of chromatin rewiring for gene activation and repression. A)** EWSR1-FLI1 binds to *de novo* enhancers enriched in GGAA repeats microsatellites, and recruits chromatin remodeling complexes and histone modification enzymes to activate enhancer regulated gene expression. **B)** Wild-type ETS transcription factors bind to single GGAA regions and are displaced by EWSR1-FLI1 to inactivate enhancers regulating tumor suppressor and differentiation genes. BAF, BRG1/BRM associated factor; ETS WT, wild-type ETS; MLL, mixed lineage leukemia. Created with Biorender.com.

On the contrary, when monomers of EWSR1-FLI1 bind to single GGAA elements, wild-type ETS transcription factors are replaced by the fusion protein (**Figure 10B**). These regions correspond to conserved enhancers that are transformed from an active to a repressed state due to chromatin compaction, denominated heterochromatin. Active chromatin marks and regulators are also displaced and tumor suppressor and differentiation genes are silenced [29, 31, 85]. However, epigenome mapping of the EWSR1-

FLI1 rewiring showed that there was no widespread localization of the oncogene in repressive chromatin. This suggests that part of the transcriptional repression by EWSR1-FLI1 might be indirect or caused by removing of active histone marks [86].

As mentioned before, SEs are key regulatory regions for the expression of cell identity genes. In this respect it has been demonstrated that the oncogene generates new SEs in genes important for tumorigenesis [74] and for the establishment of a specific transcriptional program [95]. In Ewing sarcoma, SE regions regulate lineage-specific genes that are highly enriched in EWSR1-FLI1 and the H3K27ac histone mark. These regions display repetitive GGAA microsatellites, suggesting cell-type specificity. Among the genes regulated by SEs in Ewing Sarcoma, there are genes that are highly expressed in Ewing sarcoma cell lines, are regulated by EWSR1-FLI1, promote tumorigenesis and favor the maintenance of the cancer cell state. For instance, *CAV1*, *NKX2.2* and *CCND1* are specific SE-regulated genes in Ewing sarcoma [85, 86, 95]. Another example is *MEIS1*, a gene with a pro-survival role by activating the also SE-regulated gene *APCDD1* in association with *EWSR1-FLI1* [96]. Therefore, EWSR1-FLI1 is able to rewire the epigenetic state of Ewing sarcoma cells not only by regulating gene promoters and creating *de novo* enhancers, but also creating *de novo* lineage specific super-enhancers [53, 95].

5.4. DNA methylation

DNA methylation is another epigenetic mechanism required to establish the appropriate chromatin state. In Ewing Sarcoma, *de novo* enhancers

regulated by EWSR1-FLI1 are localized in regions of DNA hypomethylation (**Figure 11**). Thus, DNA methylation distribution anti-correlates with H3K27ac active enhancer marks [29, 86, 97]. This supports the notion that transcription factors can reduce DNA methylation at their binding sites in order to form euchromatin regions [86]. Besides, DNA methylation signature of Ewing Sarcoma tumors is a specific [97].

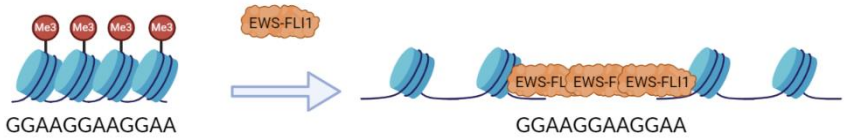


Figure 11. **DNA methylation in Ewing Sarcoma.** EWSR1-FLI1 reduce DNA methylation to allow the formation of active chromatin states. Me3, trimethylation. Created with Biorender.com.

5.5. Non-coding RNAs

A crosstalk between DNA methylation, histone modifications and their modifying enzymes, and ncRNAs is indispensable for an adequate epigenetic regulation. A number of ncRNA have been implicated in Ewing sarcoma tumorigenesis. EWSR1-FLI1 targets the expression or maturation of miRNAs such as miR-30 or miR-145 by direct or indirect mechanisms [29, 98]. Also, the oncogene activates the expression of other miRNAs. Both activation and repression mechanisms contribute to Ewing Sarcoma pathogenesis [98]. One mechanism to avoid maturation of miRNAs is through TAR RNA-binding protein 2 (TARBP2) disruption [29, 99]. By inhibiting miRNA processing, RISC complexes are not able to target and degrade mRNAs from genes associated to the plasticity and self-renewal of normal tissue stem cells such as OCT4 or SOX2. Thus, cell differentiation is impaired and

the expression of pluripotency genes facilitates tumorigenesis [29, 48]. Furthermore, EWSR1-FLI1 can also regulate the expression of lncRNA that promote tumorigenesis. For instance, lncRNA EWSAT1 repress the expression of genes that favors Ewing Sarcoma progression [29, 100].

6. Alternative splicing: a transcription-linked regulatory process of gene expression hijacked by EWSR1-FLI1

6.1. Definition

DNA is transcribed into mRNA forming a precursor sequence (pre-mRNA) that contains exons and introns. Exons are going to be part of the mature RNA, while introns are sections that will be removed from the functional mRNA sequence [101, 102]. Introns forming the pre-mRNAs have short consensus sequences in both 5' and 3' ends, denominated 5' and 3' splice sites (SS) respectively. As pre-mRNA splicing occurs co-transcriptionally, changes in the transcription elongation rate or epigenetic marks can influence exon splicing [103]. When exons are spliced together and introns are eliminated following the same order as they have in a gene, the process is called constitutive splicing. However, when this order is swapped the process is denominated alternative splicing [104]. This results in different forms of mature mRNA proceeding from the same gene. Thus, alternative splicing is a mechanism that expands the complexity of gene expression and regulation in high order eukaryotes [101, 102, 104, 105]. About 90-95% of

human genes undergo alternative splicing giving rise to different mRNA transcript variants. By excluding the variants that undergo nonsense-mediated mRNA decay or by maintaining them in the nucleus, cells can produce distinct protein isoforms that will increase phenotypic diversity. Proteins can change their domain architectures, binding sites, stability, activity or localization depending on the protein isoform generated by alternative splicing [105]. In addition, introducing premature stop codons by alternative splicing can regulate gene expression levels [102].

Alternative splicing must be strictly regulated depending on the cell type and the development stage. Splicing is an important mechanism during development and differentiation that, among other functions, enables the response to environmental changes. Therefore, time and space of splice-site needs to be controlled for a correct functioning of cells [102, 104, 105]. In some cancers, misfunctioning of this mechanism is the basis of the disease[104].

6.2. Alternative splicing type of events

Besides constitutive splicing (**Figure 12A**), mRNA variants can be formed by five main types of alternative splicing events [101, 102, 104-106]. Exon skipping or cassette exons is the most common mechanism in mammalian cells (**Figure 12B**). It consists in deletion of one or more exons. Mutually exclusive exons correspond to two or more splicing events that take place coordinately to exclude a different exon in each event (**Figure 12C**). This is the most unusual type of event. Alternative 5' splice sites, also called alternative donors, are the use of an alternative 5' donor site that changes

the 3' site of the upstream exon (**Figure 12D**). On the contrary, alternative 3' splice sites or alternative acceptors (**Figure 12E**) use the 3' splice site to change the 5' end of the downstream exon. Finally, intron retention consists in maintaining one or more introns in the mature mRNA variant (**Figure 12F**) [106].

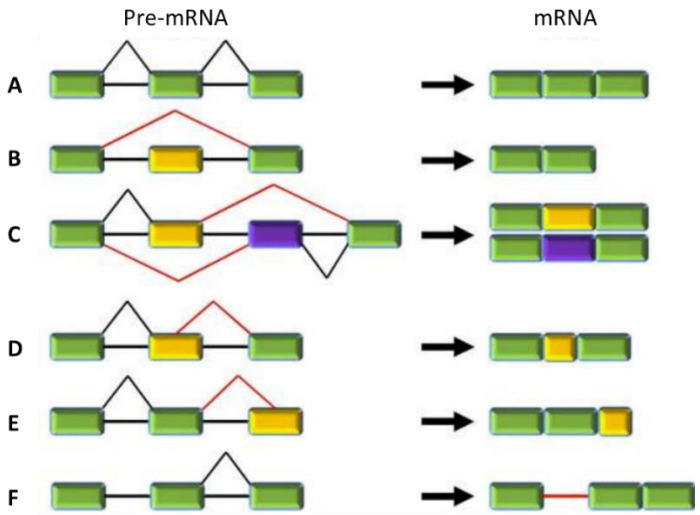


Figure 12. **Classification of splicing events**. Schematic representation of the splicing events that may occur in eukaryotic cells including constitutive splicing (**A**), exon skipping/cassette exons (**B**), mutually exclusive exons (**C**), alternative 5' splice sites /alternative donors (**D**), alternative 3' splice sites/alternative acceptors (**E**), and intron retention (**F**). Adapted from Jiang *et al.,* (2021) [106].

6.3. The spliceosome

Splicing of pre-mRNAs is performed by the spliceosome, a multi-megadalton flexible complex formed by small nuclear ribonucleoproteins (snRNPs). snRNPs contain snRNAs, the RNA-binding proteins involved in biogenesis, function and turnover of target RNAs; Sm proteins, of less than 20 kDa, and like-Sm (LSm) proteins, and several associated proteins that work synergistically [101, 102, 104-108]. Introns forming the pre-mRNAs have, in addition to the 5' and 3' SS, a conserved branch point sequence (BPS)

situated 18 to 40 bp upstream of the 3' splice site (**Figure 13**) [101, 102, 105, 107]. Strong splice sites correspond to the more similar sequences to the consensus. They are easily recognized by the spliceosome and usually promote constitutive splicing. However, when a splice site sequence is different from the consensus, and interactions with the spliceosome are weak, alternative splicing takes place. Apart from *cis*-regulation of splicing depending on the splice site sequences, proteins known as splicing factors regulate splicing in *trans*. Splicing factors recognize specific sequences in exons and introns called splicing enhancers and silencers [101]. Also, in higher eukaryotes a polypyrimidine tract (PPT) after the BPS is necessary for splicing factors recognition of the 3' splice site [107].

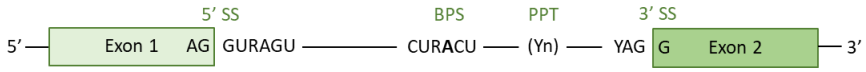


Figure 13. **Characteristics of the splicing sites.** Consensus sequences in both 5' and 3' SS, and in the BPS regulate splicing. GU for 5' SS, AG for 3' SS, and A for BPS are nearly invariant. PPT sequence is also necessary for recognition of splicing factors. BPS, branch point sequence; PPT, polypyrimidine tract. Created with Power Point.

In eukaryotes, there are two types of spliceosomes, the snRNP U2-dependent, and the snRNP U12-dependent. They both function in a similar manner, but the snRNAs are specific to each type. The U2 spliceosome has U1, U2, U4, U5, and U6 snRNAs, while the U12 is composed of U11, U12, U4atac, U5, and U6atac. The U5 snRNA is common to both spliceosomes, as well as to the rest of associated proteins. Remarkably, the U2 spliceosome is in charge of about 96% of splicing in eukaryotic cells [107].

Splicing consists in two consecutive transesterification reactions to remove introns from the pre-mRNA and join exons covalently [101, 102, 107]. As it is detailed in **Figure 14**, the early step of spliceosome assembly corresponds to the recognition of consensus sequences in 5' and 3' intron ends when

U1 snRNP is recruited to the 5’ splice site, SF1 splicing factor bind to the BPS, and U2AF subunits recognize the PPT and 3’ splice site, respectively. Afterwards, U2 snRNP recognizes the BPS and displaces SF1. The tri-snRNP complex U4/U6.U5 is recruited and destabilization of U1 and U4 snRNPs takes place to allow U6 snRNA to bind to 5’ splice site. RNA helicase DHX16 performs the first catalytic step of splicing. This first reaction consists of the cleavage of the 5’ splice site, and 5’ intron end ligates to the branch adenosine of the intron. Subsequently, DHX38 catalyze the second transesterification reaction. In the second reaction, the 5’ and the 3’ ends of exons are ligated after 3’ splice site is cleaved by the 3’ OH group of the 5’ exon. DHX8 releases the mature mRNA (ligated exons) generating the intron lariat spliceosome and DHX15 dissociates the spliceosome complex to enable snRNPs to be used in another splicing process [102, 104, 106, 107].

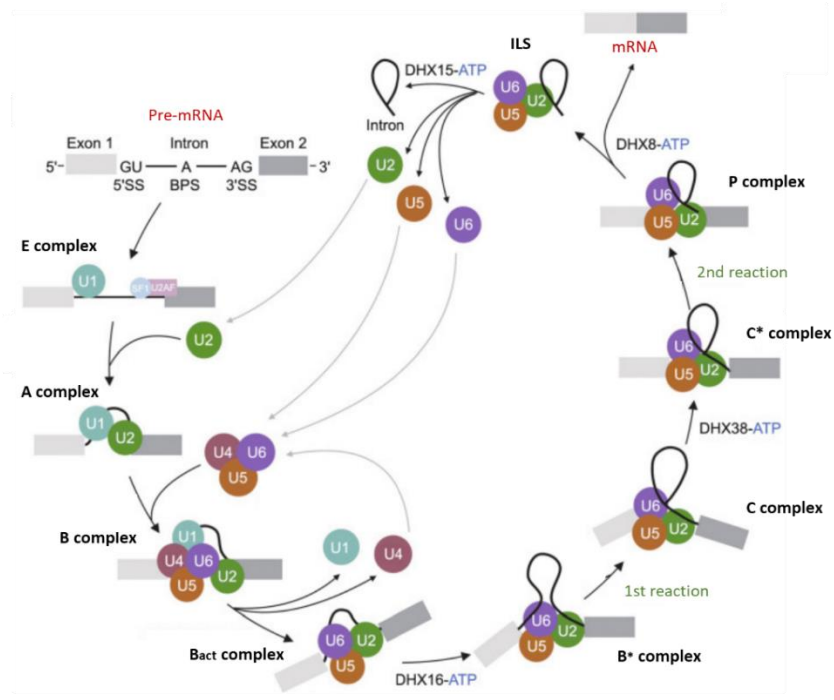


Figure 14. **Representation of the eukaryotic U2-dependent splicing cycle.** Two consecutive transesterification reactions are necessary to remove introns from the pre-mRNA and join

exons covalently. After binding of U1, U2, U4, U5 and U6 snRNPs, and other splicing factors as the non-snRNP SF1 and U2AF in different steps of the process, the mRNA is formed and released. In U12 spliceosome U1 is replaced by U11, U2 by U12 and U4/U6.U5 by U4atac/U6atac.U5. DHX8, DHX15, DHX16 and DHX38 are DEAH-box ATP-dependent RNA helicase. SS, splice site; BPS, branch point sequence; ILS, intron lariat spliceosome. Adapted from Yang *et al.,* (2022) [107].

6.4. Alternative splicing in Ewing Sarcoma

FET family of genes encode RNA-binding proteins that can mediate transcriptional activation by binding to hyperphosphorylated RNA pol II through the prion-like domain of the N-terminal part. In addition, they have a role in RNA splicing due to the recruitment of serine-arginine (SR) splicing factors by the C-terminal domain of FET. As mentioned before, the fusion protein EWSR1-FLI1 contains the N-terminal of EWSR1 and replaces the EWSR1 RNA binding domain by the DBD of FLI1. Therefore, it maintains the capacity to bind to RNA pol II [24, 108]. The N-terminal domain of EWSR1, as well as EWSR1-FLI1, bind to the splicing factor SF1 that recognizes the BPS during splicing (**Figures 13 and 14**) [109, 110]. EWSR1-FLI1 binds with higher affinity than EWSR1 to U1C, one of the three snRNAs that are part of the U1 snRNP forming the E complex in splicing reaction (**Figure 14**) [110]. Furthermore, EWSR1-FLI1, but not EWSR1, is able to alter the 5' splice site selection to form different variants of the adenoviral gene E1A [111]. Also, EWSR1-FLI1 interacts with RNA helicase A (RHA). RHA is a protein implicated in RNA metabolism, whose activity is decreased by the oncogene [112, 113].

Consistently, it has been reported that 43% of EWSR1-FLI1 proteomic interactions correspond to the GO term "RNA splicing and processing" and, by KEGG pathways analysis, 29% of interaction relate to members of the

spliceosome complex [114]. Interestingly, the oncogene induces aberrant alternative splicing of different genes involved in tumorigenesis [114]. For instance, EWSR1-FLI1 regulates the splicing of the SWI/SNF complex protein ARID1A. Since the SWI/SNF complex is required for chromatin relaxation, the oncogene uses a specific isoform of ARID1A that enables Ewing Sarcoma growth and maintenance [29, 94, 115]. Thus, apart from its role as a transcriptional regulator, EWSR1-FLI1 modulates alternative splicing. These alternative splicing events usually are characterized for having enriched GC content, weaker splice sites, and exons with greater distance between 3'SS and the BPS [116].

HYPOTHESIS AND OBJECTIVES

EWSR1-FLI1 acts as a pioneer factor and generates active super-enhancers *de novo*, several of which are located near genes with known roles in cell proliferation, differentiation, and apoptosis. Subsequently, proteins such as the Mediator complex and chromatin remodelers are introduced to enhancers or SEs to activate the aberrant Ewing sarcoma transcriptional program.

Therefore, the hypothesis of this project is that specific subunits of the Mediator complex transcriptionally regulated by EWSR1-FLI1-bound super-enhancers should play a key role in regulating transcription and changes in chromatin structure that enable tumorigenesis in Ewing sarcoma.

To this end, our main objective is to identify genes of the Mediator complex whose expression is regulated by the binding of EWSR1-FLI1 to their super-enhancers and to characterize their functions in Ewing sarcoma.

To achieve the principal objective of the project, the specific aims are the following:

1. To identify Mediator subunits specifically expressed in Ewing Sarcoma.

2. To determine the chromatin regions bound by Mediator in Ewing Sarcoma.

3. To investigate potential interactions of EWSR1-FLI1 with proteins of the Mediator complex.

4. To characterize the functional effects upon Mediator depletion in Ewing Sarcoma.

5. To explore Mediator-related vulnerabilities and their specific inhibitors in Ewing Sarcoma.

MATERIALS AND METHODS

1. Materials

1.1. Plasmids

DNA plasmids used are listed in **Table 2**.

Plasmid	Type	Resistance	Source
EWS-FLI1-FLAG-GFP	Lentiviral	Ampicillin	In-house
PRE	Lentiviral	Ampicillin	Addgene
RSV	Lentiviral	Ampicillin	Addgene
shMED13L	Lentiviral	Ampicillin	In-house
shRERE	Lentiviral	Ampicillin	In-house
SPARQ	Lentiviral	Ampicillin	SBI System Biosciences
Tet-pLKO-puro	Lentiviral	Ampicillin	SBI System Biosciences
VSVG	Lentiviral	Ampicillin	Addgene

Table 2. **DNA plasmids.** List of plasmids used in the experiments indicating the type, the selection antibiotic and the source from which they were obtained.

1.2. Antibodies

Antibodies used for chromatin immunoprecipitation (ChIP), immunofluorescence (IF), immunohistochemistry (IHC), immunoprecipitation (IP) and Western blot (WB) experiments are listed in **Table 3**.

Antibody	Use	Working dilution	Reference	Source
Actin B	WB	1:1000	sc-53015	Santa Cruz Biotechnology
Alexa Fluor 546	IF	1:100	A11035	Invitrogen
Alexa Fluor 647	IF	1:100	A21237	Invitrogen
Anti-Mouse	WB	1:10000/ 1:20000	P0260	Dako
Anti-Mouse HRP	IHC	100 µL	K4001	Dako
Anti-Rabbit	WB	1:10000/ 1:20000	P0448	Dako
Anti-Rabbit HRP	IHC	100 µL	K4003	Dako

Flag	IF, IP, WB	1:20, 3 µL (1 mg/mL), 1:2000	F1804	Sigma-Aldrich
FLI1	IF, IP, WB	1:20, 3 µL (200 µg/mL), 1:1000	sc-356	Santa Cruz Biotechnology
H3K27ac	ChIP	5 µg	ab4729	Abcam
IgG	ChIP, IP	5 µg, 3 µL (5mg/mL)	ab37415	Abcam
Lamin B1	WB	1:1000	sc-374015	Santa Cruz Biotechnology
MED1	ChIP, IF, IHC, IP, WB	5 µg, 1:20, 1:50, 3 µL (1 mg/mL), 1:2000	A300-793A	Bethyl
MED12	ChIP, IF, IHC, IP, WB	5 µg, 1:20, 1:50, 3 µL (1mg/mL), 1:3000	A300-774A	Bethyl
MED13L	ChIP, IF, IHC, IP, WB	5 µg, 1:20, 1:100, 3 µL (1 mg/mL), 1:2000	A300-421A	Bethyl
RBM39	WB	1:1000	sc-376531	Santa Cruz Biotechnology

RERE	ChIP, IF, IHC, IP	5 µg, 1:20, 1:100, 3 µL	HPA024093	Sigma-Aldrich
RERE	ChIP, WB	5 µg, 1:2000	PA5-90323	Invitrogen
RERE	ChIP	5 µg	Orb101371_3	BioOrbyt
RERE	ChIP	5 µg	A15255	ABclonal
STAG2	ChIP, IF, IP, WB	5 µg, 1:20, 3 µL (1 mg/mL), 1:2000	sc-81852	Santa Cruz Biotechnology
U2AF2	WB	1:1000	sc-53942	Santa Cruz Biotechnology

Table 3. **Antibodies.** List of antibodies used in the experiments indicating the technique of utilization, the working dilution and the reference and source from which they were obtained.

2. Tumor samples

Tumor tissues and clinicopathologic data from four patients with Ewing sarcoma, four patients with rhabdomyosarcoma and four patients with neuroblastoma were obtained from the Biobank of Hospital Sant Joan de Déu (Barcelona), integrated in the Spanish Biobank Network of Instituto de Salud Carlos III. Placenta and tonsil tissues used as control samples were

obtained from the Pathological Anatomy Service of Hospital del Mar (Barcelona).

3. Cell lines

Ewing Sarcoma cell lines were obtained from the American Type Culture collection (ATCC) including A673, A4573, SK-ES-1 and TC-71 cell lines. Rhabdomyosarcoma cell lines RH4 and RD were also purchased from ATCC. HEK-293FT, HeLa, Neuroblastoma cell line SK-N-SH and adult cancer cell lines Caco-2, MCF-7 and MDA-MB-231 were obtained from "Banc de línies cel·lulars tumorals" of Hospital del Mar Research Institute (Barcelona). A673 TR shFLI1 cell line was obtained from Dr. Javier Alonso's group [117].

All cell lines were cultured in DMEM medium (Sigma-Aldrich) supplemented with 10% fetal bovine serum (FBS) (Lonza), 10mM of L-glutamine (Sigma-Aldrich) and 1% of penicillin-streptomycin (Sigma-Aldrich). When cell lines were transfected using inducible short hairpin RNA (shRNA) plasmids for MED13L and RERE, DMEM medium was supplemented with 10% FBS without tetracycline (Condalab), 10mM of L-glutamine and 1% of penicillin-streptomycin. To induce the shRNA, knockdown cells were treated with 100 ng/mL for shMED13L, shRERE and shTET-pLKO; and 1 µg/mL for shFLI1 of Doxycycline hyclate (Sigma Aldrich) during 72 hours.

Cells in culture were maintained in a humidified incubator at 37^0C and with an atmosphere of 5% CO_2. Before cells reach confluence or when required for the experiments, cell cultures were passaged using 0.05% Trypsin-

Ethylenediaminetetraacetic acid (EDTA) (Gibco). Cells were frozen in FBS with 10% DMSO (Sigma-Aldrich) and kept at -80ºC or in liquid N_2.

4. Transfection and infection techniques

4.1. Oligofection of small interfering RNA

To obtain cell lines carrying a small interfering RNA (siRNA) in order to express a transient knockdown, 2.5 x 10^5 cells/well were seeded in a 6-well plate. The following day, siRNA oligonucleotides (**Table 4**) were diluted in distilled water following the manufacturer's instructions and medium of each well was changed to 1.5 mL of serum-free OPTIMEM medium (Gibco). 1 µM of siRNA oligonucleotide was mixed with 2 µL of DharmaFECT 4 transfection reagent (Horizon) in a final volume of 200 µL of OPTIMEM medium. Mixture was incubated 15 minutes at room temperature and then transferred to each well. Transfection process was repeated 24 hours later and cells were collected 72 hours after the first transfection to use for different experiments.

Gene	Forward (5' – 3')	Source
GFP	GCUGACCCUGAAGUUCAUCUU	Horizon
MED13L seq#1	CCACCAACAGUUGCAGACUUGCAAA	Invitrogen

MED13L seq#2	GACAGAAACUUUGACAGCUGUUGCA	Invitrogen
MED13L seq#3	GGGCCACCGUGAUGUUGCCUAUAUU	Invitrogen
RERE seq#1	CAGCAGGCCCAACUCGCCAUCUGAA	Invitrogen
RERE seq#2	GCCACCUACCAAGUCAGGAAGAUUA	Invitrogen
RERE seq#3	CCGUCCAGUGAAUUCUUGGACCUAA	Invitrogen

***Table 4.* siRNA oligonucleotides.** List of siRNA oligonucleotides used to induce a transient knockdown including the sequence and the source from which they were obtained.

4.2. Generation of stable cell lines by infection

Ewing sarcoma cell lines stably expressing shRNAs (**Table 2**) were generated by viral infection. Briefly, HEK-293FT cells were grown in a 10-cm plate at a very low passage and, when reached confluence, were seeded again with 1:3 split ratio. Transient transfection was performed the following day using 36 µL of X-tremeGENE HP DNA Transfection Reagent (Roche) diluted in 600 µL serum-free OPTIMEM medium. Transfection reagent dilution was mixed with 50 µL of OPTIMEM medium with 4 µg of the vector and the lentiviral packaging vectors (2 µg of vesicular stomatitis G protein (VSVG), 3 µg of PRE and 3 µg of RSV). After a 30 minutes incubation at room temperature, transfection mixture was added to each cell culture and cells were incubated at 37^0C in a humidified incubator for

48 hours changing the culture medium 24 hours after transfection. Supernatants of lentiviral viruses expressing the plasmid vectors were collected and transferred to infect A673 Ewing sarcoma cells, that had been seeded the previous before with a 1:3 split ratio. When lentiviral viruses were transferred to the appropriate cell line 2 µg/mL of hexadimethrine bromide (Polybrene) (Sigma-Aldrich) was used to enable binding of the virus to the cell membrane. 24 hours after infection cell medium was changed. Since vectors include an antibiotic-resistance gene, 48 hours after infection 2 µg/mL of puromycin (Gibco) was added to the cell culture to positively select the cells. Similarly, generation of FLAG-EWSR1-FLI1-GFP expressing cells were performed by viral infection, and infected cells were sorted in BD Influx (BD Bioscience) cell sorter to select GFP positive cells. Sequences of shRNAs to induce a stable knockdown are described in **Table 5**.

Gene	Vector	Sequence (5' – 3')	Source
MED13L	Tet-pLKO-puro	CCGGGACAGAAATTTTGACAG CTGTTGCACTCGAGTGCAACA GCTGTCAAAATTTCTGTCTTTTTG	In-house
RERE	Tet-pLKO-puro	CCGGCCGCCACCGTTTATGTT CAAACTCGAGTTTGAACATAA ACGGTGGCGGTTTTTG	In-house

Table 5. **shRNA sequences.** List of shRNA sequences used in plasmid vectors including the type of vector, the sequence and the source from which they were obtained.

5. Pharmacological treatments of cell cultures

Treatments *in vitro* to study cell sensitivity were performed in 96-well plates seeding 3000 cells/well. Drugs and concentrations used can be

found in **Table 6**. In the case of Pladienolide B treatments, drug was renewed every 24 hours since it degrades. Cells were treated for 72 hours and cell viability was assessed with luminescence cell viability assay Cell Titer-Glo (Promega). After 40 minutes of incubation at room temperature protected from the light, 15 µL of Cell Titer-Glo/100 µL of medium were added to each well and plates were incubated at room temperature protected from the light five minutes with gently shaking and 25 minutes without shaking. Luminescence signal was read in Orion II luminometer (Berthold) and signal correspond to adenosine triphosphate (ATP) indicating the presence of metabolically active cells. IC50 calculations were performed using Graphpad Prism 8.0 and synergy analysis for combination treatments of Pladienolide B and BRD6989 were done with CompuSyn software [118].

Inhibitor	Concentration used	Vehicle	Source
BRD6989	5 µM	DMSO	MedChemExpress
Pladienolide B	0.0001 – 100 nM	DMSO	From Dr. Juan Valcárcel

Table 6. **Drugs.** List of drugs used *in vitro* including the range of concentrations used, the dilution vehicle and the source from which they were obtained.

6. DNA molecular biology techniques

6.1. Plasmid DNA amplification

One Shot Stbl3 competent bacteria derived from the HB101 *Escherichia coli* strain (Invitrogen) were incubated with DNA plasmids (**Table 2**) for 30 minutes on ice. Bacteria were transformed by heat shock 45 seconds at 42^0C and 1mL of Luria Bertani (LB) medium was added. After 1 hour incubation at 37^0C, transformed bacteria with plasmids were seeded in LB agar plates with 100 µg/mL of ampicillin (Sigma-Aldrich) and incubated overnight at 30^0C or 37^0C, depending on the plasmid, to allow colonies to grow. Colonies were picked and cultured in suspension overnight in 40 mL of LB with 100 µg/mL of ampicillin at 30^0C or 37^0C in a shaking incubator. The following day, plasmid DNA was purified using QIAGEN Plasmid Midi Kit (Qiagen) proceeding with the manufacturer's instructions. Eluted DNA was quantified using the NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific) and stored at -20^0C.

6.2 Cloning Tet-pLKO-puro vector for shRNA expression

Tet-pLKO-puro vector (Addgene) contains the cis-elements necessary for packaging, reverse transcription and integration for the production of lentiviral particles. This vector also includes the components for inducible expression of shRNA when tetracycline or doxycycline is added to culture medium. Tet-pLKO-puro was amplified and isolated to further digest using AgeI and EcoRI restriction enzymes (Thermo Fisher Scientific) following manufacturer's protocols and it was gel-purified in 1% agarose D1 Low EEO

gel (Condalab). Oligos designed with the shRNA sequence included cloning sites for AgeI and EcoRI and were annealed using annealing buffer (Clontech, Takara) in boiling water. Annealed oligos and digested Tet-pLKO were ligated using 10X ligase buffer and T4 DNA ligase (New England Biolabs). Competent bacteria cells from *Escherichia coli* were transformed with the final product of the ligation reaction and bacteria were seed to obtain positive clones. Picked colonies were cultured overnight with 6 mL of LB and 100 µg/mL of ampicillin at 30^0C and DNA was purified in 50 µL using GenElute HP Plasmid Miniprep Kit (Sigma-Aldrich). Eluted DNA was digested using XhoI restriction enzyme (Thermo Fisher Scientific) and gel-purified with GFX PCR DNA and Gel Band Purification Kit (Cytiva). Sanger sequencing of the Tet-pLKO shRNA constructs to verify shRNA sequence was done by Genomics Unit of Universitat Pompeu Fabra (Barcelona). To prepare the sequencing reaction, DNA was amplified using the shRNA sequencing primer (5'-GGCAGGGATATTCACCATTATCGTTTCAGA-3') in a Polymerase Chain Reaction (PCR). PCR mix included 0.75 µL of BigDye (Thermo Fisher Scientific), 1.5 µL of 5X DNA Sequencing Buffer, 0.35 µL of 10 µM sequencing primer, 0.5 µL of DMSO and 250 ng of the template to a final volume of 12 µL. Biometra Tone thermocycler (Analytik Jena) was used 3 min at 94^0C to denature the DNA, followed by 40 cycles including 10 seconds amplification at 96^0C, 5 seconds at 52^0C and extension for 4 minutes at 60^0C. After sequencing validation, shRNA plasmid vectors were amplified and used for stable cell lines generation as described in 4.2 section.

6.3. Chromatin immunoprecipitation

Chromatin immunoprecipitation (ChIP) assays were performed using four 20-cm cell plates per condition with a confluence of 75-100%. In order to detect proteins that might not bound directly to the DNA or have a dynamic interaction, cells were treated with ChIP Cross-link Gold (Diagenode) for 30 minutes at room temperature. Then, cells were fixed adding 1/10 of 1X cross-link solution (0.25 M HEPES pH 8 (Sigma-Aldrich), 0.5 M NaCl (Merck Millipore), 5 mM EDTA (Sigma-Aldrich), 2.5 mM ethylene glycol tetraacetic acid (EGTA) (Sigma-Aldrich) and 1% formaldehyde (Sigma-Aldrich)) for 10 minutes at room temperature. Reaction was stopped with 1/10 stop solution (1gr of glycine (ITW Reagents) in 10 mL of 10 mM Tris-HCl pH 8 (Thermo Fisher Scientific) for 5 minutes. Cells collected with ice-cold 1x PBS with 0.5 mM EDTA and 1X Complete protease inhibitors (Roche) were lysate using 6mL of lysis buffer (10 mM Tris-HCl pH 8, 0.25% Triton X-100 (Merck Millipore), 10 mM EDTA pH 8, 0.5 mM EGTA, 20 mM β-Glycerol-phosphate (Sigma-Aldrich), 0.1 mM sodium orthovanadate (Sigma-Aldrich) and 10 mM sodium butyrate (Sigma-Aldrich)) for 30 minutes at room temperature. Samples were washed with 100 µL of NaCl 5M in 5 mL of sonication buffer (10 mM Tris-HCl pH 8, 100 mM NaCl, 1 mM EDTA pH 8, 0.5 mM EGTA, 10 mM sodium butyrate, 20 mM β-Glycerol-phosphate and 100 µM sodium orthovanadate). After centrifugation, lysates were resuspended in 800 µL of sonication buffer and 10% of sodium dodecyl sulfate (SDS) (Sigma-Aldrich). Sonication was performed in a Bioruptor Pico (Diagenode) in 15 cycles (30 seconds on and 30 seconds off each cycle) to obtain chromatin fragments from 200 to 500 bp. Soluble chromatin was recuperated by centrifugation and 50 µL were taken off to measure by NanoDrop 1000

spectrophotometer the concentration and were used as the input. Using 30 µg of chromatin per immunoprecipitation (IP), chromatin was pre-cleared for a minimum of two hours rotating at 4^0C by adding 1% of bovine serum albumin (BSA) (New England Biolabs), 10mg/mL of salmon sperm DNA (Invitrogen), 4 µl of IgG (Abcam, 5 mg/mL) and 60 µL protein A/G-Sepharose (Invitrogen, 30 mg/mL). Then, chromatin was incubated overnight also rotating at 4^0C with IgG as a negative control and with the corresponding antibodies (**Table 3**). The following day, 60 µl of protein A/G-Sepharose were used to capture the chromatin fragments bound to each antibody. Immunoprecipitated fragments were washed three times using RIPA buffer (0.1% deoxycholic acid sodium salt (DOC) (Sigma-Aldrich), 10 mM Tris-HCl pH 8, 150 mM NaCl, 1% Triton X-100, 0.1% SDS, 1 mM EDTA pH 8, 0.25 mM EGTA, 10 mM sodium butyrate, 20 mM β-Glycerol-phosphate and 0.1 mM sodium orthovanadate), three times with RIPA sodium buffer (11.2 mL of RIPA buffer with 2.8 mL of 5M NaCl), two times with lithium buffer (250 mM LiCl (Sigma-Aldrich), 1% NP-40 (Merck Millipore) , 1% DOC, 10 mM Tris-HCl pH 8, 1 mM EDTA, 1 mM EGTA, 10mM sodium butyrate and 0.1 mM sodium orthovanadate) and two times with Tris-EDTA (TE) at pH 7.5 (10 mM Tris-HCl and 1 mM EDTA). Chromatin-protein complexes were eluted with 100 µl of Elution buffer (1% SDS, 30 mM NaCl, 20 mM β-Glycerol-phosphate and 10 mM sodium butyrate in TE buffer) rotating at room temperature for 30-60 minutes. Cross-link was reversed incubating overnight at 65^0C with 2 µL of RNase A (Thermo Fisher Scientific, 10 mg/mL). Protein was digested using 2.5 µL for immunoprecipitations and 5 µL for input sample with Proteinase K (New England Biolabs) at 55^0C during 2 hours. Finally, chromatin was purified using the MiniElute PCR Purification Kit (Qiagen) and eluted in 50 µL for ChIP-qPCR and 25 µL for sequencing.

6.4. ChIP-Quantitative Polymerase Chain Reaction (ChIP-qPCR)

Eluted chromatin after each corresponding immunoprecipitation and input samples were diluted 1:4 with distilled water. A mix of 0.3 µL of 10 µM forward primer and 0.3 µL of 10 µM reverse primer (**Table 7**) with 2.4 µL of distilled water was distributed in each well of a 384-well plate (Applied Biosystems). Hereinafter, 2 µL of diluted chromatin and 5 µL of SYBR Select Master Mix (Applied Biosystems) were added. qPCR was performed in 7900HT Fast Real-Time PCR (Applied Biosystems) or QuantStudio 5 (Applied Biosystems). Each sample was run in triplicates and data was analyzed calculating the percent of input for each ChIP: %Input = $2^{(-\Delta Ct\ [normalized\ ChIP])}$ x 100 where normalized ChIP is Ct[ChIP] - (Ct[Input] - Log2(Input Dilution Factor)).

Gene	Forward (5' – 3')	Reverse (5' – 3')
ACAT1	TGAGTTTGTGCTGGGTAGGC	GGGCCGGTGGTTTATGTGTA
ACTIN B	AGTGTGTGTCGCAATAGGCA	GCCTTGTACCGTGAACTGGA
CACNB2	TACTCCAGTAAATCAGGAGG	TGATGGAGGTGTTGATTTTC
DANT1	AGATGCTGATCCGCCATGTG	GCCAAGCCAGTTCCTCTTGA
DICER1	TCTTTTCCCCGATCTGTTGC	TTGCTTCAGCCCAGGATTTC
EEF1A1	ACTTTCCCAGTTTACCCCGC	CCACAGTCCCCGAGAAGTTG
LINGO1	GGGGACAGAAAAACCAACGTG	GGCCCAACAACTTTTTGCCA

RERE	CGCATTAAAAGCCGACCTGG	GCACAGGGCTCATGTCTGAT
tRNAs (Chr. 1)	AGCACACCCTCCAAACACC	GACCTTTTGCGGAGGTCCC
U2	CTCGACTGCCAAGCTGAGAT	ATGTGGCACGCTTTCGTTTC
U2SURP	GCCTAGGCTGGAGTTTCCTTT	CAGAAGACCCTCCACTTGGT

Table 7. **Primer sequences for ChIP-qPCR.** List of forward and reverse primers used for ChIP-qPCR. All primers were obtained from Sigma-Aldrich.

6.5. ChIP-sequencing and bioinformatic analysis

ChIP-sequencing (ChIP-seq), as well as library preparation and quality controls, was performed at the Genomics Unit of the Center for Genomic Regulation (CRG, Barcelona). Sequencing was carried out in a HiSeq 2500 platform (Illumina) with 30 – 40 million reads/sample using a 1 x 50 flowcell sequencer. All samples were included in technical triplicates.

Data obtained was analyzed by Dr. Luca Cozzuto from the Bioinformatics Unit at the CRG (Barcelona) and by Dr. Júlia Perera and Ariadna Acedo from the Bioinformatics Units of Hospital del Mar Research Institute (Barcelona). Annotation of the peaks was done using the ChIPseeker package [119]. GENCODE version 39 was used to annotate the peaks using version 3.15.0 from TxDb.Hsapiens.UCSC.hg38.knownGene package. The consensus peak set was obtained by overlapping the three replicates and retaining peaks found in at least two replicates using the function findOverlapsOfPeaks of the ChIPpeakAnno package (version 3.30.1) using different windows in the *maxgap* argument. Consensus Peak Annotation

was performed by annotatePeak. Promoter region was defined from 5kb upstream to +100bp downstream of the transcription start site (TSS). The position and strand information of nearest genes were reported, as well as the distance from the peak to the TSS of its nearest gene and the genomic region. Since some annotation overlapped, ChIPseeker adopted the following priority in genomic annotation: Promoter > 5' UTR > 3' UTR > Exon > Intron > Downstream > Intergenic. Downstream was defined as the downstream of gene end. EWSR1-FLI1 consensus peaks and annotations were analyzed from publicly available raw data of EWS-FLI1 ChIP-seq in A673 cell line[120] found in NCBI Sequence Read Archive (SRA; http://www.ncbi.nlm.nih.gov/sra) under accession number SRA096176.

Annotation of chromatin states was done with BEDtools (version 2.30.0) using the chromatin states from the Epigenome Roadmap [121]. Histone marks of E003 (H1 Cells), E008 (H9 Cells), E006 (H1 Derived Mesenchymal Stem Cells), E026 (Bone Marrow Derived Cultured Mesenchymal Stem Cells) and E025 (Adipose Derived Mesenchymal Stem Cell Cultured Cells) were used to check the overlap with the consensus peaks of each immunoprecipitation. Chromatin states categories were grouped as active TSS (Active TSS and Flanking Active TSS), transcription (transcription at gene 5' and 3', strong transcription, and weak transcription), enhancers (genic enhancers and enhancers), ZNF genes and repeats, heterochromatin, bivalent TSS (bivalent/poised TSS and flanking bivalent TSS/enhancers), bivalent enhancer, and repressed Polycomb (repressed Polycomb and weak repressed Polycomb). Quiescent state was discarded for data representation.

DNA binding motif of each consensus peak set and of the overlapping peaks between different consensus peak sets was analyzed using MEME-

ChIP web service [122]. The University of California Santa Cruz (UCSC) Genome Browser was used to visualize the genomic data obtained [123].

7. RNA molecular biology techniques

7.1. RNA extraction

Cells were pelleted to proceed to RNA extraction using GenElute Mammalian Total RNA Miniprep Kit (Sigma-Aldrich) following the manufacturer's instructions. Lysis mixture was prepared using 10 µL of β-mercaptoethanol per 1 mL of Lysis Solution and 350 µL were added to each sample. In order to remove cellular debris, cell lysates were filtered using a filtration column centrifuging 2 minutes at maximum speed. An equal volume of 70% ethanol was added to the filtered lysates and loaded into the binding columns and centrifuged at maximum speed for 15 seconds. Flow-through was discarded and 500 µL of Wash Solution I was added into the columns followed by a centrifugation 15 seconds at maximum speed. After this first washing step, binding columns was transferred to a new collection tube and 500 µL of Wash Solution II pre-diluted with ethanol was added. Columns were centrifuged at maximum speed for 15 seconds and flow-through was discarded. A second wash with Wash Solution 2 was done with an additional centrifugation step of 2 minutes at maximum speed. Binding columns were transferred to a new collection tube and 50 µL of Elution Solution was loaded into each column and were centrifuged at maximum speed for 1 minute. Eluted RNA was analyzed using NanoDrop

1000 spectrophotometer to check quantity and quality and RNA was stored at -80^{0}C and kept on ice when used for experiments.

7.2. PCR amplification and gel visualization

RNA extracted from cells in culture as described in section 7.1 was retrotranscribed using from 500 to 2000 ng of RNA in a final volume of 12 µL. 0.5 µL of hexamers (Roche, 600 µM) or 0.5 µL of Oligo (dT)s (Roche, 50 µM) were added to each sample. Samples were transferred to a thermocycle 10 minutes at 65^{0}C to denature. Reverse transcription (RT) mix included per sample 4 µL of Buffer 5X (Roche), 2 µL of deoxynucleotide triphosphates (dNTPs) (Roche, 10 mM), 0.5 µL of RNase inhibitor (Roche, 40 units/µL) and 0.5 µL of transcriptase reverse (Roche, 500 units). 7 µL of the RT mix were added to each sample and thermocycle conditions for reaction included 10 minutes at 25^{0}C to anneal primers, 1 hour at 50^{0}C for amplification and 5 minutes at 85^{0}C to inactivate enzymes.

Gene amplification was performed using 0.5-1 µL of the RT product, 0.5 µL of 1 mM dNTPs, 2.5 µL of Buffer 5X (Promega), 1 µL of the corresponding forward primer and 1 µL of the reverse primer (10 µM) (**Table 8**), 0.4 µL of BIOTAQ™ DNA Polymerase (Bioline, 5 units/µL), 0.2 µL of DMSO and 0.75 µL of MgCl 25 mM (Promega). PCR conditions for different genes to validate splicing modifications are detailed in **Table 9**.

Gene	Event	Forward (5' – 3')	Reverse (5' – 3')	Band length (bp)
APAF1	HsaEX0005018	TTGGGATGCGACATCAGCAAA	CCATCAGGAGAAAACATCACACCA	246-375
BIN1	HsaEX0008093	GACAACGCGCCTGCAAAAG	GGAGGCTGCTTCACTTGCC	234-363
BRD8	HsaEX0008428	CGTCAAGCAGTAAAAACACCCC	CGTCAAGCAGTAAAAACACCCC	278-497
CLASP1	HsaEX0015474	TCTTGAAGCTGCTGTTGCTGA	CTTGAGGCATCACTGTTGGCA	102-126
EZH2	HsaEX0023622	GTTGTGGGCTGCACACTGC	TTTTGGAGCCCCGCTGAATAC	218-344
MICAL3	HsaEX0071906	AGCTTCTCCGAGGACTCAGAC	CTGTCTCCGAGCTGCCTTTTG	144-195
SOX6	HsaEX0061075	GTGATAACTACCCCGTACAGTTCA	ACTTGAGTTACAGGGCTGGTC	226-349
U2SURP	HsaEX0061619	TCGCTCCAAAGACAAGAAGGA	AGACCGTTCAGGCACAAAACA	255-432

Table 8. **Primer sequences for PCR.** List of forward and reverse primers used for PCR validation of splicing events. Length for the bands obtained is also included. All primers were obtained from Sigma-Aldrich.

Gene	APAF1	BIN1	BRD8	CLASP1	EZH2	MICAL3	SOX6	U2SURP
Sample volume (µL)	1	1	1	1	0.5	1	1	0.5
Initial denaturation	96^0C 3 minutes							
Number of cycles	35	30	30	30	30	35	35	30
Denaturing stage	96^0C 30 seconds							
Annealing stage	52^0C 30''	52^0C 30''	58^0C 30''	58^0C 30''	58^0C 30''	60^0C 30''	58^0C 30''	58^0C 30''
Extension stage	68^0C 30 seconds							
Final step	68^0C 30 minutes							

Table 9. **PCR conditions.** Conditions for PCR validation of splicing event changes in each gene. Sample volume, number of cycles and annealing temperature are included.

Following amplification, PCR products were loaded into a gel to perform electrophoresis. Gels used were agarose or polyacrylamide (Bio-Rad). Agarose gels were used at 1% of agarose in 1X Tris-Acetate-EDTA (TBE) buffer (Invitrogen) and 4 µL of SYBR Safe DNA Gel Stain (Invitrogen). Electrophoresis was run also in 1X TBE buffer at 100 V for 30 minutes. DNA ladder (Sigma-Aldrich) and samples with 3 µL of 6X Orange G loading buffer (10 mM Tris-HCl pH 7.5, 15% Ficoll (Merck), 0.15% Orange G (Sigma-Aldrich) and 50 mM EDTA) were loaded. In the case of polyacrylamide gels, 8% of polyacrylamide was polymerized using 1X TBE, 0.17% ammonium persulfate (APS) (Sigma-Aldrich) and 10 µL of TEMED (Sigma-Aldrich). Gels were run in 1X TBE buffer 25 minutes at 200 V. Gel was stained using 10 µL

of SYBR Safe DNA Gel Stain in 100 mL of distilled water for 30 minutes protected from the light. Both agarose and polyacrylamide gels visualization were done by UV light exposure in Gel Doc™ XR+ Gel Documentation System (Bio-Rad) using Image Lab™ software (Bio-Rad).

7.3. Quantitative reverse transcription PCR (RT-qPCR)

Purified RNA was used to perform and RT with 1 µg of RNA. Methods used for RT were explained in section 7.2. In addition, GoScript™ Reverse Transcriptase system (Promega) was used in some experiments. RNA in a final volume of 10 µL was denatured at 70^0C for 5 minutes. A mix including 4 µL of nuclease-free water, 4 µL of GoScript™ Reaction Buffer, Random Primer (Promega) and 2 µL of GoScript™ Enzyme mix were added to each sample. Thermocycle conditions were 5 minutes at 25^0C to anneal primers, 1 hour at 42^0C for extension and 15 minutes at 70^0C for enzymes inactivation. cDNA obtained after RT was diluted 1:10 in RNase free water and analyzed by quantitative PCR (qPCR). qPCR was performed in 7900HT Fast Real-Time PCR or in QuantStudio 5 in 384-well plates. In each well, 2 µL of diluted cDNA, 5 µL of SYBR Select Master Mix, 0.3 µL of 10 µM forward primer and 0.3 µL of 10 µM reverse primer (**Table 10**) were mixed in a final volume of 10 µL. Samples were loaded in triplicates and final data was obtained using the fold change method $2^{-\Delta\Delta Ct}$ where ΔΔCt corresponds to ΔCt (treated sample) – ΔCt (untreated sample). ΔCt is the Ct of the gene of interest – Ct of the housekeeping gene.

To analyze expression of the small nuclear RNAs (snRNAs) RNU2-1 by qPCR, RT products were diluted 1:20 and 3 µL/well were used. qPCR mix including

0.44 µL of TaqMan Gene Expression Assay 20X for RNU2-1 (Hs02786874_gH, Applied Biosystems) and 4,4 µL of TaqMan Universal PCR Master Mix (Applied Biosystems) to a final volume of 7 µL were added to each well. Samples were also loaded in triplicates in 7900HT Fast Real-Time PCR or in QuantStudio 5 and data was analyzed using the fold change method.

Gene	Forward (5' – 3')	Reverse (5' – 3')
45sRNA	GAACCTTCCGACCCCTCTC	TACGAGGTCGATTTGGCGAG
5.8sRNA	CTTAGCGGTGGATCACTCGG	AGTGCGTTCGAAGTGTCGAT
ACAT1	TCCAATTGGGATGTCTGG	CTCCTCCATTGCAAATACTG
ACTIN B	GACGACATGGAGAAAATCTG	ATGATCTGGGTCATCTTCTC
BCAT2	GGGAACCATGAACATCTTTG	CATGTCCAGTAGACTCTGTC
BCL2L15	AACTGGTCACCCAACATCCG	CTGGGACTTGCCATTGCTTC
CLK1	AGAGACCATGAAAGCCGGTAT	CATGTGAACGACGATGTGAAGT
CLK4	CGAGAGAAAGCTGGGGACAT	TGCACTGGATTCGATACCCG

CXCL8	GGAAAGTTCCAGGTGTTAGGA	CACCCAGTTTTCCTTGGGGT
CXCR4	AGGTAGCAAAGTGACGCCGA	TAGTCCCCTGAGCCCATTTCC
EWS-FLI1	CCAAGTCAATATAGCCAACAG	GGCCGTTGCTCTGTATTCTTA
GAPDH	ACAGTTGCCATGTAGACC	AAATCCGTTGACTCCGACCTT
GRID2	GAAAGTGTTGGTGCAGCTCG	ACAGCACAAGTCAGTCGAGG
MED12	GGTCCGGCAACTTCAACAAC	CAGAACTCGCACACTCCCTT
MED13L	GTTGATGGCTTTTCAGCGGG	AGTTCAGGTAACCCACAGGC
PRPF39	CAGGATTTTACAGGCTGGGTA	TCCTGGCAGCCATCAAGT
PTBP2	AGTGGGTATGCCTGGAGTCTCA	TCTTCACACGCTGCACATCTCC
RBM10	CTCTACTATGACCCCAACTCCCA	GTCCGCCTCTCCCCATCCCA
RBM39	GTTCGTCGATGTTAGCTCAGTGC	AGCCTCATAGGTCCAGCACTTC
RERE	CTTGCAGATCGGGGGCATAG	TCACAATTAGTGACATGGCTGGA

SF1	CCGCAGCATTACCAACACCACA	CATCCGTGCTTTATCCTGAGCTG
SRSF10	GTCCCACTTGATTTCTACACTCG	TTCTCCTTTCATAACTCCGGCT
TERT	AGAGTGTCTGGAGCAAGTTG	GATGAAGCGGAGTCTGGAC
UPP1	AGTCACAATGATTGCCCCGT	GGCTGGGAAATTGTGTCTGC
U2AF2	TACGGGCTTGTCAAGTCCATCG	CTGGCAGTCAAACACAGAGGTG

Table 10. **Primer sequences for RT-qPCR**. List of forward and reverse primers used for RT-qPCR. All primers were obtained from Sigma-Aldrich.

7.4. RNA-sequencing (RNA-seq) and functional analysis

Biological triplicates including three different siRNA sequences targeting MED13L and three different sequences for RERE (**Table 4**) were used to perform an RNA-seq in order to study the functional effects induced by depletion of each gene. RNA samples were obtained as described in section 7.1 and library preparation using PolyA capture and quality controls were processed by Mar Genomics Unit from Hospital del Mar Research Institute (Barcelona). Samples were sequenced in a HiSeq 2500 platform (Illumina) with a paired-end coverage and ~30 million reads/sample using a 2 x 75 flowcell sequencer at the Genomics Unit of the Center for Genomic Regulation (CRG, Barcelona).

Data obtained was also analyzed by Mar Genomics Unit. Raw library size differences between samples were treated with the weighted trimmed mean method (TMM) [124] implemented in the edgeR package [125]. Raw sequencing reads were mapped with STAR (version 2.7.1a) [126]. Gencode release 36 based on the GRCh38.p13 reference genome was used to annotate transcripts. The table of counts was obtained with FeatureCounts function in the Subread package (version 1.6.4) [127]. The differential gene expression analysis (DEG) between conditions was assessed with voom+limma in the limma package (version 3.48.0) [128]. Genes having less than 10 counts in at least 2 samples were excluded from the analysis. The normalized counts were used in order to make unsupervised analysis, PCA and clusters. For the differential expression (DE) analysis, read counts were converted to log2-counts-per-million (logCPM) and the mean-variance relationship was modelled with precision weights using voom approach in limma package.

Pre-Ranked Gene Set Enrichment Analysis (GSEA) [129] implemented in clusterProfiler package (version 4.0.0) [130] was used in order to retrieve enriched functional pathways. The ranked list of genes was generated using the -log(p.val)*signFC for each gene from the statistics obtained in the DE analysis with limma [128]. Functional annotation was obtained based on the enrichment of gene sets belonging to gene set collections in Molecular Signatures Database (MSigDB). The collections used in this project are c5.bp: Gene sets derived from the Biological Process Gene Ontology (GO) (version 7.2).

Heatmaps and the corresponding Z-scores were calculated with DEG using a threshold of adj.P.Value < 0.05 and |logFC| > 1 on the logCPM normalized counts.

Data analysis performed with R (version 4.1.0) included limma, genefilter, gplots, Vennerable, and hugene20sttranscriptcluster.db packages. Results were validated by RT-qPCR.

7.5. Splicing analysis

RNA-seq with high sequencing depth was used to perform a splicing analysis after MED13L and RERE depletion by siRNA (section 4.1). Biological triplicates with three different siRNA sequences for MED13L and RERE respectively were used. Analysis was carried out by Dr. Sophie Bonnal from Dr. Valcárcel's lab at the Center for Genomic Regulation (CRG, Barcelona). Results were analyzed using the pipeline VAST-TOOLS [131] to calculate the metric percent spliced-in (PSI) and compare the control with the studied condition. This index is the ratio between transcript reads when including or excluding exons [132]. It indicates the splicing efficiency in a specific exon into the transcript population of a gene [133]. A public RNA-seq database of EWS-FLI1 knockdown by endoribonuclease-prepared siRNA (esiRNA) [134] was used to compare the splicing profile with the one of MED13L and RERE knockdowns.

After VAST-TOOLS output, splicing events were classified into microexons and longer alternative exons, which are types of cassette exon splicing; intron retention, alternative 3' splice site (or alt. acceptor) and alternative 5' splice site (alt. donor) events. A filter of delta PSI ≥ 15 was applied for each splicing category. The number of events higher in control, higher in MED13L, RERE, or EWS-FLI1 upon knockdown were obtained. Also, the

number of events assessed in total (total EV), and of events with dPSI≥15 (total AS).

8. Protein molecular biology techniques

8.1. Protein extraction from cell cultures

Protein extraction was performed directly in cell culture plates or with cells already harvested and pelleted. When extraction was in cell culture plates, plates were placed in ice. Cells were rinsed with PBS and 1 mL of RIPA-M lysis buffer (1% NP-40, 0.1 % DOC, 150 mM NaCl, 1 mM EDTA, 50 mM Tris-HCl pH 7.5 and 1X Roche's protease inhibitors) was added to the plate and cells were scraped. For pelleted cells, 1 mL of RIPA-M lysis buffer was added to homogenate cells. After lysis, cells were centrifuged for 15 minutes at maximum speed at 4^0C. Supernatant was collected and kept on ice and pellet was resuspended with 200 µL of UREA-T buffer (50 mM Tris-HCl pH 8.1, 75 mM NaCl, 8M Urea pH 8.66 (Sigma-Aldrich) and 1X Roche's protease inhibitors) by sonication with Digital Sonifier 450 (Branson) during 5 seconds at 30% amplitude. Samples were centrifuged again for 15 minutes at maximum speed at 4^0C and supernatant was collected and mixed with the previous collected supernatant. Protein samples were quantified using the BCA Protein Assay Kit (Thermo Scientific) following the manufacturer's protocol at 562 nm in the Infinite M Nano+ (Tecan) spectrophotometer. Storage of samples was at -80^0C and they were kept on ice whenever used for experiments.

In experiments where nuclear and cytoplasmic fractions were needed to be separated, cells were pelleted and resuspended with 500 µL of fresh Buffer A (0.25 M Sucrose (Sigma-Aldrich), 10 mM HEPES pH 7.5, 3 mM CaCl2 (Sigma-Aldrich), 10 mM NaCl, 1 mM Phenylmethanesulfonyl fluoride (PMSF) (Sigma-Aldrich), 1 mM Dithiothreitol (DTT) (Cytiva), 0.25% NP-40 and 1X Roche's protease inhibitors). Samples were incubated for 30 minutes at 4^0C and centrifuged 10 minutes at 3000 rpm at 4^0C. Supernatant corresponded to the cytoplasmic fraction and was collected. Pellets were rinsed twice with 125 µL of Buffer A and supernatants were collected as cytoplasmic fraction. Nuclear fraction was extracted using UREA-T buffer to break the nuclei (pellets), as described above with 250 µL of UREA-T so the nucleus:cytoplasm proportion was 1:3.

8.2. Western blot (WB)

20 to 50 µg of protein extracts obtained and quantified as described in section 8.1 were diluted with 5 µL of 4X Laemmli buffer (155 mM Tris pH 6.8, 31% glycerol (Fluka), 3.1% SDS, 0.01% bromophenol (Sigma-Aldrich) and 3,5% β-mercaptoethanol (Sigma-Aldrich)) to a final volume of 20 µL of RIPA-M buffer. Diluted samples were boiled at 100^0C for 3 minutes in order to denature proteins and centrifugated 1 minute at 8000 rpm. Depending on the weight of the protein of interest, samples were loaded in a sodium dodecyl sulfate polyacrylamide gel (SDS-PAGE) with 6% or 8% of acrylamide composition in the separating gel (380 mM Tris pH 8.8, 6/8% acrylamide, 0.1% SDS, 0.1% APS and 12 µL TEMED). Stacking gel composition was 125 mM Tris pH 6.8, 5.1% acrylamide, 0.1% SDS, 0.1% APS and 10 µL TEMED. Prestained protein ladder PageRuler™ (Thermo

Fisher Scientific) was run in parallel in the gel as a molecular weight marker. Electrophoresis was run at 70V for the stacking gel and at 100V for the separating gel with running buffer (Tris-Glycine-SDS (TGS) buffer including 25 mM Tris, 192 mM glycine and 1% SDS) in a chamber from Bio-Rad. Proteins were transferred from the SDS-PAGE gel to a PVDF membrane (Merck Millipore) previously activated with methanol (Merck Millipore). Transference was carried out at 250mA with transfer buffer (20% methanol, 25 mM Tris and 192 mM glycine) during 2.5 hours at 4^0C. Membranes were then incubated for 1 hour with 3% BSA (Sigma-Aldrich) in Tris-buffered saline-Tween 20 (TBS-T) (0.1X Tween 20 (Sigma-Aldrich), 20 mM Tris and 150 mM NaCl) at room temperature in order to block unspecific binding. Hybridization with primary antibody (**Table 3**) was done overnight at 4^0C. The following day, membranes were washed 3 times for 5 minutes with TBS-T and incubated 1 hour at room temperature with the indicated secondary antibody (**Table 3**) coupled to horseradish peroxidase (HRP). Membranes were washed every 10 minutes during 2 hours with TBS-T. Working solution of Immobilon ECL Ultra Western HRP Substrate (Merck Millipore) with mixing equal parts of the peroxide solution and the luminol enhancer solution was prepared to cover membranes to react with the HRP enzyme conjugated to the secondary antibody. Membranes were incubated for 1 minute at room temperature, excess liquid was removed and membranes were covered with a plastic protector. In a dark room using a film cassette (Cytiva), X-ray films (Agfa) were placed on top of membranes during the optimal exposure time for each protein. Films were developed with an SRX-101A medical film processor (Konica-Minolta). Relative protein expression was quantified using ImageJ software.

8.3. Protein immunoprecipitation

Cells were cultured in a 20 cm plate and when they reached confluence, they were collected and washed twice with PBS. After centrifugation 5 minutes at 1200 rpm, cells were resuspended with 1 mL of lysis buffer including 0.5% Triton X-100, 1 mM EDTA, 1X Roche's protease inhibitors and 1X Roche's phosphatase inhibitors all in PBS. Cells were lysed for 10 minutes on ice and centrifuged at maximum speed for 15 minutes. Supernatant was collected and pellet was resuspended with 300 µL of lysis buffer. After a second centrifugation, supernatant was collected and added to the first supernatant. Cell lysates had a concentration of 10µg/µL approximately and 50-100 µL were used per IP, corresponding to 0.5-1 mg of protein. 1/10 part of the volume used for IP was used as the input fraction. A step of pre-clearing before immunoprecipitation was done using 0.5% Triton X-100, 1mM EDTA, 1% BSA, 1 µg rabbit IgG, 25 µL of Dynabeads A (Invitrogen) and + 25µL of Dynabeads B (Invitrogen) previously washed with lysis buffer, 1X Roche's protease inhibitors and 1X Roche's phosphatase inhibitors in PBS during 2 hours rotating at 4^0C. For immunoprecipitation, beads were precipitated and supernatant was mixed with 25 µL/25 µL of Dynabeads A/ B, 3 µL of the specific antibody (**Table 2**) and 100 µL of lysis buffer at 4^0C overnight. IP with IgG was used as a negative control of protein binding. The following day, samples were washed three times with lysis buffer and supernatant was discarded. Beads were resuspended in 100 µL of 1X Laemmli Buffer diluted from 4X in lysis buffer and boiled at 100^0C for 3 minutes to denature proteins and unbind the beads. With a magnetic separation rack, beads were discarded and samples were loaded in an 8% SDS-PAGE gel to analyze protein-protein interactions by WB.

9. Microscopic image techniques

9.1. Immunohistochemistry

Immunohistochemistry (IHC) experiments were performed in slides with 2.5 µM sections from paraffin blocks of tumor tissues from Biobank of Hospital Sant Joan de Déu (Barcelona). To deparaffinize samples, they were heated at 60^0C for 45 minutes and rehydrated in an alcohol battery with xylol baths for 10 minutes and ethanol for 5 minutes following standard protocols. Slides were washed with distilled water and antigen retrieval was achieved with Citrate buffer at pH 6 (10 mM trisodium citrate (Merck Millipore)) for 20 minutes in boiling water in a pressure cooker. Slides were washed with distilled water to decrease temperature and washed three times with 1X PBS. Endogenous peroxidase was inhibited with 300 µL/slide of 4% hydrogen peroxide (Sigma-Aldrich) and 1% of sodium azide (Sigma-Aldrich) in 1X PBS for 10 minutes. Slides were washed three times with 1X PBS and tissue was blocked with 1% BSA in 1X PBS for 1 hour. 100 µL of the primary antibody used (**Table 3**) were added in a 1:50 dilution for MED1, 1:50 for MED12 and 1:100 for MED13L with the blocking solution and incubated overnight at 4^0C. The following day, samples were washed three times with 1X PBS and incubated 1 hour at room temperature with 100 µL of the prediluted specific secondary antibody (Dako) (**Table 3**). After three washes with 1X PBS, peroxidase activity was enzymatically detected with DAB (Dako) according to manufacturer's instructions and tissues were incubated until they were properly stained. Sections were washed 3 times and counterstained in filtrated hematoxylin for 15 seconds. Then, they were dehydrated with 2 minutes alcohol baths and 5 minutes xylol baths. Finally, slides were mounted with dibutylphthalate polystyrene xylene

(DPX) (Sigma-Aldrich). Results were observed using Olympus BX61 microscope and IHC images were taken with the digital camera Olympus DP71 using CellSense software (Olympus).

9.2. Immunofluorescence

Cells in culture were seeded directly on a slide inside a culture plate with a concentration of 100.000 cells/mL. The following day, samples were washed three times with 1X PBS containing calcium (Ca^{+2}) and magnesium (Mg^{+2}) and fixed in a freshly prepared 3.7% formaldehyde solution in PBS (+Ca, +Mg) for 10 minutes at room temperature. Slides were washed two times with PBS and cells were permeabilized with 0.2% Triton X-100 in PBS for 5 minutes at room temperature. After two more washing steps, non-specific epitopes were blocked using 1% BSA and 0.03% Tween 20 in 1X PBS in a humid box 30 minutes at room temperature. Primary antibody (**Table 2**) was added 1:20 diluted in blocking buffer and was incubated for 1 hour or longer. Before adding the secondary antibody conjugated to a fluorochrome, samples were washed extensively 5 times with 0.02% Triton X-100 in PBS. Secondary antibody (**Table 2**) was added 1:100 diluted in blocking buffer and incubated in a humid box at room temperature for 1 hour. Slides were washed again 5 times with 0.02% Triton X-100 in PBX, liquid excess was removed and a drop of glycerol with 4',6-diamidino-2-phenylindole (DAPI) (Invitrogen) (40 µL of 1 mg/mL of DAPI/1 mL glycerol) was added to mount slides. Samples were stored at 4^0C protected from the light. Images from four channels were captured with Confocal multiespectral Leica TCS SP8 microscope at the Microscopy Confocal Unit of Hospital Sant Joan de Déu (Barcelona). Super-resolution images were

obtained by image deconvolution using HyVolution tool from Leica Microsystems and were processed and analyzed with LAS X (Leica) and Fiji/ImageJ softwares. Colocalization experiments were quantified by Overlap Coefficient and Förster resonance energy transfer (FRET) [135] using LAS X software (Leica).

10. Statistical analysis

Statistical analyses were conducted using the software Graphpad Prism 8.0. All data is expressed as mean values ± standard deviation (SD). Tests performed in order to compare two groups are specified in each experiment. Statistical significance was determined by a p value of less than 0.05.

RESULTS

1. MED13L and RERE expression in Ewing sarcoma

1.1. Data mining to identify genes regulated by EWSR1-FLI1 binding to super-enhancers in Ewing sarcoma

To identify important genes regulated by super-enhancers that have a role in Ewing Sarcoma tumorigenesis we reviewed data from a public database (Tomazou *et al*. (2015) [86]) including super-enhancer-regulated genes bound by EWSR1-FLI1 in A673 cell line. To determine which of these genes might be expressed in Ewing sarcoma, we performed a hierarchical unsupervised analysis using hESCs and hMSCs jointly as a reference for Ewing Sarcoma tumors transcriptional profiles deposited in GEO. In addition, we extracted transcriptome data from embryonic chondrogenic progenitors expressing EWSR1-FLI1, the murine Ewing sarcoma model that best recapitulates some features of the human disease (Tanaka *et al.* (2014) [49]). The crossing of these three databases identified forty-five common genes (**Figure 15A**). Among them, MED13L, one of the Mediator complex subunits, and the transcriptional regulator RERE were selected since they are potent regulatory proteins of super-enhancers. The enrichment in the histone mark H3K27ac and EWSR1-FLI1 binding characteristic of super-enhancer regions for *MED13L* and *RERE* are shown in **Figure 15B**.

MED13L is the paralogue of MED13, a gene that is part of the kinase module of the Mediator complex. Both proteins are responsible for the reversible binding of the kinase module to the Mediator core, and may

have functions as transcriptional activators or repressors [71, 72, 75-77]. MED13L is reported to be involved in important neural and cardiac developmental pathways [71, 75, 76, 136, 137], and has been shown to be involved in p300 recruitment to enhancers and promoters of different oncogenes in non-small-cell lung cancer [138].

RERE is a nuclear coregulator that can act as a transcriptional activator or repressor depending on the cellular context and associated factors. It is involved in embryonic development and positively regulates retinoic acid signaling and neural differentiation through Notch-mediated Hes expression [139, 140]. Individuals with RERE deficiency have a phenotype that overlaps with that of individuals affected by CHARGE syndrome. These patients have alterations in embryonic development generally due to deficiency in *CHD7* (chromodomain helicase DNA binding protein 7), a chromatin remodeler involved in the formation of DNA loops between regulatory regions and the transcription machinery enriched at super-enhancers [74, 141-143].

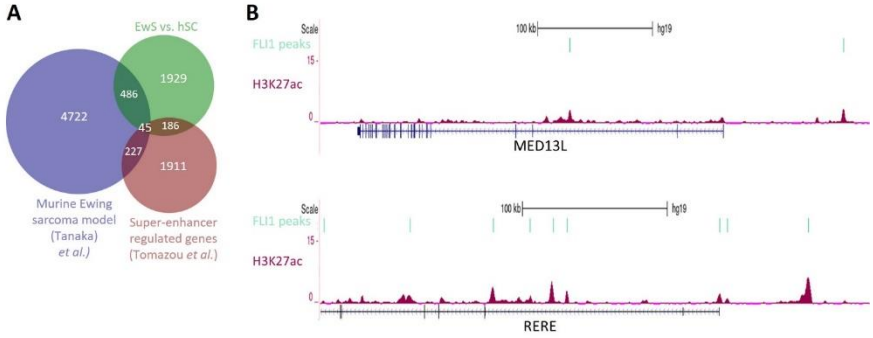


Figure 15. **Identification of genes regulated by EWSR1-FLI1 binding to their super-enhancers in Ewing sarcoma. A)** Venn diagram showing the strategy of *in silico* data mining to identify important genes for Ewing sarcoma tumorigenesis. **B)** Signal tracks of chromatin immunoprecipitation sequencing (ChIP-seq) displaying the enrichment of FLI1 [120], and H3K27ac [86] in *MED13L* and *RERE* regions in the A673 cell line, obtained from the University of California Santa Cruz (UCSC) Genome Browser.

1.2. MED13L and RERE expression in pediatric tumors and cell lines

Super-enhancer regulated genes in Ewing sarcoma tend to be highly expressed [86, 95]. To determine the expression of MED13L and RERE in Ewing sarcoma and other tumor cell lines, we explored the Cancer Dependency Map (DepMap) database (https://depmap.org/portal/). The mean expression of MED13L mRNA in Ewing sarcoma was like the expression levels in other adult and pediatric cancer cell lines (**Figure 16A**). In contrast, RERE expression was significantly higher in Ewing sarcoma cell lines ($P < 0.0001$).

To characterize their expression at the protein level, representative histological sections of Ewing sarcoma and other pediatric tumors were analyzed by immunohistochemistry (IHC) to assess MED13L and RERE expression in the context of different developmental tumors. MED13L expression could be detected in some cells of rhabdomyosarcoma and neuroblastoma specimens, but the intensity was higher in Ewing sarcoma (**Figure 16B**). A similar pattern was observed for RERE, whose expression in Ewing sarcoma was more intense than in rhabdomyosarcoma or neuroblastoma (**Figure 16B**). Analysis of the expression of other Mediator subunits not directly regulated by the oncogene, such as MED1 (from the core complex) and MED12 (from the CDK8 kinase module) showed that MED1 was equally expressed in the three development tumors studied, as well as MED12, with an intense expression in these tumors. Therefore, MED13L and RERE, but not other Mediator subunits, are differentially expressed in Ewing sarcoma.

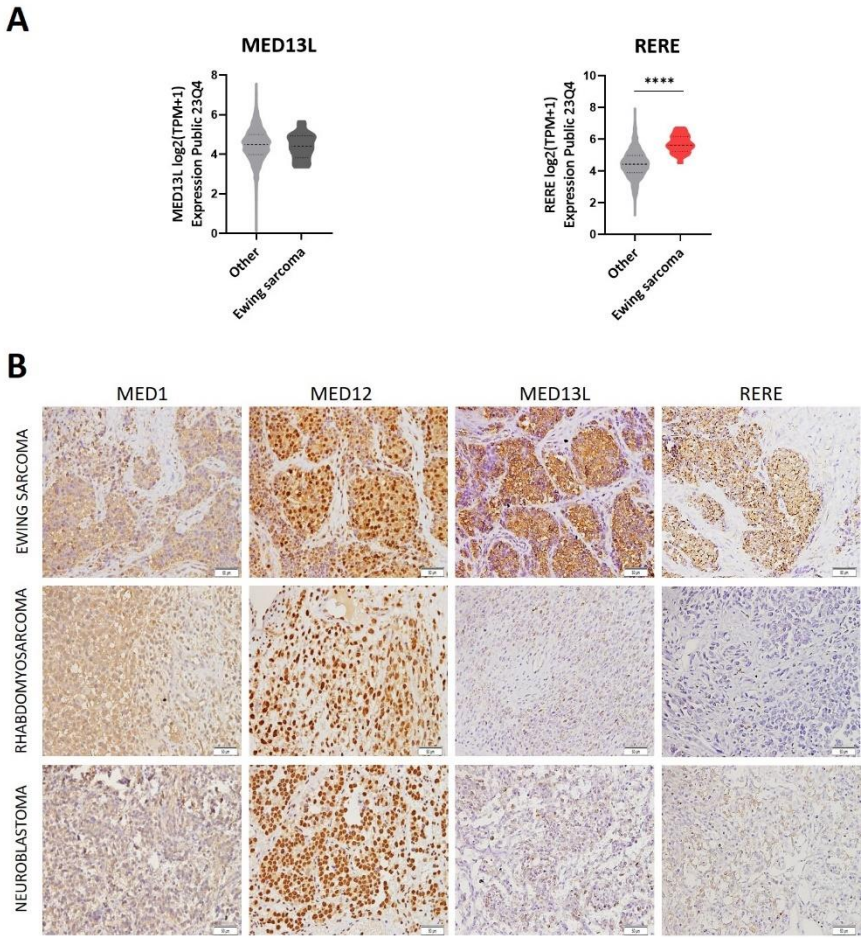


Figure 16. **MED13L and RERE are differentially expressed in Ewing sarcoma. A)** mRNA expression of MED13L and RERE in adult (n=1457) and pediatric (n=22) cancer cell lines, extracted from the Expression Public 23Q4 database from the Cancer Dependency Map (DepMap) database (https://depmap.org/portal/) (https://depmap.org). Unpaired t test with Mann Whitney test was performed to evaluate the differences. ****P <0.0001. **B)** Immunohistochemistry (IHC) assessing MED1, MED12, MED13L and RERE protein expression in representative samples of Ewing sarcoma, rhabdomyosarcoma and neuroblastoma tumors specimens obtained from the Biobank of Hospital Sant Joan de Déu (Barcelona). Scale bar: 50 µm.

1.3. MED13L and RERE expression depends on EWSR1-FLI1 levels

To determine whether EWSR1-FLI1 binding to MED13L and RERE super-enhancers is required for their sustained expression in Ewing sarcoma, we investigated the dependency of their expression to EWSR1-FLI1 levels. To this end, we took advantage of the Ewing sarcoma cell line A673 carrying a doxycycline-inducible short hairpin RNA (shRNA) targeting FLI1. This shFLI1 exclusively knocks down the levels of EWSR1-FLI1 in these cells, since Ewing cell lines do not express wild-type FLI1 [117]. Induction of shFLI1 with doxycycline effectively reduced EWSR1-FLI1, as shown in (**Figure 17A**). Downregulation of the oncogene expression resulted in a decrease of RERE mRNA expression, while MED13L and MED12 mRNAs remained unaffected (**Figure 17A**). In contrast, the reduction of EWSR1-FLI1 was followed by a more than 50% decrease in MED13L and RERE protein levels, whereas MED12 was unaffected (**Figure 17B**). Regulation of MED13L protein levels by EWSR1-FLI1 suggested post-transcriptional mechanisms involved.

These results indicate that MED13L and RERE expression depends on EWSR1-FLI1, consistent with their identification as EWSR1-FLI1 bound super-enhancers [86], but not for the other Mediator subunits like MED12.

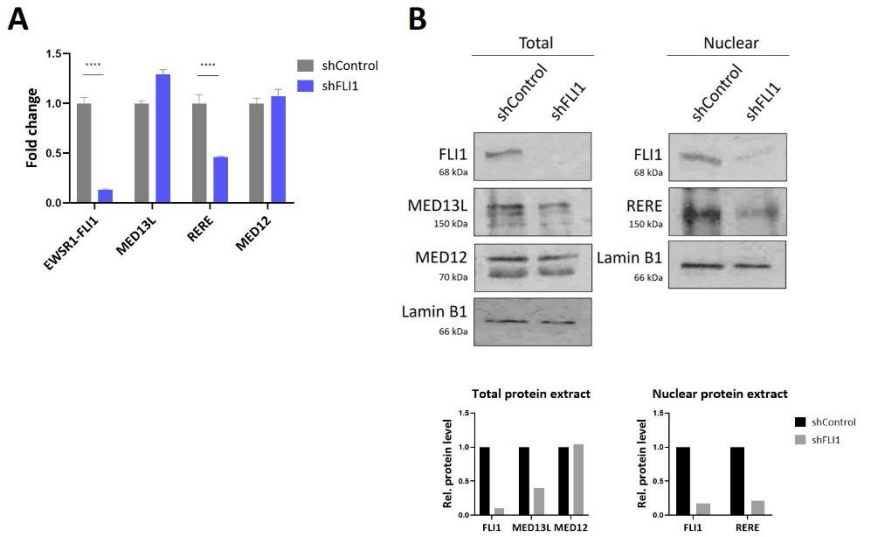


Figure 17. **The expression of MED13L and RERE depends on EWSR1-FLI1. A)** mRNA expression levels for EWSR1-FLI1, MED12, MED13L and RERE in A673 cells expressing the inducible shFLI1 and treated with 1 µg/mL doxycycline, and determined by RT-qPCR. Error bars indicate the standard deviation of three replicates. Two-way ANOVA was used to analyze differences and Sidak's multiple comparisons test was performed to compare expression. ****P value < 0.0001. **B)** Western blot (WB) showing the expression levels of MED12, MED13L and RERE in A673 cells after inducing the expression of the EWSR1-FLI1 knockdown with 1 µg/mL of doxycycline. Lamin B1 was used as a loading control. At the bottom, WB quantification of band intensities relatives to Lamin B1, performed with Image J software.

2. Chromatin occupancy of MED13L and RERE

2.1. Mediator and RERE bind mainly to intergenic regions

The Mediator complex has an important role as a transcriptional regulator by interacting with different chromatin cofactors and forming DNA loops between regulatory regions and promoters of target genes [71, 72, 75]. As for RERE, the similarity of phenotypes due to deficiencies in this protein and CHD7 suggested that they might share chromatin distribution in gene regulatory elements [74, 141, 142]. To explore chromatin regions specifically

bound by MED13L and the putative binding sites of RERE in Ewing sarcoma, we performed chromatin immunoprecipitation experiments followed by sequencing (ChIP-seq) in the A673 cell line. We determined the binding sites of MED13L and RERE, and of the MED1 and MED12 subunits, whose expression is not regulated by the oncogene. In addition, we determined chromatin regions enriched in acetylated histone lysine 27 (H3K27ac), which characterizes enhancers, super-enhancers and promoters, and STAG2 binding, since this cohesin plays a key role in chromatin loop formation in Ewing sarcoma cells [80]. The chromatin sites bound by EWSR1-FLI1 were obtained from Bilke *et al*., (2013) [120], and were used to identify the regions shared by Mediator, RERE and the oncogene in A673 cells.

Three replicates of each chromatin immunoprecipitation condition were sequenced and overlapping peaks in at least two replicates were used for the ChIP-seq analysis. Each peak was annotated to a gene (**Figure 18A**) and genomic distribution was classified as distal intergenic, intergenic, intron, exon, downstream and promoter regions (see Materials and Methods, 6.5 section). H3K27ac and EWSR1-FLI1 signals were enriched at promoter regions (**Figure 18B**). In contrast, peaks of the Mediator subunits, RERE and STAG2 were more commonly found in distal intergenic regions. With respect to binding to gene regions, MED1, MED13L, RERE and STAG2 displayed a preferential enrichment in introns. In contrast, EWSR1-FLI1, H3K27ac and MED12 were equally enriched in intronic and exonic regions. These data indicated that genome-wide distribution of MED13L, RERE and STAG2 differs from EWSR1-FLI1. It is important to note that MED12 displayed a genomic distribution similar to EWSR1-FLI1, suggesting that some Mediator subunits may have specific functions. Indeed, the binding pattern of MED12 –but not that of other Mediator subunits- resembles

that of H3K27ac, a widely distributed signal that decorates active promoters. This distribution of MED12 is consistent with its high expression in all the pediatric tumors studied (**Figure 16B**), and with the role of MED12 in maintaining the active state of specific stem cell enhancers [83].

Genome-wide analysis of DNA binding motifs enriched in chromatin regions recognized by EWSR1-FLI1 [120], MED13L, RERE and STAG2 revealed that MED13L, RERE and STAG2 bind to sequences distinct from EWSR1-FLI1. Whereas the oncogene binding sites are enriched in the GGAA repetitive motifs already described [87-91], the binding peaks of MED13L, RERE and STAG2 were enriched in CA microsatellite regions. These results confirm that MED13L, RERE and STAG2 display a different genomic distribution than the oncogene (**Figure 18C**).

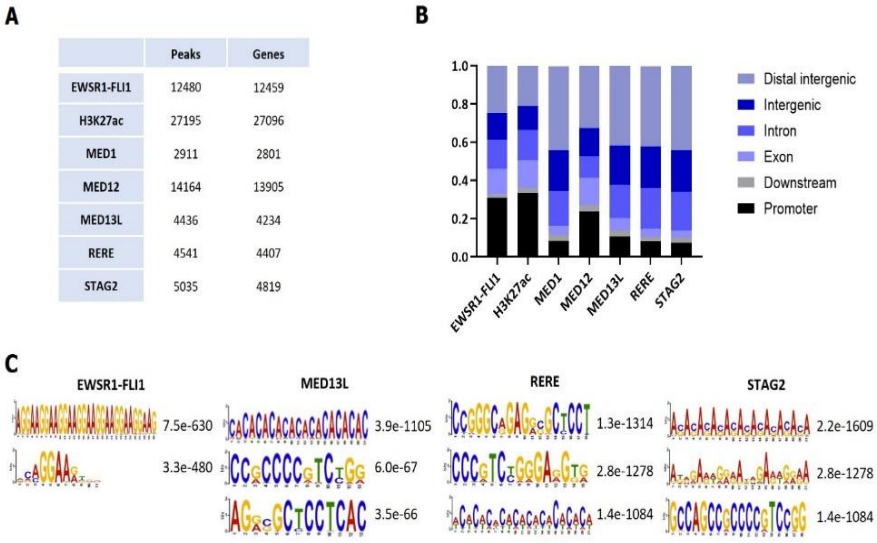


Figure 18. **Chromatin regions bound by Mediator, RERE and STAG2 in Ewing sarcoma. A)** Table displaying the number of peaks and associated genes obtained by ChIP-seq for EWSR1-FLI1, H3K27ac, MED1, MED12, MED13L, RERE and STAG2 in A673 cells. EWSR1-FLI1 peaks and associated genes in A673 cells were extracted from Bilke *et al.*, (2013) [120]. **B)** Bar plots depicting the genomic distribution of EWSR1-FLI1 [120], H3K27ac, MED1, MED12,

MED13L, RERE and STAG2 in A673 cells. **C)** Top MEME DNA binding motif for EWSR1-FLI1 [120], MED13L, RERE and STAG2 peaks and their corresponding E value.

In order to elucidate the transcriptional states associated to chromatin regions enriched by EWSR1-FLI1 [120], H3K27ac, MED1, MED12, MED13L, RERE and STAG2, we annotated the chromatin state of each peak in the genome of H1 and H9 human embryonic stem cells (hESC), and H1- and bone marrow (BM)-derived mesenchymal stem cells (MSC) [121] (see Materials and Methods, 6.5 section). Similar categories were grouped to represent the results obtained, considering that the quiescent state, which corresponds to low levels of all histone marks, was excluded from the representation for the sake of clarity. This analysis showed that EWSR1-FLI1 and H3K27ac binding was enriched in chromatin states corresponding to active transcription start sites (TSSs), and to a lesser extent to transcriptionally active regions and enhancers in the four stem cell lines (**Figure 19**). In contrast, the binding sites of the MED13L, MED1, RERE and STAG2 in hESCs mainly corresponded to transcription-associated chromatin, followed by heterochromatin regions and zinc finger protein (ZNF)-regulated genes that changed to states associated with Polycomb-repressed regions in more differentiated cells such as H1- or BM-derived MSCs. Thus, the chromatin status associated with MED13L and RERE binding diverges from that of EWSR1-FLI1, further confirming their different genomic distribution.

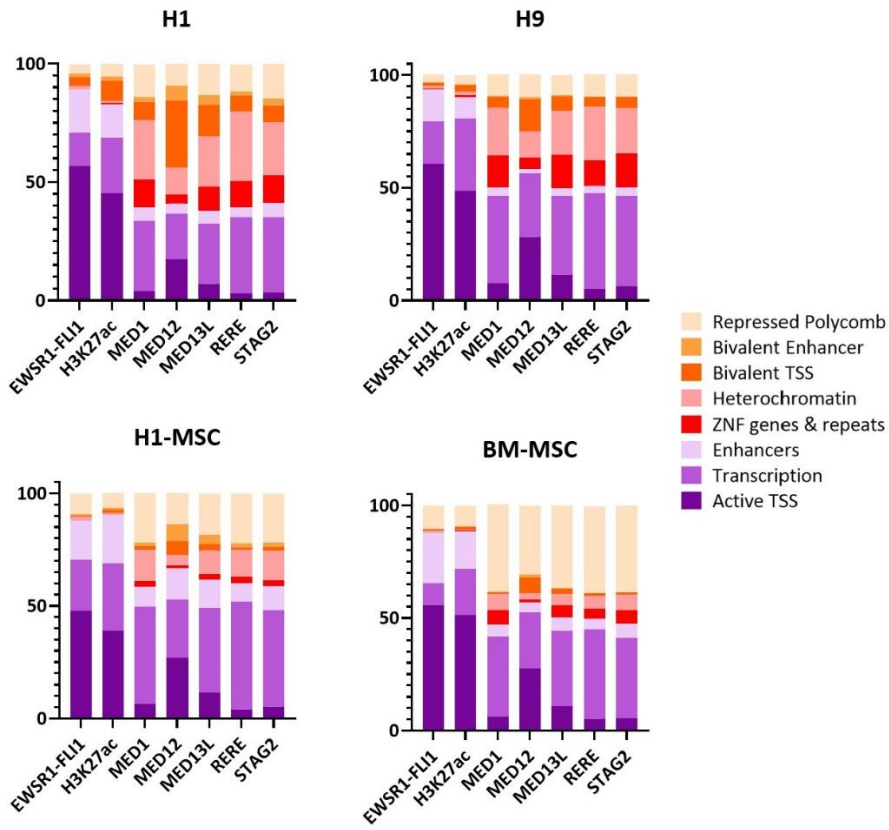


Figure 19. **Chromatin states associated to the binding sites of EWSR1-FLI1, H3K27ac, Mediator, RERE and STAG2.** Bar plots showing the classification of the chromatin states annotated in the Epigenome Roadmap [121]. Cell lines used were ESC lines H1 and H9, H1-MSC, and BM-derived MSC. BM, bone marrow; ESC, embryonic stem cell; MSC, mesenchymal stem cell; TSS, transcription start site; ZNF, zinc finger.

2.2. MED13L and RERE share binding in a high percentage of chromatin regions

As mentioned above, RERE was initially identified as a nuclear receptor that functions as a transcription factor to modulate the expression of its target genes [139, 140], but to date a relationship with Mediator has not been described. Since our ChIP-seq analysis revealed a striking similarity in the genomic distribution of RERE, Mediator subunits and STAG2, we

intersected the binding peaks of each protein to elucidate whether they bind to the same chromatin regions. Eighty per cent of the RERE peaks (corresponding to 3876 peaks) overlapped with peaks bound by at least one of the three Mediator subunits analyzed (MED1, MED12 and MED13L), and 44% of the RERE peaks (1948 peaks) overlapped with all three units simultaneously (**Figure 20A**). These data indicate that RERE mostly binds to the same chromatin regions as the Mediator complex, and therefore, may cooperate with its function. On the other hand, 78% of the STAG2 bound regions (3780 peaks) were also enriched in MED13L or RERE (**Figure 20B**). Two representative examples of these genomic regions are *TAF1* and *LINGO1*, where Mediator, RERE and STAG2 are found in the enhancer and promoter regions, respectively (**Figure 20C-D**). Since the conformational change induced by the Mediator complex that allows the formation of DNA loops requires interaction with STAG2, these data suggest that MED13L and RERE could also be part of the complex protein-DNA loop.

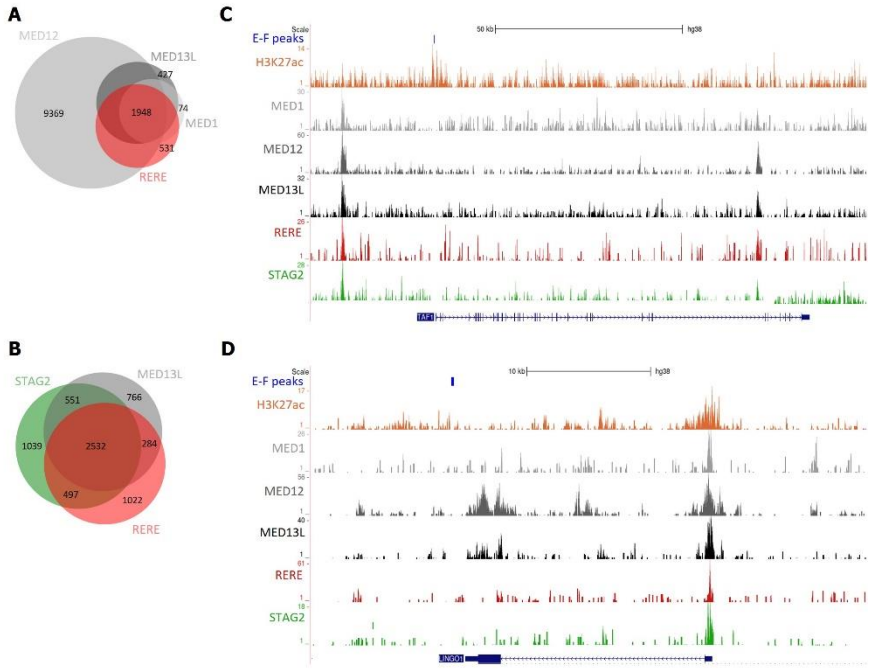


Figure 20. **Mediator, RERE and STAG2 bind to overlapping chromatin regions. A)** Venn diagram showing the intersection of ChIP-seq peaks for MED1, MED12, MED13L, and RERE in A673 cells. **B)** Intersection of MED13L, RERE and STAG2 ChIP-seq peaks in A673 cells depicted in a Venn diagram. **C-D)** ChIP-seq signal tracks for FLI1 [120], H3K27ac, MED1, MED13L, RERE and STAG2 in *TAF1* and *LINGO1* loci, generated with the University of California Santa Cruz (UCSC) genome browser. E-F, EWSR1-FLI1.

2.3. MED13L and RERE binding to chromatin is specific to Ewing sarcoma

Next, we investigated whether MED13L and RERE recognize the same targets in other cell lines or, in contrast, are specific of Ewing sarcoma cells. We chose CACNB2, DANT1, EEF1A1, tRNAs, U2 and RERE among the genes that have peaks of MED13L and RERE. Although having MED13L and RERE peaks, U2SURP could only be validated for RERE binding. LINGO1 gene only present binding peaks of MED13L. ChIP-qPCR in the Ewing sarcoma cell lines A673 and SKES1 and rhabdomyosarcoma cell lines RH4 and RD

showed enrichment of MED13L (**Figure 21A**) and RERE (**Figure 21B**) in these targets in Ewing sarcoma, but not in rhabdomyosarcoma cell lines, indicating that chromatin binding sites of MED13L and RERE are specific of Ewing sarcoma.

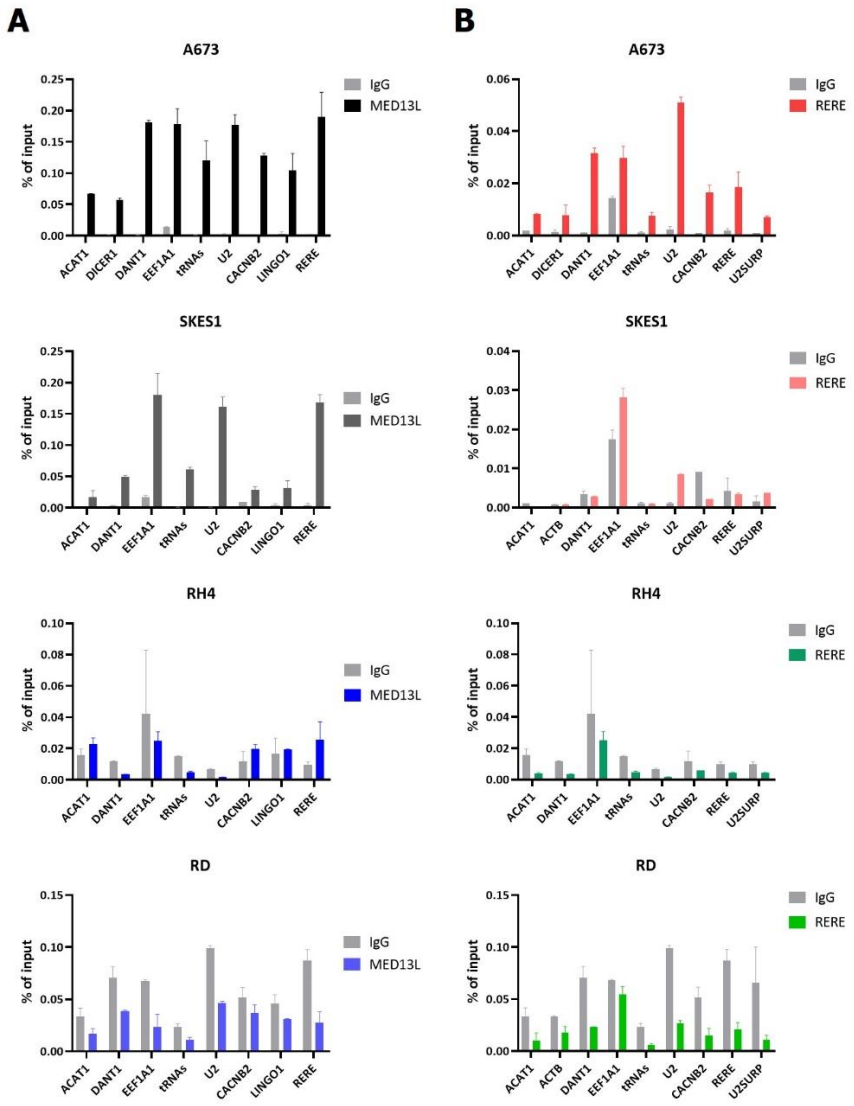


Figure 21. **MED13L and RERE binding is specific to Ewing sarcoma.** ChIP-qPCR experiments of MED13L **(A)** and RERE **(B)** targets in the Ewing sarcoma cell lines A673 and SKES1 and in the rhabdomyosarcoma cell lines RH4 and RD. Data represent the enrichment ratio of

immunoprecipitated samples relative to input. IgG was used as a negative control for each immunoprecipitation and ACAT1, ACTIN B and DICER1 were used as control regions.

3. Characterization of chromatin regions enriched in MED13L, RERE and EWSR1-FLI1

3.1. Mediator, RERE and the oncogene share their binding to a subset of genes

The above results indicated that chromatin regions enriched in MED13L and RERE largely differed from EWSR1-FLI1 bound regions. To determine whether, albeit in a minority, there were chromatin sites where the oncogene could coincide with MED13L and RERE recruitment, we crossed MED13L and RERE peaks with the EWSR1-FLI1 peaks [120]. This analysis identified only 80 and 51 common peaks with MED13L and RERE, respectively (**Figure 22A**). Examples of those few regions with enrichment in MED13L, RERE and the oncogene include *CASC3* and *ZMIZ1*, as shown in **Figure 22B**. This poor overlap is consistent with the divergent genomic distribution of the peaks and the chromatin states associated (**Figure 18-19**), where MED13L and RERE bind to different chromatin regions compared to EWSR1-FLI1. For STAG2, the overlapping regions with EWSR1-FLI1 were also very limited (**Figure 22A**).

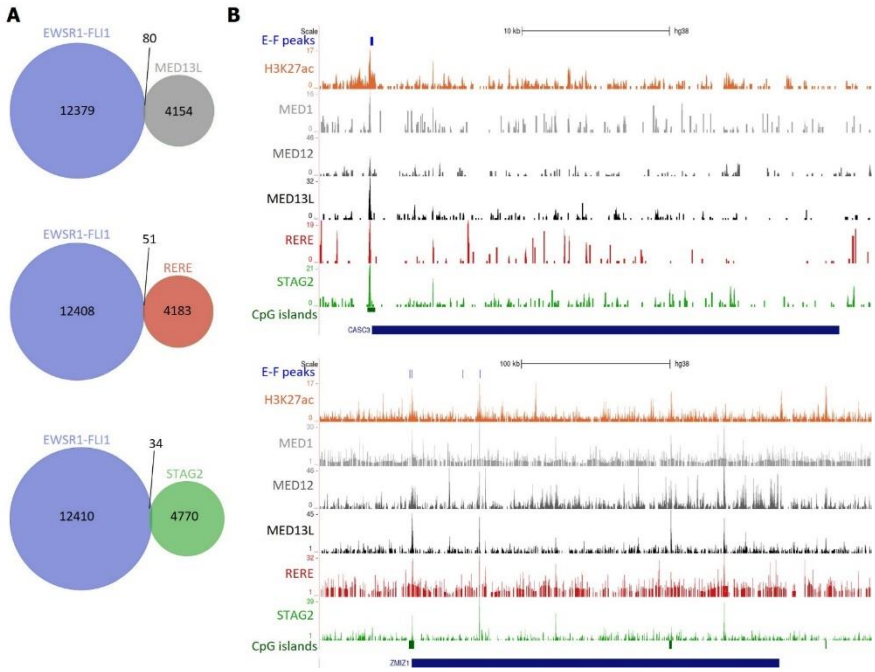


Figure 22. MED13L, RERE and STAG2 overlap in discrete regions with EWSR1-FLI1. **A)** Venn diagram depicting the overlap between ChIP-seq peaks for MED13L, RERE or STAG2 and EWSR1-FLI1 in A673 cells. **B)** ChIP-seq signal tracks for FLI1 [120], H3K27ac, MED1, MED12, MED13L, RERE and STAG2 in CASC3 and ZMIZ1 genic regions, generated with the UCSC genome browser. E-F, EWSR1-FLI1.

Genome browser visualization of the EWSR1-FLI1 binding peaks and the chromatin regions enriched in Mediator, RERE and STAG2 showed that although most of the oncogene peaks did not coincide with the other proteins, they were often located in the same genes. Crossing the genes associated with MED13L, RERE, or STAG2 peaks and genes associated to EWSR1-FLI1 peaks revealed 34% of MED13L-bound genes (1108 genes), 34% of RERE-bound genes (1160), and 35% of STAG2-bound genes (1265) were coincident with genes associated to EWSR1-FLI1 peaks (**Figure 23A**). This result suggests that, in co-occupied genes, Mediator and RERE could be contributing to the formation of DNA loops or stabilization between the EWSR1-FLI1-binding regulatory regions and the promoter of target genes, potentially mediated by STAG2. An example of co-occupied genes by

Mediator, RERE, STAG2 and EWSR1-FLI1 is *JARID2*, as shown in Figure 23B. In this genomic region, EWSR1-FLI1 binding corresponds to a distal region enriched in H3K27ac, characteristic of (super)enhancer regions, whereas Mediator, RERE and STAG2 are bound closer to the gene promoter.

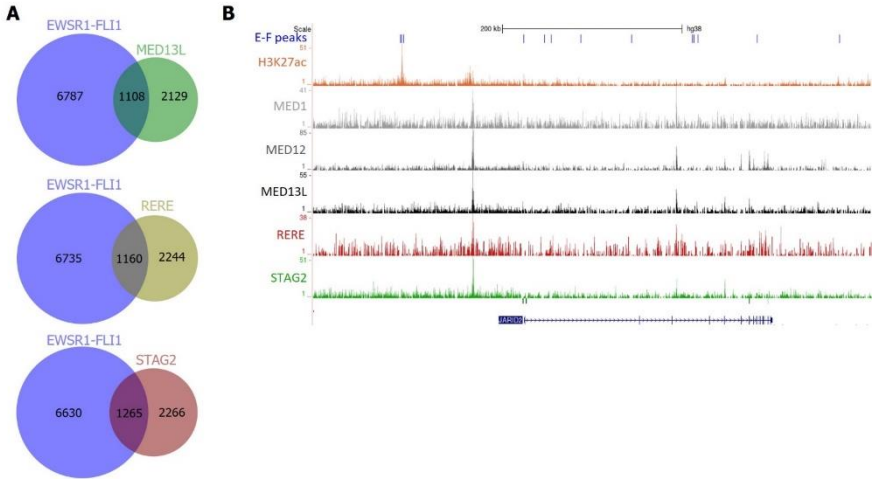


Figure 23. **MED13L, RERE and STAG2 binding to EWSR1-FLI1-associated genes. A)** Intersection between genes associated to ChIP-seq peaks for MED13L, RERE, and STAG2 with genes associated to EWSR1-FLI1 binding peaks [120] in A673 cells. **B)** ChIP-seq signal tracks for FLI1 [120], H3K27ac, MED1, MED13L, RERE and STAG2 in *JARID2*, from the UCSC genome browser. E-F, EWSR1-FLI1.

3.2. Chromatin regions bound by Mediator, RERE, STAG2 and EWSR1-FLI1 are characterized by an enrichment of tandem repeats

Next, we wished to investigate the few discrete regions where MED13L and RERE overlap with EWSR1-FLI1 binding. These genomic regions were mainly characterized by tandem peaks of Mediator, RERE, STAG2 and the oncogene and by being non-coding regions regulated by CpG islands (**Figure 24A**). Examples of these regions include the lncRNA DANT1, that is formed within the DXZ4 microsatellite repeat region of chromosome X; the

genomic regions coding for the ribosomal subunits RNA5S and RNA5.8S; and the tRNA loci. In addition to these regions, Mediator, RERE, STAG2 and EWSR1-FLI1 were coincident in their binding to the *EEF1A1* protein-coding gene. While ribosomal RNAs and tRNAs are directly involved in ribosome function, *EEF1A1* encodes a protein responsible for the enzymatic delivery of aminoacyl tRNAs to the ribosome. Furthermore, Gene ontology (GO) analysis of the genes associated with these peaks revealed an enrichment in the category "Ribosome" for cellular component and "Structural constituent of ribosome" for molecular function (**Figure 24B**), supporting the role of Mediator, RERE and EWSR1-FLI1 in the co-regulation of translational rewiring in Ewing sarcoma.

The consensus DNA-binding motif of MED13L is the IRF7 motif (e values: 1.7e-006) and the GGAA single FLI1 motif (e value = 2.3e-002), while the consensus DNA binding motifs of RERE are the SP2 and SP3 binding motifs (e value = 2.4e-055) (**Figure 24C**). These motifs are separated from the GGAA microsatellite repeats specifically bound by EWSR1-FLI1 in *cis*-regulatory regions [87-91].

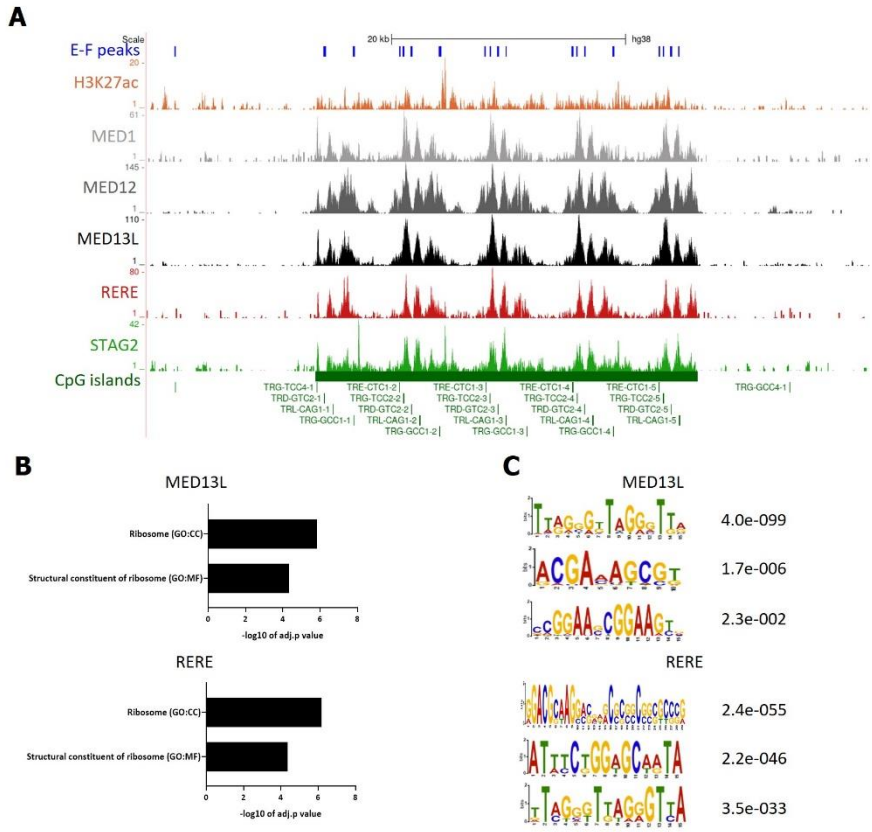


Figure 24. **Mediator, RERE, STAG2 and EWSR1-FLI1 form tandems of enriched peaks in CpG islands that correspond to non-coding regions. A**) ChIP-seq signal tracks from UCSC Genome Browser for FLI1 [120], MED1, MED12, MED13L, RERE and STAG2 in *tRNA*s regions (chromosome 1). **B)** Bar plots displaying statistically significant categories from Gene Ontology (GO) analysis of the genes bound by MED13L or RERE and EWSR1-FLI1 in A673 cells with adjusted P value <0.05. Analysis was performed with gProfiler online database [144] (https://biit.cs.ut.ee/gprofiler/gost). CC, cellular component; GO, Gene ontology; MF, molecular function. **C)** Top MEME DNA binding motif for overlapping peaks in the same genomic region of EWSR1-FLI1 with MED13L or RERE and their corresponding E value. E-F, EWSR1-FLI1.

3.3. MED13L, RERE and EWSR1-FLI1 regulate ribosomal subunit expression in Ewing sarcoma cell lines

Ribosomal subunits 45s and 5.8s are one of the singular regions where Mediator, RERE, STAG2 and EWSR1-FLI1 colocalize in chromatin. To explore

whether MED13L and RERE binding to these regions influences their expression, we used two different approaches. The first consisted in depleting MED13L and RERE by transient short interfering RNA (siRNA) in the A673 cell line, using three different siRNA sequences for each protein. Validation of the siRNA efficacy was determined by RT-qPCR and WB (**Figure 25A-B**). The second approach was the expression of inducible shRNAs in A673 cells. For this, A673 cells were infected with a lentivirus carrying doxycycline-inducible shRNAs and a puromycin selection cassette, and culture in the presence of puromycin to generate stable cell lines. Expression of the shRNAs was induced using 100 ng/mL of doxycycline after culturing the cells for 72 hours, and knock-down efficacy was also validated by RT-qPCR and WB (**Figure 25C-D**).

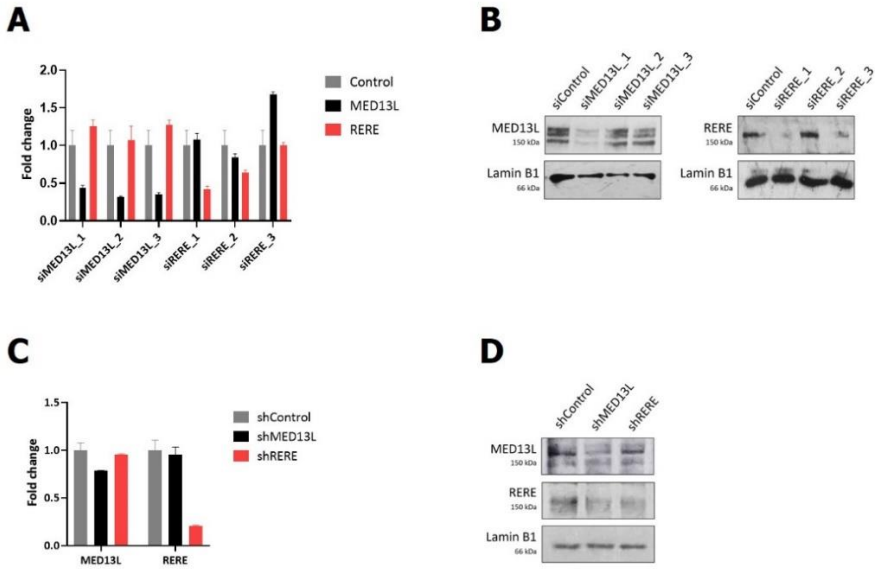


Figure 25. **Knockdown of MED3L and RERE by small interfering RNA (siRNA) and by inducible short harpin RNA (shRNA) in the A673 cell line. A)** Validation by RT-qPCR of MED13L and RERE depletion by siRNA using three different sequences for each target. Values were normalized to Actin B. **B)** WB validation of MED13L and RERE depletion by siRNA. Lamin B1 was used as a loading control. **C)** RT-qPCR validation of MED13L and RERE knockdown by inducible shRNA using 100 ng/µL of doxycycline. Values were normalized to

Actin B. **D)** WB showing MED13L and RERE downregulation by inducible shRNA using 100 ng/µL of doxycycline. Lamin B1 was used as a loading control.

Once the knockdown approaches were validated, we determined the ribosomal RNA levels in A673 cells by inducing depletion of MED13L, RERE and EWSR1-FLI1, and in cells transiently oligofected with MED13L and RERE siRNA sequence #1. Downregulation of FLI1, MED13L and RERE significantly reduced 45S RNA and 5.8S RNA levels in A673 cells (**Figure 26A**). In contrast, oligofection of the rhabdomyosarcoma cell line RH4 with the same siRNAs for MED13L and RERE did not result in reduced expression of these ribosomal subunits (**Figure 26B**). These results suggest that MED13L, RERE and EWSR1-FLI1 regulate the expression of these ribosome subunits in Ewing sarcoma.

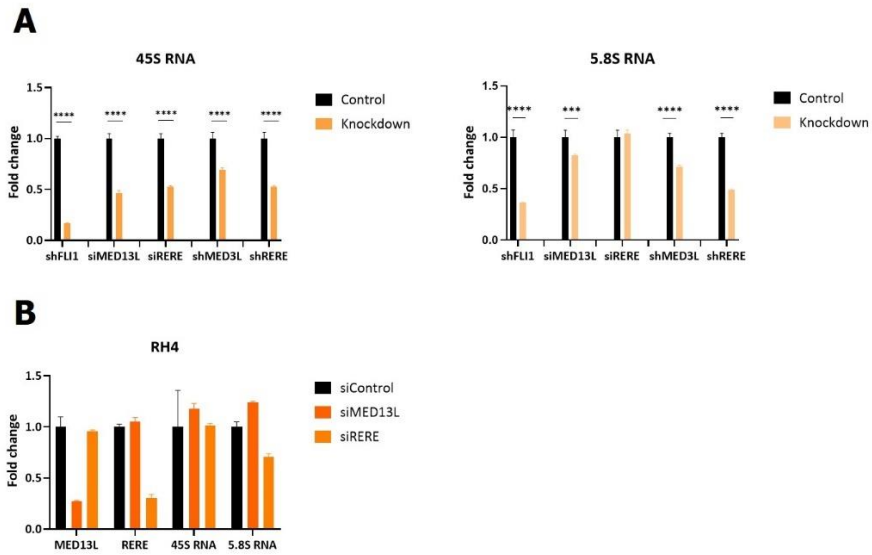


Figure 26. **The expression of ribosomal RNAs depends on MED13L and RERE. A)** 45s and 5.8 ribosomal RNA expression in A673 cells depleted from FLI1, MED13L and RERE. Values were normalized to Actin B. **B)** RT-qPCR of 45s and 5.8s ribosomal RNA in the rhabdomyosarcoma cell line RH4 depleted from MED13L and RERE by siRNA. Values were normalized to Actin B. Two-way ANOVA was used to analyze differences. Sidak's multiple comparisons test was performed to compare expression after FLI1, MED13L and RERE knockdowns. ***P value 0.0005, ****P value < 0.0001.

3.4. EWSR1-FLI1 binding to U2 chromatin regions but not MED13L and RERE result in transcriptional changes

Other chromatin regions that display enrichment of Mediator, RERE, STAG2 and EWSR1-FLI1 in tandem peaks of CpG islands is the U2 snRNA locus (**Figure 27A**). U2 is an essential component of the splicing machinery that is organized as a nearly perfect tandem array containing 5 to 22 copies of a 5.8 kb repeated unit [145]. As with ribosomal RNAs, we investigated U2 expression after MED13L and RERE depletion in A673 cell line by inducing the shRNAs with 100 ng/mL of doxycycline, and by siRNAs (**Figure 27B**). We did the same for FLI1 knockdown by inducible shRNA with 1 µg/mL of doxycycline in A673 cell line. Although we observed that U2 expression tends to decrease upon MED13L and RERE depletions, downregulation only reached statistical significance after FLI1 knockdown suggesting that the oncogene could be regulating U2 transcription.

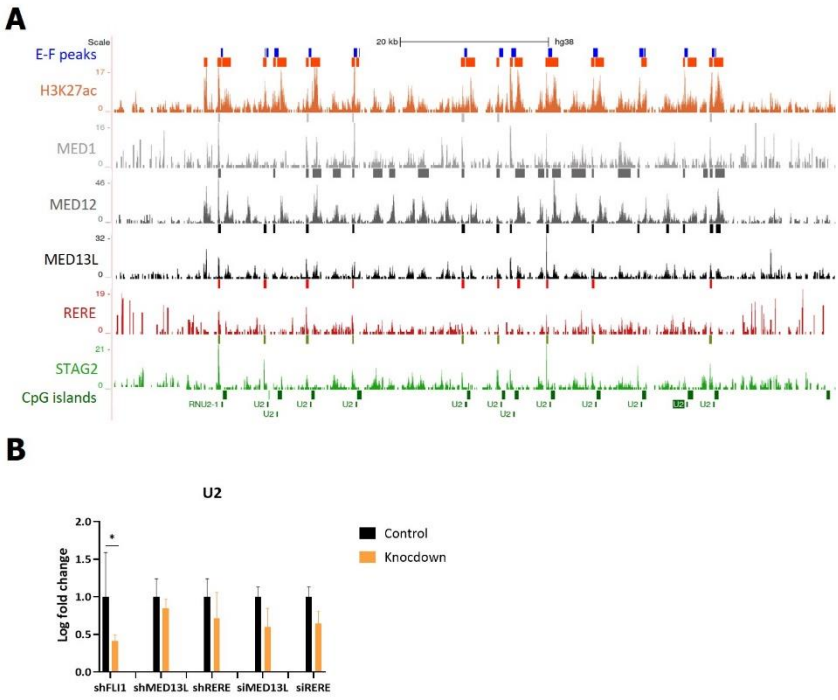


Figure 27. **MED13L and RERE bind to and regulate expression of U2 spliceosome factor.** **A)** UCSC Genome Browser ChIP-seq signal tracks for EWSR1 [120], H3K27ac, MED1, MED12, MED13L, RERE and STAG2 at the snRNA *U2* genomic region in chromosome 17 in A673 cell line. **B)** RT-qPCR determination of snRNA U2 expression upon FLI1, MED13L and RERE knockdowns in A673 cell line when compared to their control. Values are normalized to Actin B. Two-way ANOVA and Sidak's multiple comparisons test were used to analyze expression differences. *P value < 0.0314.E-F, EWSR1-FLI1.

3.5. EWSR1-FLI1 is required for MED13L and RERE binding to shared chromatin regions

Our *in-silico* screening identified MED13L and RERE as super-enhancer regulated genes bound by EWSR1-FLI1, and we showed that their expression is dependent on oncogene levels. These data suggested that EWSR1-FLI1 ensures that the Ewing sarcoma cell produces enough of these proteins to mediate its tumorigenic effects. To determine whether the oncogene is also required for the recruitment of MED13L and RERE to their

co-bound chromatin regions, we performed a ChIP-qPCR experiment using the A673 cell line carrying an inducible shFLI1 [117]. After inducing FLI1 knockdown with 1 µg/mL of doxycycline for 72 hours, ChIP-qPCR using MED13L and RERE antibodies was performed. As shown in **Figure 28**, depletion of EWSR1-FLI1 impaired MED13L and RERE binding to target genes, suggesting that the oncogene is required for MED13L and RERE binding to the unique chromatin regions where they coincide. Furthermore, as RERE lies in a genomic region bound by MED13L and RERE but not by the oncogene, the decrease of MED13L and RERE binding after FLI1 depletion suggests that the oncogene is also responsible for their binding to regions where EWSR1-FLI1 is not present.

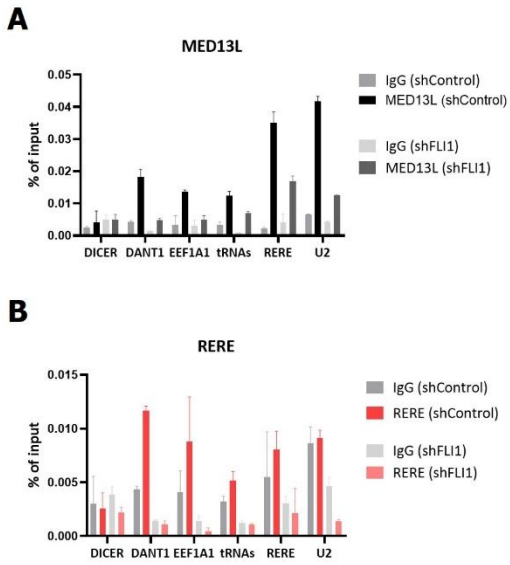


Figure 28. **EWSR1-FLI1 is necessary for MED13L and RERE binding to their target regions.** ChIP-qPCR of MED13L **(A)** and RERE **(B)** in the A673 cell line bearing the inducible FLI1 knockdown or the control vector, and treated with 1 µg/mL of doxycycline. Data correspond to the enrichment ratio of immunoprecipitated samples relative to input. IgG was used as a negative control for each immunoprecipitation and DICER1 was used as control region.

4. Mediator and RERE interactions in Ewing sarcoma

4.1. Proteins of the Mediator complex colocalize with EWSR1-FLI1

ChIP-seq data showed that the oncogene binds to up to 30 percent of the genes to which MED13L and RERE bind, even though they do not share the same chromatin binding sites. To investigate the possibility that EWSR1-FLI1 localizes to nuclear regions near MED13L and RERE either because they are in the same chromatin region or because they are in contact through chromatin loops, we studied their subcellular distribution by immunofluorescence (IF). To overcome the technical limitation due to both MED13L and FLI1 antibodies being produced in rabbit, A673 cells were transduced with a FLAG-tagged EWSR1-FLI1 in a GFP-bearing lentiviral construct (see Materials and Methods section 4.2). Thus, subcellular localization of the ectopic oncogene could be detected with a mouse FLAG antibody. This approach was validated by IF using confocal microscopy, as the signal corresponding to the FLAG-tag oncogene effectively overlapped with the endogenous FLI1 in the discrete nuclear domains (**Figure 29A**).

Once proper localization of the tagged oncogene was confirmed, FLAG-EWSR1-FLI1-transduced A673 cells were fixed and subjected to immunostaining to simultaneously detect MED13L and FLAG in transfected cells (GFP positive) (**Figure 29B**). Confocal imaging revealed the subcelullar localization of MED13L being mainly nuclear, with a small fraction seen in the cytoplasm. With respect to the oncogene (FLAG), the subcellular

distribution was nuclear, as expected, but the signal was weak, and few nuclei had an intensity that would not allow conventional colocalization analyses. Therefore, we recorded the intensity histogram of each fluorochrome in a selected nuclear region. Plot profiling showed that the fluorescent signals of FLAG and MED13L were coincident, suggesting that FLAG-EWSR1-FLI1 and MED13L colocalize in discrete nuclear regions (**Figure 29C**). Technical limitations precluded the study of FLAG-EWSR1-FLI1 colocalization with other Mediator subunits and RERE.

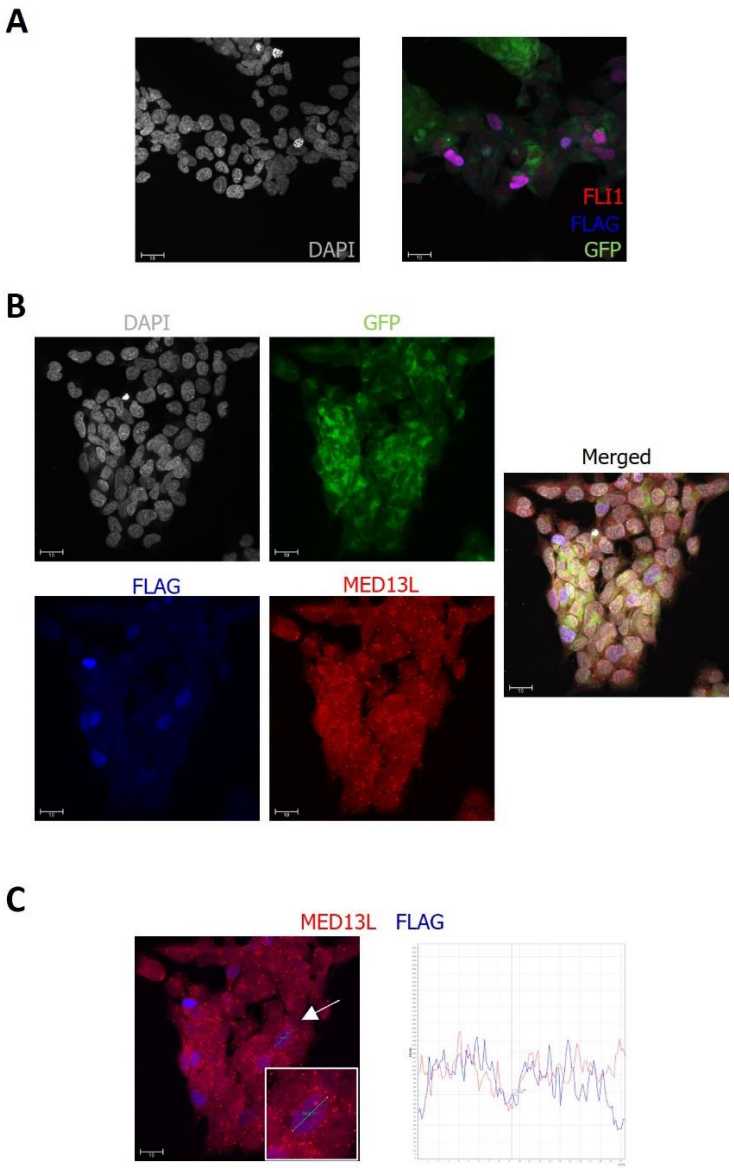


Figure 29. **MED13L colocalize with FLAG-tagged EWSR1-FLI1. A)** EWSR1-FLI1 was Flag-tagged and cloned into a lentiviral vector with GFP. A673 cells were infected with lentiviral FLAG-EWSR1-FLI1 supernatants, and double immunofluorescence (IF) to detect FLAG (blue) and FLI1 (red) was performed. Cell nuclei were identified by DAPI staining (grey). GFP (green) indicated cells infected with FLAG-EWSR1-FLI1. Scale bar: 10 µm. **B)** IF detection of FLAG (blue) and MED13L (red) in A673 cells expressing FLAG-EWSR1-FLI1-GFP. Cell nuclei, in grey. Scale bar: 10µm. **C)** Plot profile (right) depicting the histogram intensities of MED13L (red) and FLAG (blue) in a region of the cell nucleus of an A673 cells expressing FLAG-EWSR1FLI1-GFP (left). Obtained with LAS X (Leica) software. Scale bar: 10 µm.

4.2. Proteins of the Mediator complex interact with EWSR1-FLI1

The subcellular distribution of EWSR1-FLI1 and MED13L suggested that the two proteins may physically interact at least in some chromatin regions. To investigate this possibility, we performed immunoprecipitation (IP) assays in A673 extracts with antibodies for FLI1, MED1, MED12, MED13 and RERE, and the interacting proteins were analyzed by WB. MED13L and MED12 immunocomplexes effectively pulled down endogenous EWSR1-FLI1, although MED12 and MED13L could not be detected when the oncogene was pulled down with a FLI1 antibody, probably due to steric hindrance effects (**Figure 30A**). To overcome this limitation, we transduced A673 cells with the FLAG-tagged oncogene, and cell extracts were immunoprecipitated with a FLAG antibody. WB confirmed the presence of endogenous MED13L in the FLAG-EWSR1-FLI1 complex (**Figure 30B**). This result reveals a direct interaction of EWSR1-FLI1 with this subunit of the kinase module of the Mediator complex. In addition, MED1 was detected when MED12 and MED13L were immunoprecipitated, the MED12 subunit was observed in immunoprecipitations of MED1 and MED13L and, conversely, MED13L was detected in MED12-bound immunocomplexes. These results indicate that the Mediator subunits analyzed interact among themselves as part of the same complex.

Unsuccessfully we tried to detect any potential interaction between endogenous RERE and the subunits of the Mediator complex or EWSR1-FLI1, likely due to the limited availability of commercial RERE antibodies.

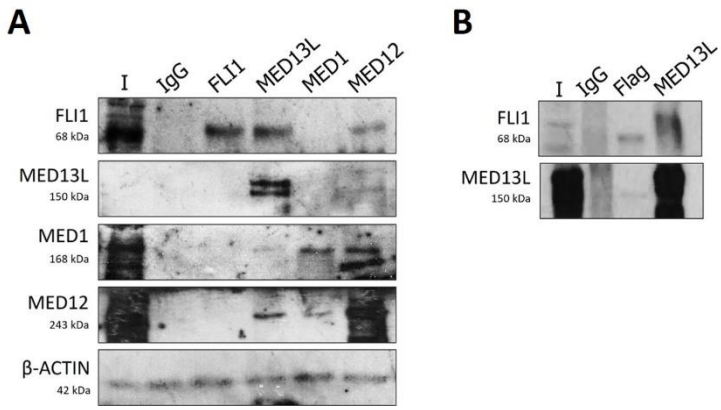


Figure 30. **The Mediator complex proteins MED12 and MED13L interact with EWSR1-FLI1. A)** Equal amounts of A673 cell extracts were immunoprecipitated with antibodies against FLI1, MED1, MED12 and MED13L, and bound proteins were detected by WB with the indicated antibodies. IgG was used as a negative control. Ten percent of the cell lysates were used as input. β-Actin was used as loading control. **B)** A673 cells were infected with lentiviral FLAG-EWSR1-FLI1 supernatants and cell extracts were generated. Immunoprecipitations using antibodies against FLAG and MED13L were performed, and bound proteins were detected.

4.3. RERE colocalizes with STAG2 in discrete nuclear domains

As mentioned above, Mediator, RERE and STAG2 bind to overlapping chromatin regions. Since technical problems prevented the detection of direct interactions between RERE and the other proteins by IP techniques, we investigated the subcellular localization of RERE, whose antibody was generated in rabbit, and STAG2, detected with a mouse antibody, therefore compatible for colocalization studies. Merging images with fluorochromes corresponding to cell nucleus (blue), RERE (red) and STAG2 (green) showed yellow spots that indicate RERE and STAG2 colocalization in discrete nuclear domains (**Figure 31A**). Confocal quantitation resulted in an Overlap Coefficient correlation of 0.77, indicating a less than 290 nm distance between the two proteins (**Figure 31B**). This data suggests that RERE is part

of the complex formed by STAG2 in DNA loops linking regulatory regions and promoters of target genes.

Study of the colocalization between STAG2 and the Mediator complex subunits MED1, MED12 and MED13L was challenging due to signal disbalance. In addition, RERE colocalization with Mediator subunits was not feasible since the antibodies for all these proteins were of rabbit origin.

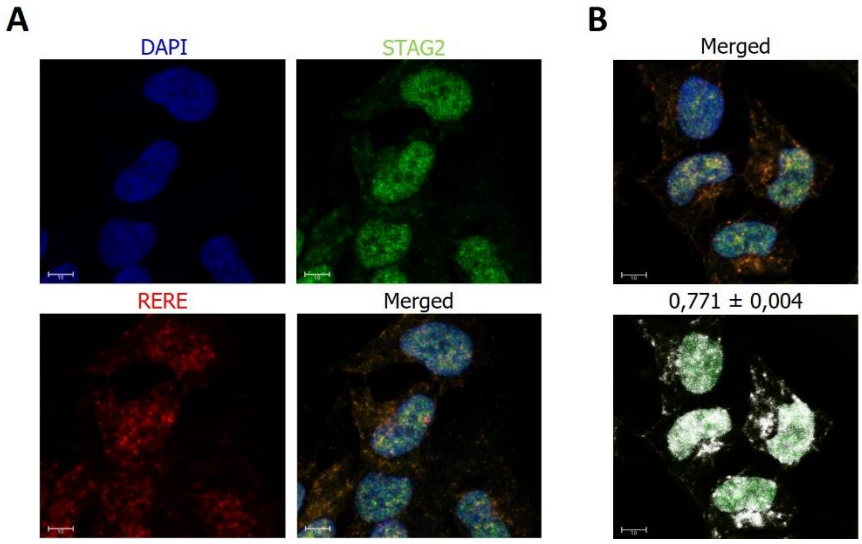


Figure 31. **RERE colocalizes with STAG2 in the nucleus of A673 cells. A)** IF staining of STAG2 (green) and RERE (red) in A673 cells. Cell nuclei were stained with DAPI staining (blue). Scale bar: 10µm. **B)** Merged dual IF image detecting STAG2 and RERE (top) and mask image (bottom) showing quantification of colocalization by Overlap Coefficient in A673 cells using Leica LAS X software. Scale bar: 10µm. Data is represented as mean ± S.E. of 2 fields.

4.4. STAG2 and EWSR1-FLI1 stand in close proximity in the nuclei of A673 cells.

In the literature is described how the Mediator complex forms a DNA-loop between gene regulatory regions and promoters with the help of the cohesin STAG2 [80]. In A673 cells, EWSR1-FLI1 cooperates with STAG2 to

stablish part of the oncogenic neo-enhancer or super-enhancer-promoter structure [146, 147]. Altogether, these data suggested that STAG2 and the oncogene, if not directly interacting, could stand in close proximity in the nuclei of Ewing sarcoma cells. To explore this possibility, subnuclear colocalization of STAG2 and EWSR1-FLI1 was studied by IF (**Figure 32A**). Merging of the three fluorochrome signals revealed yellow spots in discrete regions corresponding to FLI1 and STAG2 colocalization. Confocal quantification of both signals resulted in an Overlap Coefficient of 0.85, indicating that FLI1 and STAG2 stand at <290 nm distance (**Figure 32B**). To better characterize this result, we performed Fluorescence Resonance Energy Transfer (FRET), which determines the interaction between to molecules within several nanometers. FRET is a process by which transfer of energy occurs from an excited state fluorophore to a second chromophore in close proximity [135]. FRET mean efficiency for FLI1 and STAG2 interaction was 0.24 (**Figure 32C**), which states that FLI1 and STAG2 co-localize at less than 100 Å, a distance close enough for molecular interactions to occur.

Altogether, these data indicate that EWSR1-FLI1 is in the same nuclear domains as STAG2, and probably Mediator and RERE, and suggest that the interaction between these proteins and the oncoprotein results in the formation of DNA loops leading to the approximation of (super)enhancers to promoters and thus the transcriptional machinery.

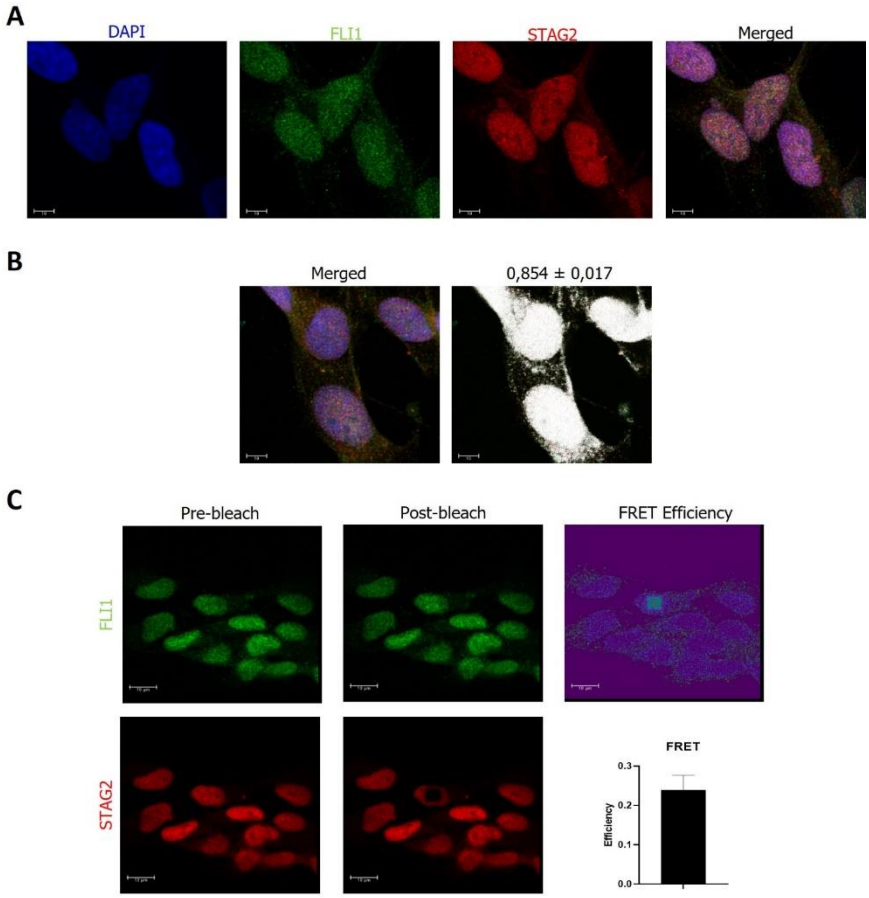


Figure 32. **STAG2 colocalizes with EWSR1-FLI1 in the A673 nucleus. A)** Immunofluorescence (IF) detection of STAG2 expression (red) and EWSR1-FLI1 (green) in A673 cells. Cell nuclei were identified by DAPI staining (blue). Scale bar: 10µm. **B)** Merged IF image of STAG2 and EWSR1-FLI1 in A673 cells and its corresponding mask image from the quantification of STAG2 and FLI1 colocalization by Overlap Coefficient using Leica LAS X software. Scale bar: 10µm. Data represent mean ± S.E of 4fields. **C)** Colocalization analysis of STAG2 and EWSR1-FLI1 by fluorescence resonance energy transfer (FRET). Left and middle images depict EWSR1-FLI1 and STAG2 before and after photobleaching. Sixteen regions of interest (ROI) were analyzed from different cells. FRET efficiencies were calculated with Leica LAS X software during STAG2 photobleaching. Data represent mean ± S.E.

5. Transcriptional changes in MED13L- and RERE-depleted Ewing sarcoma cells

5.1. MED13L and RERE depletion affects the same subset of genes

To determine the transcriptional changes mediated by MED13L and RERE, A673 cells transiently expressing three different siRNA sequences for each MED13L and RERE were used as biological replicates, and their RNA was purified and sequenced. Bioinformatic analysis identified upregulated and downregulated transcripts in MED13L- and RERE-depleted cells when compared to their expression in control (siGFP) cells (**Figure 33A**). The number of differentially expressed genes (DEGs) in MED13L-depleted cells was slightly higher than that in RERE (4866 vs. 4295, respectively; $p < 0.05$). Interestingly, 2766 (60%) of these DEGs were shared by both conditions (**Figure 33B**). These data agreed with ChIP-seq results, in which about 71% of MED13L and RERE binding sites overlapped (**Figure 20A**), suggesting that these two proteins share a functional role in Ewing sarcoma. It is important to remark that RNA sequencing was conducted using PolyA capture methodology, which precludes the detection of non-coding genes such as tRNAs, rRNAs or snRNAs, genes the oncogene binds to as shown by ChIP-seq. When the transcriptomes of MED13L and RERE knockdown cells were independently analyzed using control A673 cells (siGFP) as a reference, similar proportions of up- and down-regulated genes were observed (**Figure 33A**), further supporting the shared function of MED13L and RERE.

A

	Adj.P.Value < 0.05	|logFC| > 1 and Adj.P.Value < 0.05	P.Value < 0.05	|logFC| > 1 and P.Value < 0.05
siMED13L (up/down)	1184/670	361/216	3004/1862	786/374
siRERE (up/down)	837/455	216/173	2638/1657	516/369
siMED13L_RERE (up/down)	1421/877	314/259	3136/2078	557/349

B

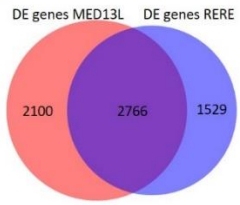


Figure 33. **Differentially expressed genes (DEGs) in MED13L- and RERE-depleted A673 cells by siRNA. A)** Descriptive table depicting the number of DEGs in MED13L- and RERE-depleted A673 cells, identified by RNA-seq. Data represents the mean of three replicates corresponding to each siRNA sequence. DEGs were obtained considering the indicated bioinformatics parameters. **B)** Venn diagram showing intersection between DEGs in MED13L- and RERE-depleted A673 cells (p val <0.05). DE, differentially expressed; FC, fold change.

Validation of the transcriptional changes was performed by RT-qPCR in A673 cells oligofected with three different sequences of siRNA of MED13L and RERE (**Figure 34**).

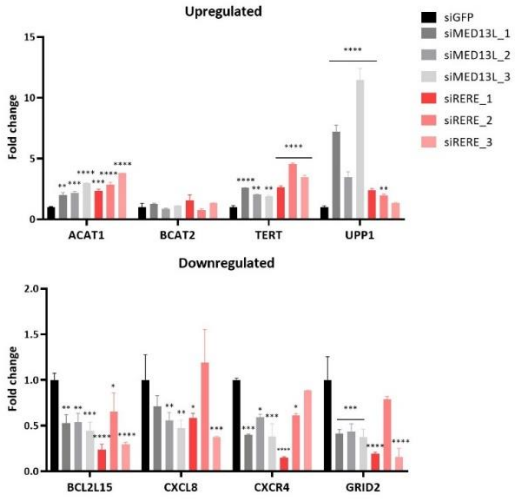


Figure 34. **Transcriptional changes in MED13L- and RERE-depleted A673 cells.** Validation by RT-qPCR of some of the transcriptional changes identified by RNA-seq in MED13L- or

RERE-depleted cells. Values are normalized to GAPDH and Actin B1. Two-way ANOVA and Dunnett's multiple comparisons test were used to statistically analyze the differences in expression. ****P value< 0.0001, ***P value <0.0005**P value <0.01, and *P value <0.05.

5.2. MED13L and RERE depletion impairs the expression of genes involved in RNA metabolism and ribosome function

DEGs in MED13L- and RERE-depleted cells compared to control cells were represented in heatmaps (**Figure 35A-B**). Gene Set Enrichment Analysis (GSEA) revealed that genes downregulated after MED13L depletion corresponded to ribonucleoproteins, protein translation and mRNA catabolic processes (**Figure 35C**), whereas RERE downregulation resulted in reduced expression of genes involved in mRNA splicing and aerobic electron transport chain (**Figure 35D**). Up-regulated genes in the MED13L knockdown condition included genes involved in acid catabolic processes, endothelial cells apoptosis and cell adhesion. For RERE-depleted cells, acid catabolic processes, intraciliary and microtubules in RERE-depleted cells (**Figure 35C and 35D**, respectively).

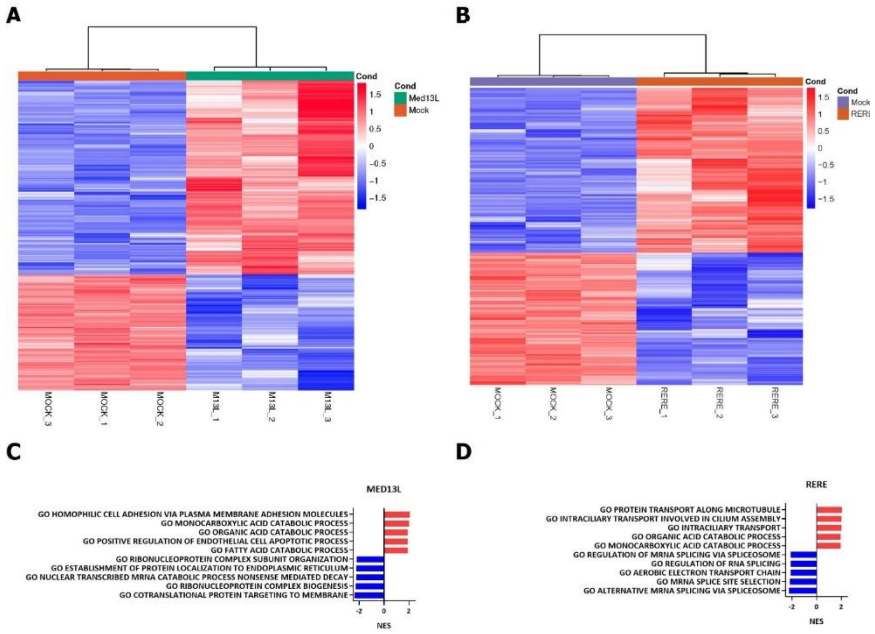
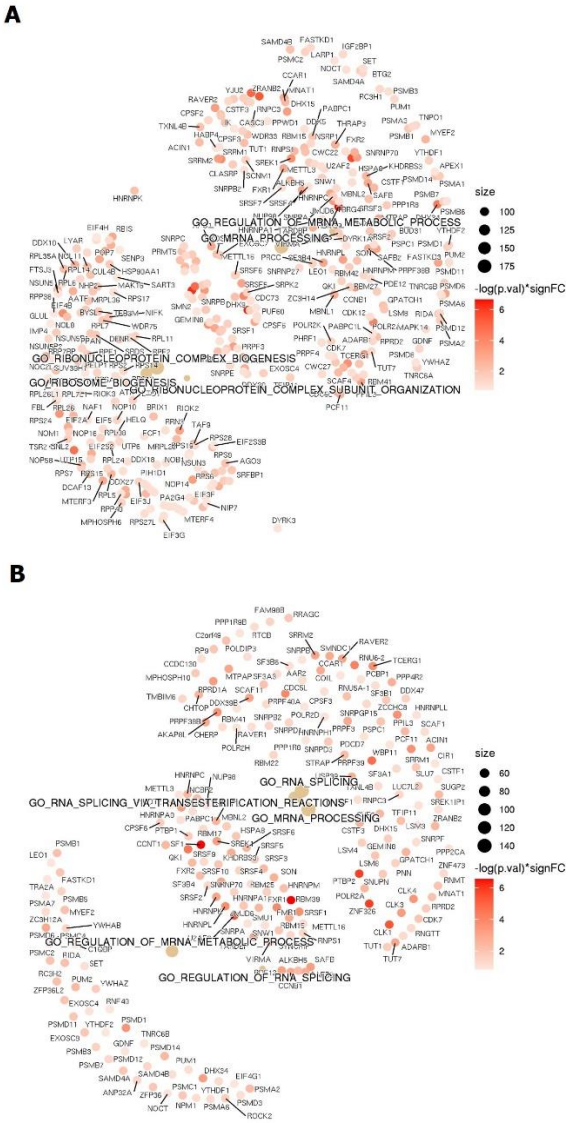


Figure 35. **MED13L and RERE depletion impairs the expression of genes involved in RNA metabolism. A-B)** Heatmaps depicting changes in gene expression in A673 cells oligofected with three different siRNAs for MED13L **(A)** and RERE **(B)**. Heatmaps display DEGs with an adjusted P value < 0.05 and |log FC| > 1 on the logCPM normalized counts. Z-scores are shown in the heatmaps. Control and MED13L/RERE replicates are indicated below. **C-D)** Bar plot of top five significantly reduced terms in MED13L **(C)** and RERE **(D)** depleted cells (NES positive), and top five significantly enhanced terms in MED13L **(C)** and RERE **(D)** depleted cells (NES negative). Adjusted P value <0.05. FC, Fold change; CPM, counts per million; NES, normalized enrichment score.

Gene concept network representation of the genes that contributed most to the enrichment in each category confirmed that the top 5 enriched terms for genes whose expression was reduced in MED13L-depleted cells were related to mRNA metabolic process, mRNA processing, ribonucleoprotein complex biogenesis, ribosome biogenesis and ribonucleoprotein complex subunit organization (**Figure 36A**). For RERE-depleted cells, RNA splicing and mRNA processing (**Figure 36B**). The top categories for genes induced in MED13L defective cells included extracellular structure organization, homophilic cell adhesion via plasma membrane adhesion molecule, positive regulation of GTPase activity,

regulation of GTPase activity and organic acid catabolic process (**Figure 36C**). For RERE-depleted cells, cilium organization, monocarboxylic, organic acids, small molecules catabolic processes and protein transport along microtubules (**Figure 36D**). In summary, MED13L and RERE depletion resulted in impaired expression of genes involved in RNA metabolism and enhanced expression of genes related to cell structure and environment, and catabolic metabolism.



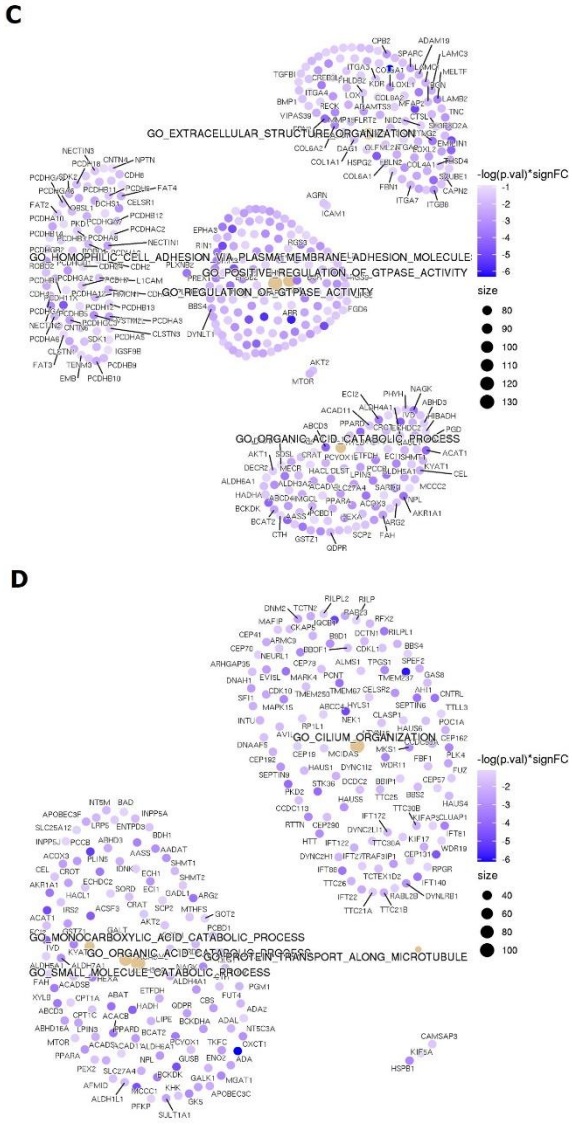


Figure 36. **MED13L and RERE depletion mainly reduce the expression of genes involved in splicing.** Gene-Concept Networks of the top five significantly downregulated terms in MED13L **(A)** and RERE **(B)** depleted cells, and upregulated terms in MED13L **(C)** and RERE **(D)** depleted cells. The subset of represented genes contributing most to the enrichment result was ordered by -log(P. value)*significant FC. Thus, in positively enriched gene sets (NES > 0) the genes are ordered in decreasing order of importance, and in negatively enriched gene sets (NES < 0) in increasing order of importance. FC, fold change; NES, normalized enrichment score.

To determine whether the transcriptional changes were due to direct binding of MED13L or RERE to the corresponding gene regions, RNA-seq and ChIP-seq data were intersected. This overlap revealed binding of MED13L and RERE to the genomic regions of 8-12% of upregulated and downregulated DEGs in MED13L- and RERE-depleted cells (**Figure 37A**). GO analysis of directly regulated genes showed that upregulated genes in MED13L- and RERE-depleted cells are involved in the organization of the cytoskeleton and cellular components, whereas downregulated genes participate in mRNA metabolic processes (**Figure 37B**). Thus, the functionality of genes directly regulated by MED13L and RERE is like that of global DEGs identified by RNA-seq. As indicated above, the number of DEGs directly regulated by MED13L and RERE is possibly underestimated, because of the limitation of RNA-seq to detect only changes in protein-coding genes.

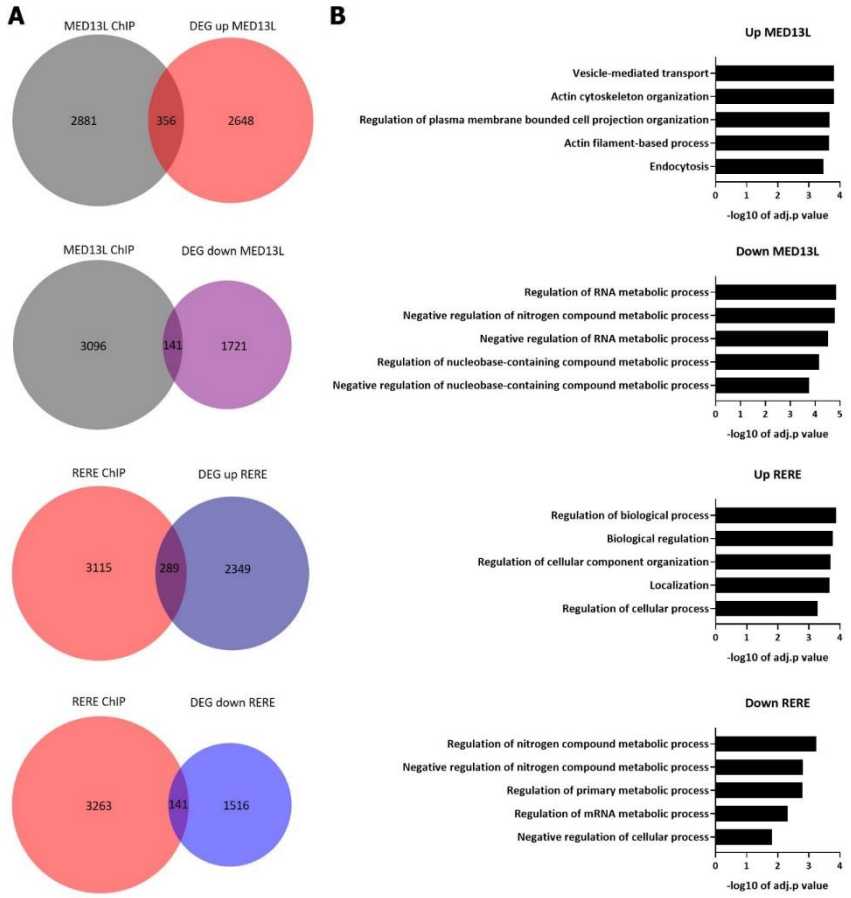


Figure 37. **RNA-seq and ChIP-seq data integration. A)** Venn diagrams depicting the overlap between genes associated to a ChIP-seq peak of MED13L or RERE and the upregulated or downregulated DEGs after MED13L and RERE siRNA knockdown. **B)** Bar plots showing statistically significant biological process categories from Gene Ontology (GO) analysis of the genes bound by MED13L or RERE and EWSR1-FLI1 in A673 cells with adjusted P value <0.05. Analysis was performed with gProfiler online database [144] (https://biit.cs.ut.ee/gprofiler/gost).

5.3. Expression of spliceosome factors is dependent on MED13L and RERE levels

MED13L- and RERE-depleted cells were defective in the expression of genes involved in mRNA metabolism (see above). GSEA from our RNA-seq data revealed an impoverishment in genes belonging to the biological process category of regulation of RNA splicing (**Figure 38A**). These effects were validated for selected genes by RT-qPCR and WB (**Figures 38B and 38C**, respectively).

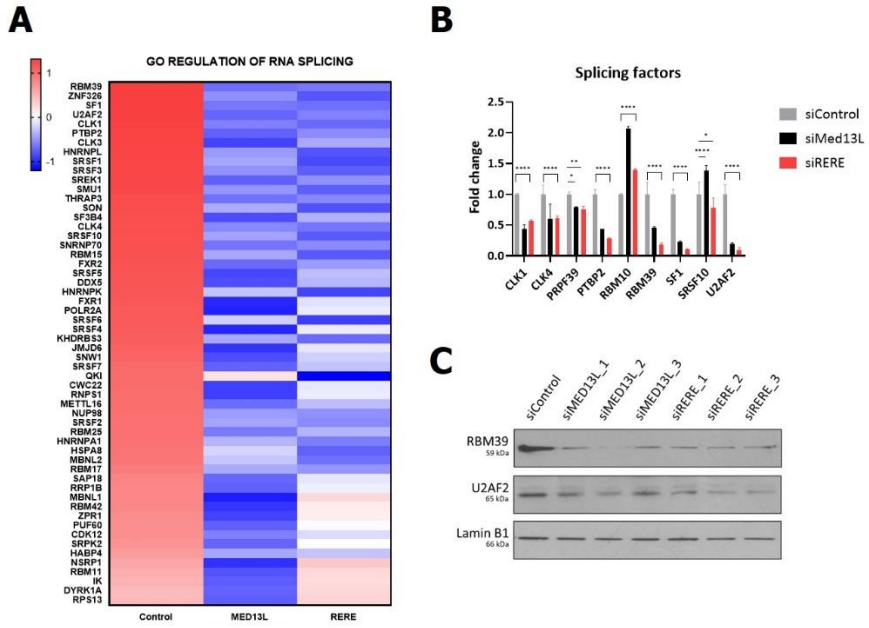


Figure 38. **MED13L and RERE depletion affects the expression of splicing factors. A)** Heat map depicting significant expression changes of GSEA genes corresponding to regulation of mRNA splicing induced by RERE and MED13L knockdown by siRNA in the A673 cell line. **B)** mRNA levels of different splicing factors after MED13L and RERE depletion by siRNA in the A673 cell line. Values were normalized to GAPDH. Statistical test to study significant differences was performed with two-way ANOVA and Dunnett's multiple comparisons test. ****P value< 0.0001, **P value <0.01, and *P value <0.05. **C)** Western blot showing the changes in protein expression of splicing factors after MED13L and RERE knockdown using three different siRNA sequences in A673 cells. Lamin B1 was used as a loading control.

6. MED13L and RERE regulate the alternative splicing machinery

6.1. MED13L and RERE depletion favors long exon inclusion in genes involved in chromatin organization and cell cycle

Alternative splicing (AS) generates transcriptomic complexity through differential selection of cassette alternative exons, alternative 5' and 3' splice sites, mutually exclusive exons, and alternative intron retention [106]. To better explore the role of MED13L and RERE in splicing, we carried out an analysis of the RNA-seq data using the pipeline VAST-TOOLS [131] (see Materials and Methods, 7.5 section). Public RNA-seq database including EWSR1-FLI1 esiRNA in A673 cell line was analyzed [134]. The events differentially occurring in each condition (control vs. MED13L-, RERE-, or EWSR1-FLI1-knockdown) were quantified according to the percent spliced-in (PSI). The number of events assessed in total (total EV) and events with a dPSI≥15 (total AS) was also categorized (**Figure 39A**). Classification of splicing categories revealed that both MED13L and RERE knockdown specifically increased events belonging to the category of longer alternative exons and had limited effects on other splicing categories. As described with ChIP-seq and RNA-seq data, we observed an overlap between the splicing regulation profiles in MED13L- and RERE-depleted cells. Events comprising alternative exons, intron retention and alternative 3' splice site regulated by MED13L and RERE overlapped significantly ($R > 0.9$), whereas alternative 5' splice sites exhibited weaker overlapping (**Figure 39B**). The splicing profiles in MED13L- and RERE-depleted cells were different from those of EWSR1-FLI1, whose downregulation did not

affect particular categories of splicing, suggesting that the oncogene would specifically regulate one type of splicing through the action of MED13L and RERE (**Figure 39A**, and data not shown). In addition, in accordance with the higher number of alternatively spliced long exons, most of the events evaluated after MED13L and RERE knockdown correspond specifically to exon inclusion in both conditions (**Figure 39C**).

A

AS TYPE	Higher in contol	Higher in MED13L	TOTAL EV	TOTAL AS
Microexons	26	11	1713	135
Long AltEx	128	1047	94140	6107
Intronret	143	13	80117	1576
Alt 3ss	NA	108	67824	2763
Alt5ss	NA	23	46401	1944

AS TYPE	Higher in contol	Higher in RERE	TOTAL EV	TOTAL AS
Microexons	26	22	1680	130
Long AltEx	116	1068	92461	6015
Intronret	105	11	78729	1590
Alt 3ss	NA	108	66247	2663
Alt5ss	NA	17	45192	1861

AS TYPE	Higher in contol	Higher in EWS-FLI1	TOTAL EV	TOTAL AS
Microexons	4	18	1807	114
Long AltEx	205	324	102646	4572
Intronret	59	36	88497	1466
Alt 3ss	NA	195	74882	2963
Alt5ss	NA	144	47781	2072

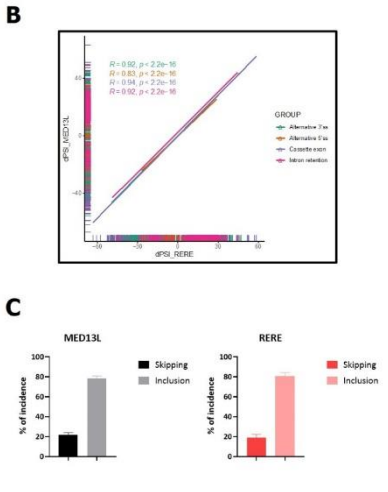


Figure 39. **MED13L and RERE depletion promote inclusion of long exons. A)** Summary of the VAST-TOOLS output comparing three controls (siControl) *vs*. three knockdown samples (siMED13L, siRERE and esiEWSR1-FLI1 [134]). The events were classified as microexons, longer alternative exons, intron retention, alternative 3' splice site (or alt. acceptor) and alternative 5' splice site (alt. donor) events. For each category, the number of events higher in control, higher upon knockdown, the number of events assessed in total (total EV), and events with differential percent spliced-in index (dPSI) ≥15 (total AS) are shown. **B)** Linear plot displaying the correlation between dPSI of genes affected by MED13L and RERE knockdown considering the four splicing datasets. **C)** Bar plot depicting the percentage of events related to exclusion *versus* inclusion upon MED13L and RERE knockdown by siRNA in A673 cells. AS, alternative spliced; EV, events; dPSI, delta percent-spliced-in.

Alternative splicing alterations in the form of inclusion of exons lead to larger mRNA isoforms. Inversely, when exons are skipped, they result in shorter isoforms. To validate the splicing profile originated due to MED13L and RERE depletion by siRNA, mRNA isoforms of some genes were analyzed by PCR and ran in agarose or polyacrylamide gels (see Materials

and Methods, 7.2 section). The splicing profile of some of the affected genes was determined in A673 cells. APAF1, BRD8, EZH2, and U2SURP transcripts underwent exon inclusion after MED13L and RERE depletion (**Figure 40**). On the contrary, the larger exons of BIN1, CLASP1, MICAL3, and SOX2 were excluded and smaller transcripts increased (**Figure 40**).

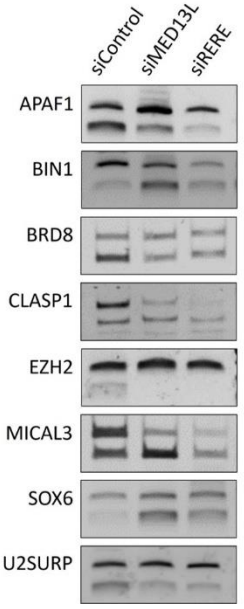


Figure 40. **Alternative splicing profile changes upon MED13L and RERE knockdown by siRNA.** Validation by PCR of alternative splicing changes in selected genes induced by MED3L and RERE depletion by siRNA in Ewing Sarcoma A673 cell line. Sequence #1 for MED13L and RERE siRNA knockdown was used to perform these experiments.

To elucidate which pathways were affected by alternatively spliced genes in MED13L- and RERE-depleted cells, we conducted Gene Ontology analysis. In the MED13L and RERE knockdown conditions, the main categories included chromosome and chromatin organization, histone modifications, organelle regulation and cell cycle process, and in RERE-depleted cells also macro autophagy (**Figure 41A-B**). These results indicate that both MED13L and RERE regulate splicing of a specific subset of genes that influence general cellular functions. Ontology analysis of genes whose

splicing was affected by EWSR1-FLI1 silencing included DNA repair, cytoskeleton organization, histone modification, neuronal development and protein modification by conjugation or removal of small proteins (**Figure 41C**). Therefore, splicing regulation by MED13L and RERE would be different from that of EWSR1-FLI1. The oncogene regulated splicing would affect genes with a wide range of functions, whereas MED13L and RERE controlled splicing would specifically regulate chromatin organization and the cell cycle.

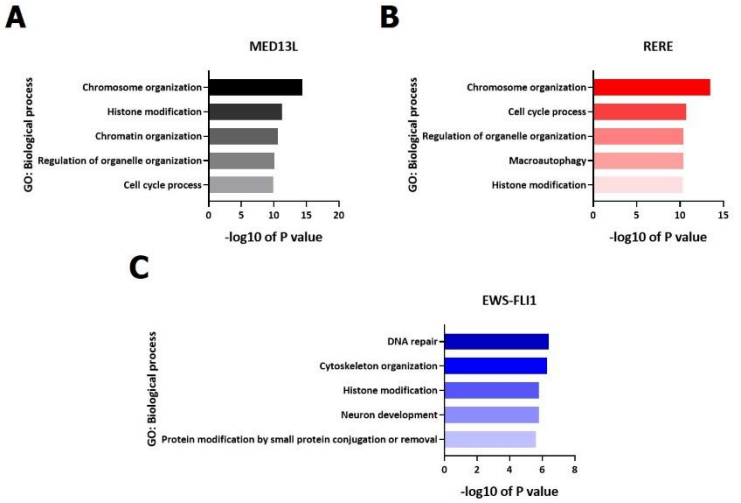


Figure 41. **MED13L and RERE knockdown affects the splicing of genes involved in chromosome organization and cell cycle.** GO biological process categories for 1191 genes associated to AS events induced by siRNA MED13L knockdown **(A)**, 1162 genes associated to AS events induced by RERE knockdown by siRNA **(B)** and 846 genes with altered splicing due to EWSR1-FLI1 siRNA [134]. **(C)**. Analyses were performed with the TopGene online database [148] (https://toppgene.cchmc.org/). GO, gene ontology.

6.2. Molecular mechanisms involved in AS in MED13L- and RERE-depleted cells

Transcription by RNA polymerase II gives rise to all nuclear protein-coding mRNAs and a large set of non-coding RNAs, and is strictly regulated and

coordinated by RNA maturation processes, including splicing [149]. To investigate the possibility that MED13L and RERE directly binds to the genomic regions affected by co-transcriptionally regulated AS, we crossed the coordinates of the regulated events with the binding sites of MED13L and RERE, previously determined by ChIP-seq. We found that only two of about 1500 regulated events had binding of MED13L and RERE in the same coordinate of the splicing event. Furthermore, only sixty-six MED13L and RERE binding sites corresponded to genes displaying an altered splicing profile (data not shown). These data indicates that the effects of MED13L- and RERE-depleted cells on AS might be mediated by splicing factors and not directly, co-transcriptionally regulated.

To assess this hypothesis, we generated lists of splicing [150] and spliceosome regulatory genes [151] from public databases. RNA-seq data from MED13L- and RERE-depleted cells were interrogated to identify AS factors whose expression was directly regulated by MED13L and RERE. 24-22% of splicing regulators were DEGs in MED13L- and RERE-depleted cells, while 13-15% of spliceosome genes were DEGs in cells with MED13L and RERE knockdowns (**Figure 42A-B**). Among the genes to which MED13L and RERE bind, only 5% were splicing factors and, of these, 2% were differentially expressed in cells expressing MED13L and RERE knockdowns (**Figure 42A-B**). In contrast, 39% of the AS regulatory network genes were found to be direct EWSR1-FLI1 targets, along with enrichment of the oncogene in 29% of spliceosome components (**Figure 42C**). These data indicate that the effects of MED13L and RERE depletion in AS are mediated by the controlled expression of splicing factors. This is consistent with GSEA from the RNA-seq showing regulation of mRNA splicing and the binding of MED13L and RERE to genes involved in mRNA metabolism that are transcriptionally

altered (**Figure 35-38**). Instead, EWSR1-FLI1 regulation of splicing is accomplished through binding to the splicing network.

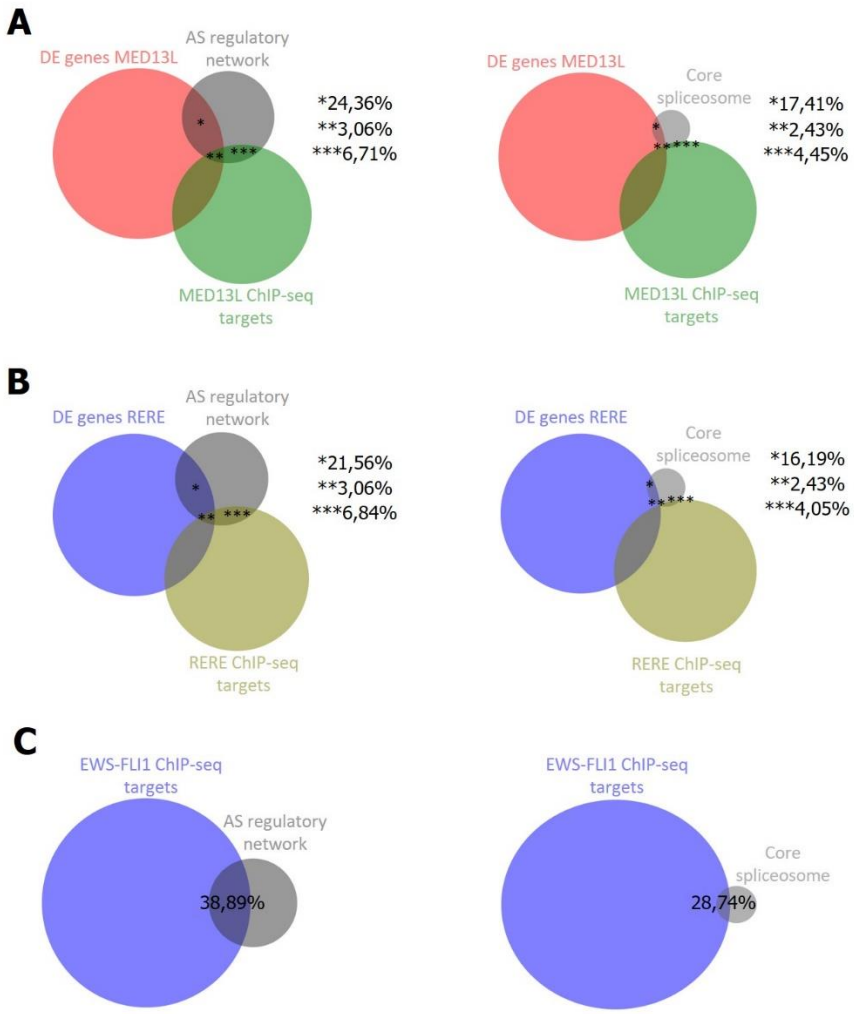


Figure 42. **MED13L and RERE depletion affects the expression of AS regulatory genes and spliceosome components.** Venn Diagrams displaying the overlap between AS regulatory network genes (left) and core spliceosome genes (right) with DEGs in MED13L **(A)** and RERE **(B)** depleted cells identified by RNA-seq and ChIP-seq annotated genes for MED13L **(A)**, RERE **(B)** and EWSR1-FLI1 **(C)** in A673 cells. AS regulatory genes were obtained from Han *et al.*, (2017) [151] and core splicesome genes from Wahl and Lührmann (2015) [150]. AS, alternative splicing.

6.3. Global impact of AS events in MED13L- and RERE-depleted cells on protein-coding features

To determine the impact of the AS events on protein diversity due to MED13L and RERE depletion, we used the VastDB protein impact database (https://vastdb.crg.eu). AS events were classified as exon inclusion or skipping and effects were grouped as generation of alternative 3' or 5´ untranslated region (UTR), alternative protein isoforms, uncertain impact with changes in the coding sequence (CDS), open reading frame (ORF) disruption, protein isoforms when splice site is used, and non-coding isoforms. Prediction of the effects due to inclusion and skipping events favored by MED13L and RERE knockdowns involved alternative protein isoforms and ORF disruption (**Figure 43**). Hence, MED13L and RERE would mediate AS events that will end up in the formation of new protein isoforms that might contribute to Ewing sarcoma tumorigenesis.

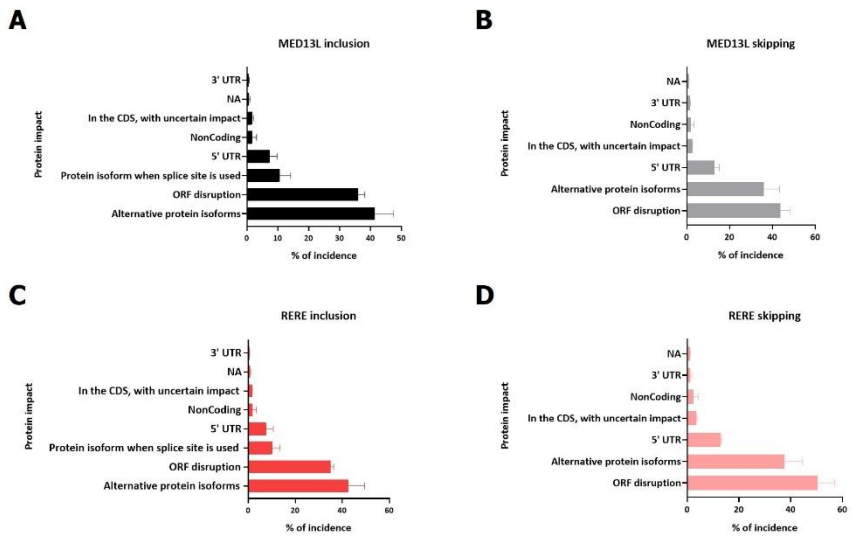


Figure 43. **Prediction of AS impact in MED13L- and RERE-depleted cells on protein-coding features.** Bar plots representing the prediction of the percentage of incidence of the protein impact caused by AS changes in exon inclusion and skipping after MED13L **(A and B, respectively)** and RERE **(C and D, respectively)** depletion by siRNA in A673 cells. Protein

impact dataset was downloaded from VastDB online database (https://vastdb.crg.eu/). CDS, coding sequence; ORF, open reading frame; UTR, untranslated region.

7. AS-based vulnerabilities in Ewing Sarcoma

7.1. Splicing inhibitor Pladienolide B

Direct regulation of MED13L and RERE expression by the oncogene suggested that tight control of AS is critical in Ewing sarcoma. Pladienolide B is a splicing inhibitor that targets the spliceosome factor SF3B1, preventing U2 assembly into the splicing stie since it stabilizes U2 in the BPS [107, 152, 153]. To assess splicing related vulnerabilities, we used Pladienolide B inhibiting U2 function as the main splicing factor and a target of EWSR1-FLI1, MED13L and RERE. We determined the IC50 in Ewing sarcoma cell lines A673, A4573, SKES1, and TC71, and other tumor types cell lines to test the specificity and sensitivity. Cell viability assays indicated that Ewing sarcoma cell lines were more sensitive to Pladienolide B treatment compared to any other cell lines tested (**Figure 44**).

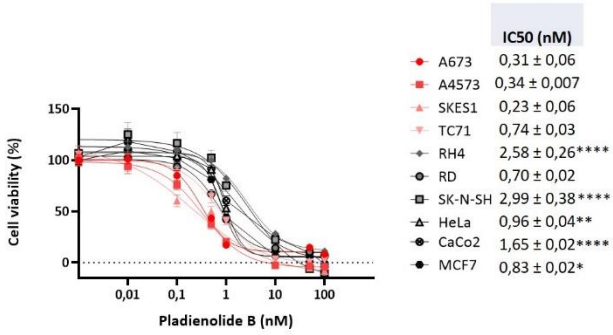


Figure 44. **Ewing sarcoma cell lines are more sensitive to Pladienolide B inhibitor.** Cell viability and IC50 calculation for the cell lines A673, A4573, SKES1 and TC71 (Ewing sarcoma), RH4, RD (rhabdomyosarcoma), SK-N-SH (neuroblastoma), HeLa (cervix cancer), CaCo2 (colon cancer) and MCF7 (breast cancer) with different concentrations of

Pladienolide B after 72 hours. Ordinary one-way ANOVA and Dunnett's multiple comparisons test were used to compare treatment differences between A673 and the other cell lines. ****P value <0.0001, **P value 0.088, *P value 0.0374.

7.2. Effect of Pladienolide B on MED13L- and RERE-depleted cells

To corroborate MED13L and RERE role in splicing regulation, A673 cells expressing inducible shRNAs for MED13L or RERE were treated with Pladienolide B for 72h. Since splicing is an essential mechanism for all the cells, low doses of the splicing inhibitor were used for the experiment as the inhibitory effect of high doses would mask the knockdown effect. We observed that control cells were significantly more sensitive than MED13L- and RERE-depleted cells (**Figure 45**) suggesting that MED13L and RERE drive splicing in the A673 cell line.

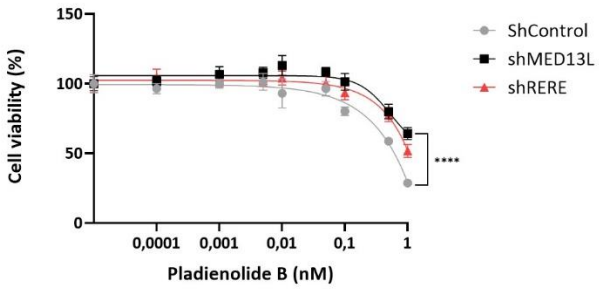


Figure 45. **MED13L and RERE depletion decreases Pladienolide B inhibitory effect.** Calculation of cell viability of A673 cell line after MED13L (shMED13L) and RERE (shRERE) inducible knockdown with 100 ng/mL of doxycycline, as compared to shControl with different Pladienolide B concentrations. Cells were treated for 72 hours. Two-way ANOVA and Dunnett's multiple comparisons test were used to compare treatment differences upon MED13L and RERE depletion. ****P value <0.0001.

7.3. Treatment with Pladienolide B in combination with CDK8 inhibitor BRD6989

Pladienolide B derivatives have been tested in clinical trials. Although treatment with these inhibitors affected splicing in tumors, no therapeutic advantages were observed. In addition, ocular toxicities appeared in some cases [152, 154, 155]. Considering this, we sought to find combined treatments. As MED13L is part of the Mediator kinase module, we used BRD6989, an analog of cortistatin A which is the natural inhibitor of the CDK8 Mediator kinase [156]. Ewing sarcoma and other pediatric and adult cancer cell lines were treated for 72 hours with increasing doses of Pladienolide B from 0.01 to 100 nM. Since BRD6989 has low solubility, ultrasonication and high DMSO amounts were required causing an impact on cell viability. Therefore, combination experiments were designed to test different concentrations of Pladienolide and a fixed concentration of BRD6989 of 5 µM. Cell viability assays showed that the combination was more effective than Pladienolide B as single therapy in Ewing sarcoma cell lines and in HeLa and MCF7 cell lines. In the case of the rhabdomyosarcoma cell lines RH4 and RD, the combination was not more effective than single Pladienolide B treatment. Nevertheless, sensitivity was higher for the Ewing sarcoma cell lines A673 and SKES1 (**Figure 46**). CompuSyn software analysis revealed that combination of Pladienolide B and BRD6989 was synergistic in A673 and in MCF7 cell lines (data not shown).

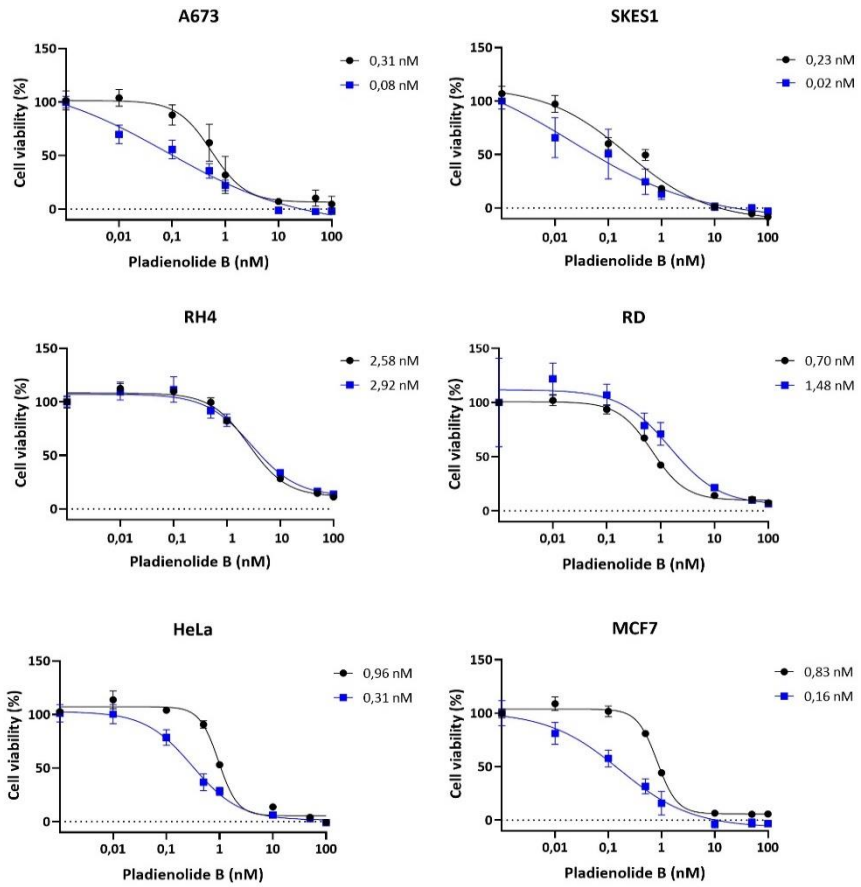


Figure 46. **Ewing sarcoma sensitivity to Pladienolide B treatment in combination with BRD6989.** Cell viability and IC50 calculation for the cell lines A673, SKES1, RH4, RD, HeLa and MCF7 with different concentrations of Pladienolide B inhibitor (black) and with Pladienolide B in combination with 5 µM of BRD6989 (blue) after 72 hours. IC50 values are indicated for each treatment.

DISCUSSION

1. MED13L and RERE expression in Ewing sarcoma

EWSR1-FLI1 is, in most cases, the only genetic alteration found in Ewing sarcoma tumors since they have a low mutation rate (0.15 mutations/megabase) [37]. Considering this and the activating and repressing transcriptional abilities of the oncogene, it is widely assumed that the oncogene is the sole responsible for tumorigenesis by epigenetic deregulation in the right cellular context [29,85, 86, 94]. Changes in histone modifications and recruitment of diverse proteins orchestrated by the oncogene contribute to alter the chromatin conformation to generate *de novo* enhancers and super-enhancers. As in many other types of tumors, super-enhancers regulate the expression of important genes for cell state maintenance and tumorigenesis [73, 74, 86]. In this project, we have identified MED13L and RERE as EWSR1-FLI1 super-enhancer-regulated targets in the A673 cell line [86], which are differentially expressed in Ewing sarcoma.

MED13L is a subunit of the kinase module of the Mediator complex, which acts as a coordinator of cell lineage specificity by interacting with master transcription factors that will regulate gene expression [88-90]. In the case of RERE, the phenotypic similarities of patients with RERE and CHD7 deficiencies [141, 142, 143] prompted us to consider that RERE might also be part of the complex of DNA loops formed between regulatory regions and their target genes. Thus, RERE would have a role in transcription regulation and chromatin architecture.

Significant enrichment of chromatin regulators and cofactors associated to RNA pol II in super-enhancer regions leads to high expression of associated genes [74]. For instance, super-enhancer-regulated genes such as MEIS1 or CCND1 are highly expressed in Ewing sarcoma [95, 96]. According to the DepMap database (https://depmap.org/portal/), which collects gene expression at the mRNA level in hundreds of tumor cell lines, RERE is significantly overexpressed in Ewing sarcoma cell lines, but MED13L expression is like other cancer cell lines. Mediator is found in all eukaryotic cells and plays an essential role in a plethora of complex chromatin functions. However, Mediator subunits may have altered expression patterns in many types of cancer [157]. This fact could explain why MED13L mRNA levels were similar to those of other tumors. For instance, in non-small-cell lung cancer, MED13L was more expressed than in normal lung tissue [138] and, in colon cancer cell lines, MED13L has been reported to be required for expression of cancer-associated super-enhancer genes [158]. Another explanation could be that, due to the experimental culture conditions, tumor cell lines are artificially enriched in cancer stem cells, in which the role of Mediator would be particularly important for aberrant expression of lineage master genes. Nevertheless, immunohistochemistry of primary Ewing sarcoma tumors showed that MED13L and RERE were differentially expressed compared to other pediatric tumors.

RERE mRNA and protein levels, but not MED12 levels, were dependent on EWSR1-FLI1 expression, indicating the specificity of transcriptional regulation by EWSR1-FLI1 of this gene. In contrast, while MED13L protein levels were effectively lowered, MED13L mRNA levels were not altered when the oncogene levels were reduced. This apparent inconsistency between MED13L mRNA and protein levels suggests MED13L post-

translational regulation by EWSR1-FLI1. Possibly by altering transcription-associated MED13L splicing, which generates diverse mRNA species with different translational efficiencies. Also, microRNAs regulated by the oncogene could be interfering in MED13L translation, similarly to miR-30 interacting with the 3'UTR region of CD99 to regulate its protein expression but not the mRNA levels [159]. In any case, our findings suggest a role of the two super-enhancer-regulated genes MED13L and RERE in Ewing sarcoma tumorigenesis.

2. Chromatin regions bound by MED13L and RERE

As mentioned previously, the Mediator complex functions as a dynamic bridge to bind enhancers and super-enhancers to promoters of target genes forming DNA loops [71, 73, 74, 77, 79, 80]. The chromatin remodeler CHD7 has been shown to interact with p300 and to localize to distal genomic regions corresponding to enhancers [160]. Considering the overlapping phenotype of individuals with deficiencies in CHD7 and RERE, we expected RERE to be found in similar chromatin regions. In addition, a role for RERE in chromatin regulation has been previously described, recruiting histone deacetylases HDAC 1 and 2 and histone methyltransferase G9a through their ELM2/SANT domains [161, 162]. Our ChIP-seq results demonstrate that the Mediator complex and RERE bind to genome-wide overlapping regions in A673 cells. The peaks corresponded mainly to distal intergenic regions and transcription-associated chromatin states in H1 cells. The relationship

of RERE with the Mediator complex has not previously been described in Ewing sarcoma, nor in any other cellular context.

The architecture of DNA loops formed in chromatin is arranged by Mediator binding to cohesin proteins such as STAG2 [80]. We observed that most STAG2-binding regions were also enriched in MED13L and RERE. In addition, STAG2 was, similarly to MED13L and RERE, enriched in distal intergenic regions and transcription-associated chromatin states. This suggested that MED13L, RERE and STAG2 might be regulating the same chromatin functions. In contrast, we observed that the genomic distribution of MED12 and the associated chromatin states were different from those of MED1 and MED13L. Although Mediator complex is conserved in eukaryotes and mainly functions as a transcription coordinator, subunits composition and functions can differ depending on the cell context [71, 72, 75, 76]. For instance, MED23 is involved in insulin-signaling by inducing RAS-MAPK-dependent pathways, while MED1, MED15 and MED23 have been reported to be involved in lipid metabolism. Both MED12 and MED13L are part of the kinase module of Mediator, which has been implicated in different developmental pathways[75]. MED12 is required for the activation of super-enhancer genes in colon cancer [158] and in mouse ESC (mESC) [83]. In the latter, MED12 induces super-enhancer pluripotency genes by promoting a DNA loop due to the interaction with the cohesin loading factor NIPBL [80, 83] and by stabilizing p300 binding leading to an enrichment in histone H3K27ac at specific enhancers [83]. The same function has been reported for MED13L in non-small-cell lung cancer and colon cancer cell lines, where is required for p300 to stablish H3K27ac marks and enable the expression of super-enhancer genes [138, 158]. In contrast, in mouse hematopoietic stem cells, MED12 activation of super-

enhancers was independent of MED13 activity, indicating that although both subunits are part of the kinase module, they do not consistently operate as a single unit [83]. In line with this, our ChIP-seq results indicate a different role for MED13L and RERE in Ewing sarcoma, as supported by their genomic distribution and regulation by EWSR1-FLI1.

Genome-wide analysis of the DNA binding motifs for MED13L, RERE and STAG2 showed enrichment in CA microsatellites. Microsatellites represent 3% of the human genome, with dinucleotides repeats being the most frequent. Among them, CA repeats present with length polymorphisms and are the most recurrent followed by AT, GA and GC [163]. CA microsatellites participate in the modulation of gene expression by stablishing a Z-DNA (left-handed DNA) conformation. For instance, the expression of *IFN-γ*, *EGFR* or *HSD11B2* depends on CA repeats polymorphisms [164]. Hui *et al.* demonstrated that CA microsatellites enhance intron splicing through length-dependent binding of the splicing factor hnRNP-L to genes with CA-rich enhancers such as *eNOS* [165]. Our findings show that MED13L and RERE binding to genes of the splicing machinery could not account for the observed alterations in the AS pattern (~10% and ~7% of MED13L and RERE ChIP-seq targets correspond to spliceosome components or splicing regulators, respectively). Considering the role of CA microsatellites in splicing regulation, the enrichment of MED13L and RERE binding in these repeat elements could explain, in part, the regulation of the splicing machinery by MED13L and RERE. To address this issue, experiments to document changes in the AS pattern using different lengths of CA repeats in spliceosome genes bound by MED13L and RERE should be performed.

3. Chromatin regions enriched in MED13L, RERE and EWSR1-FLI1

It is well established that EWSR1-FLI1 binding sites correspond to enhancer, super-enhancer and promoter regions enriched in single or repeat GGAA motifs [85, 86]. DNA-loops formed through Mediator connect the cis-regulatory regions and their target promoters and require the interaction with components of the cohesin complex such as STAG2 [80]. Adane *et al.* [146] and Surdez *et al*. [147] described that STAG2 loss of function, which is present in 15 to 20% of patients [31, 34, 37], induces a chromatin rewiring by promoting anchored chromatin extrusion at CTCF loop boundaries. These STAG2 knockout studies in Ewing sarcoma cell lines showed how enhancer-promoter interactions were disrupted, inducing a different EWSR1-FLI1 transcriptome that prompted metastatic features. Our ChIP-seq experiments revealed EWSR1-FLI1 overlapping in discrete regions with MED13L, RERE and STAG2. In contrast, ~35% of genes with a peak associated with MED13L, RERE, or STAG2 were shared with a gene associated to an EWSR1-FLI1 peak. Furthermore, immunoprecipitation and immunofluorescence analyses showed that EWSR1-FLI1 interacts with STAG2 and MED13L and co-localizes in the same regions. In addition, overlapping co-localization of RERE and STAG2 could be detected in the nuclei of A673 cells. Taken together, our results strongly suggest the participation of MED13L, RERE and STAG2 in the EWSR1-FLI1 regulatory elements and their target promoters in Ewing sarcoma. To test this possibility, 3D chromatin conformation experiments are planned to

decipher the role of MED13L and RERE in the chromatin architecture of Ewing sarcoma and their relationship with STAG2.

To improve our understanding of the role of MED13L and RERE in Ewing sarcoma, we examined the discrete regions of overlap between EWSR1-FLI1 and MED13L and RERE. Our data indicate that these regions correspond to CpG islands and most of them display peaks in tandem and encode ncRNAs. Annotation and mapping of tandemly repeated genomic elements, such as rRNAs or alpha satellites, by high-throughput sequencing methods are challenging tasks. These regions denominated "dark" or "black" may harbor signal enrichments due to genomic amplification, which could result in inaccurate data analysis [166]. However, the tandem regions detected in our analysis did not correspond to black regions annotated in the blacklist of the ENCODE project, except for a few regions of ribosomal subunits. These results indicate that genomic regions enriched with overlapping EWSR1-FLI1, MED13L and RERE were accurately determined.

CpG islands are usually devoid of DNA methylation, constituting transcriptionally active regions in promoters [64]. In vertebrate cells, approximately 70% of the annotated coding gene promoters are in these guanine- and cytosine-rich sites, making these regions sites of transcriptional initiation. Apart from this, CpG islands are also found in distant regions from annotated transcription start sites. For example, transcription of the ncRNA *AIRr* starts in an intragenic CpG island. In the same way, the non-coding transcript KCNQ1OT1 is transcribed in a CpG island found in intron 10 of *KCNQ1*. Therefore, CpG islands can be found initiating the transcription of ncRNA [64]. In agreement with this, binding of EWSR1-FLI1, MED13L and RERE in CpG-rich sites could be potentially

activating the expression of important ncRNAs for tumorigenesis. This possibility is supported by the enrichment in H3K27ac of these regions, which would be associated with their transcriptional activation. In this regard, our results point to an EWSR1-FLI1-orchestrated regulation of non-coding regions that involves Mediator and RERE. Since we performed RNA-seq after MED13L and RERE depletion using PolyA capture methodology, which allows the detection of coding mRNAs but not ncRNAs, we do not know whether MED13L or RERE affect their expression levels. ChIP detection of RNA polymerase I, II and III in these regions and sequencing of bulk RNA might be useful to better characterize the transcriptional changes of non-protein coding genes upon MED13L and RERE depletion.

Some of the non-coding protein genomic regions characterized by tandem peaks of Mediator, RERE, STAG2 and the oncogene include the tRNA loci and the genomic regions coding the ribosomal subunits RNA5S and RNA5-8S. Control of translation is the last layer of regulation of gene expression. Since proteins are crucial for all biological functions such as cell homeostasis, survival or differentiation, their synthesis must be fine-tuned to obtain the accurate levels in the precise context. The number of active ribosomes depends on the number of mRNAs, the mRNA coding length, the rate of ribosome attachment to the mRNA and the rate of elongation. All these processes will determine the protein synthesis rate and can be modified according to the cell needs [167]. Cancer cells take advantage of this regulatory mechanism to increase their aberrant proliferation. Therefore, the levels of the mRNAs used as template for protein synthesis, tRNAs, ribosomes, microRNAs and mRNA interacting proteins can be altered in cancer. For instance, altered levels of translation initiation factors such as eIF3 and eIF4E have been reported in lung and prostate cancer,

respectively [168]. Dysregulation of ribosome biogenesis in cancer not only increases protein levels required to meet high metabolic demands, but also facilitates cancer progression by maintaining stem-cell features. Indeed, increased rRNA production by the MYC oncogene is necessary for tumor progression in mouse models of B-cell tumors induced by this oncogene [169]. In addition, augmented pre-rRNA expression was observed in alveolar rhabdomyosarcoma [170]. In relation to the Mediator complex, it has been reported that the MED17 subunit (of the head module) interacts with RNA polymerases pol I and III, which transcribe ribosome subunits and tRNAs [72]. Interestingly, CHD7, the protein whose deficiency causes a RERE-like phenotype, is found at rRNA transcription sites in the nucleoli in the colorectal cancer cell line DLD1. Furthermore, this chromatin remodeler binds to demethylated regions of rRNA chromatin, indicating its direct regulation of rRNA transcription [171]. These results, together with our ChIP-seq data and the observed downregulation of genes involved in the ribonucleoprotein complex and ribosome biogenesis after MED13L depletion, demonstrate that MED13L and RERE are contributing to the translational rewiring that confers an advantage for Ewing sarcoma tumorigenesis.

The most significant binding motif of EWSR1-FLI1 is the GGAA microsatellite repeats [92]. However, the MED13L regions that overlap with the oncogene showed enrichment in IRF7 and FLI1 binding motifs. In the case of RERE, the binding sites shared with EWSR1-FLI1 were enriched in SP2 and SP3 binding motifs. Notably, IRF7, SP2 and SP3 were identified as part of the Ewing sarcoma regulatory network according to two expression arrays including 117 and 85 patient samples, respectively [172]. This variability in the binding motifs enriched in regions shared with the

oncogene and MED13L and RERE prompts us to hypothesize that Mediator and RERE would have functions at these discrete genomic sites other than transcriptional activation. However, the involvement of STAG2 in these regions needs to be clarified, as it also binds to regions enriched in EWSR1-FLI1, MED13L and RERE.

4. Functional effects induced by MED13L and RERE

In this work, siRNA and shRNA methodologies were employed to analyze the functional effects of MED13L and RERE. Even though protein expression was reduced after MED13L-induced knockdown, mRNA was not altered. Our results suggest that MED13L shRNA may be preventing translation by inhibiting mRNA entry into the ribosome. In addition, as cells were collected 72 hours after knockdown induction, mRNA levels might be restored to increase transcription due to the importance of the gene for cell function. Despite numerous attempts, *MED13L* and *RERE* knockouts could not be achieved in the A673 cell line using CRISPR-Cas9. Therefore, we were unable to validate in MED13L and RERE knockout cells the functional effects identified by shRNA and siRNA technologies. This failure might be attributed to technical limitations or to the possibility that these genes are essential for cell viability. Our ChIP-seq data show that most Mediator subunits overlap in the same genomic regions, except for MED12 binding, which exhibits a more extensive chromatin distribution in addition to the regions shared with MED13L. Although only MED13L and RERE are

regulated by EWSR1-FLI1, our data suggest a redundancy of Mediator subunits in bound chromatin regions, making the dependence on MED13L and RERE for cell viability in Ewing sarcoma unlikely. Therefore, new approaches with alternative methodologies to generate knockouts are needed.

MED13L and RERE depletion by shRNA and siRNA technologies resulted in impaired expression of proteins involved in RNA metabolism, and specifically of proteins of the alternative splicing machinery. Alternative splicing alterations have been described in many types of cancer. Different cancer-related mechanisms, such as nonsense-mediated mRNA decay, non-coding RNAs or aberrant post-translational modifications can alter splicing. Therefore, they cause modifications in the function or localization of splicing factors, alterations in the downstream targets of the splicing factors and changes in mRNA isoforms that will facilitate tumor initiation and progression. Indeed, transcription of the splicing factors can be modified in cancer cells [107]. It is well known that EWSR1-FLI1 plays a role in alternative splicing by interacting with the splicing machinery, and regulates the splicing of many genes that are important for tumorigenesis[29, 114, 115]. Mediator complex is also involved in the regulation of alternative splicing in other cellular contexts [71, 72]. Our data show that EWSR1-FLI1, Mediator and RERE bind to tandemly arrayed chromatin regions corresponding to the leading splicing factor U2, and EWSR1-FLI1 regulates U2 expression. We have shown that the expression of spliceosome components and regulators is deregulated by MED13L and RERE depletion, leading to the splicing pattern in the A673 cell line. Importantly, our bioinformatics analysis revealed that the alternative splicing profile induced by EWSR1-FLI1 differed from that regulated by

MED13L and RERE. Overall, our data suggest that the oncogene elicits a broader regulation of alternative splicing and, by regulating MED13L and RERE expression, specifically controls one type of splicing in specific subsets of genes. Therefore, in accordance with other groups that have previously suggested the presence of transcription factors that could be managing the utilization of specific splice sites [38, 116], we propose MED13L and RERE as mediators of some of the aberrant alternative splicing events observed in Ewing sarcoma.

5. Alternative splicing-related vulnerabilities in Ewing sarcoma

Sustained expression of genes regulated by super-enhancers is relevant to maintain essential cell functions that constitute potential therapeutic targets. In this regard, it is difficult to translate the treatment of specific cancer-related splicing isoforms by RNAi-based approaches to the clinic, whereas the treatment of the splicing machinery has emerged as a potential new therapeutic strategy. Considering the altered splicing patterns observed in many cancers, many different small molecules have been developed to block the function of specific splicing proteins. Most of them target SF3B1, since this protein complex is involved in each step of the way[173].

To assess the vulnerability conferred by aberrant alternative splicing in Ewing sarcoma conferred by both the oncogene and MED13L and RERE,

we used the splicing inhibitor Pladienolide B. This inhibitor interferes with the interaction of SF3B1 with the pre-mRNA, while the synthetic Pladienolide B analog E7107 inhibits the ATP-dependent RNA remodeling. Both inhibitors alter the recognition of BPS by U2 snRNA, causing a different splicing pattern [173]. Pladienolide B has been used *in vitro* and *in vivo* in gastric and cervical cancers and in erythroleukemia [152] and, in human cervical carcinoma cells, Pladienolide B induced cell cycle arrest, apoptosis and p73 splicing [153]. Our work demonstrated the functional relevance of splicing regulation in Ewing sarcoma by MED13L and RERE with Pladienolide B treatment, as Ewing sarcoma cell lines were more sensitive to Pladienolide B compared to any other cell lines tested. Furthermore, depletion of MED13L and RERE renders the cells less sensitive to the inhibitor. Our results support those of Grohar *et al.* [174] and Laurence *et al.*[180], who demonstrated the higher sensitivity of Ewing sarcoma cell lines to Pladienolide B. Grohar *et al.* [174] observed that disruption of the spliceosome machinery led to the formation of differently spliced EWSR1-FLI1 transcripts that downregulated the expression of classical oncogene target genes. However, they did not inquire into the different splicing isoforms of other genes that the inhibitor could be hampering to prevent expression of the target genes and tumorigenesis. In this respect, a transcriptome analysis to determine splicing alterations in MED13L and RERE knockdown cells after Pladienolide B treatment could be useful to elucidate changes in mRNA isoforms that impair cell viability.

The ocular side effects observed in phase I clinical trials with Pladienolide B derivatives [152, 154, 155] encouraged us to consider combination treatments with Pladienolide B. Lawrence *et al.* [175] reported that Pladienolide B and its derivative E7107 are the splicing inhibitors to which Ewing sarcoma cell

lines show the greatest sensitivity, and found a synergistic effect of E7107 in combination with vincristine which induced enhanced apoptosis and G1 phase accumulation compared to control cells. Since we have observed the involvement of the Mediator kinase module subunit MED3L in splicing regulation in Ewing sarcoma, we used BRD6989 to inhibit Mediator kinase CDK8 in combination with Pladienolide B. BRD6989 is the analog of cortistatin A, a natural compound that was described to inhibit Mediator kinases causing the up-regulation of super-enhancer-associated genes involved in tumor suppressor and lineage functions in acute myeloid leukemia (AML) such as *IRF8* or *ETV6* [156]. Our findings indicate that Ewing sarcoma cell lines were more sensitive to the combination, which was synergistic in the A673 cell line. To further validate the combination of Pladienolide B and BRD6989 as a promising new therapeutic strategy in Ewing sarcoma, i*n vivo* experiments would be necessary to test the effect on tumor volume after single and combined treatments.

6. Thesis main points

In summary, in this project we have identified MED13L and RERE as direct EWSR1-FLI1 target genes, which are differentially expressed in Ewing sarcoma. We have demonstrated that the oncogene directly interacts with Mediator subunits, and that the oncogene recruits specifically Mediator, RERE and STAG2 to specific genomic regions. Besides these chromatin regions shared with the oncogene, MED13L, RERE and STAG2 are enriched in discrete genomic regions that correspond to CpG islands where they are disposed as tandems in non-coding regions. MED13L and RERE depletion

results in altered splicing and translational rewiring. In addition to the role on splicing regulation, our data suggest that MED13L and RERE would be contributing to the DNA-loop formation between neo-enhancers and super-enhancers and the promoter of target genes. In these regions, where EWSR1-FLI1 is bound to GGAA repeat elements, Mediator would facilitate the acquisition of an Ewing sarcoma transcriptome. The role of Mediator in Ewing sarcoma would likely be extended to the establishment of a particular chromatin conformation, since STAG2 is also present in MED13L/RERE-enriched regions. Our future work will focus on elucidate Mediator and RERE implications in Ewing sarcoma chromatin architecture. Taken together, our results unveil a new mechanism used by EWSR1-FLI1 to promote tumorigenesis.

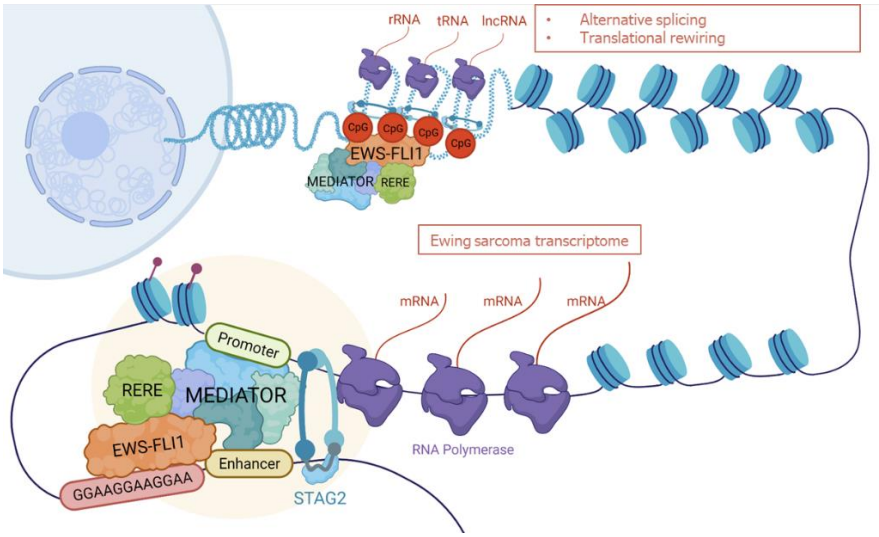


Figure 47. **Mediator and RERE functions in Ewing sarcoma.** Visual representation of the functions described for Mediator and RERE in this work. EWSR1-FLI1 chromatin rewiring requires Mediator, RERE and STAG2 to form dynamic loops between the regulatory regions and the target genes that form Ewing sarcoma transcriptome. In addition, our project also describes the recruitment by EWSR1-FLI1 of Mediator, RERE and STAG2 to non-coding regions that would result in alterations in alternative splicing and translation control. Created with Biorender.com.

CONCLUSIONS

The work compiled in this thesis demonstrates the role of super-enhancer-regulated genes MED13L and RERE in Ewing sarcoma while uncovering a new mechanism involved in EWSR1-FLI1-driven tumorigenesis. Thus, we conclude that:

1. MED13L and RERE expression is under the control of EWSR1-FLI1-regulated super-enhancers. Both genes are differentially expressed in Ewing sarcoma.

2. RERE would interact with the Mediator complex containing MED13L in the same chromatin regions.

3. EWSR1-FLI1 binds to MED13L, MED12 and STAG2, while RERE, STAG2 and the oncogene colocalize in A673 cells, indicating that Mediator, STAG2 and the oncogene interact in discrete nuclear domains.

4. Mediator, RERE and STAG2 would contribute to the formation of DNA-loops between neo-(super)-enhancers and promoters of EWSR1-FLI1 target genes in Ewing sarcoma.

5. Mediator, RERE and STAG2 bind to discrete genomic regions that overlap with EWSR1-FLI1. These regions are CpG islands where peaks are in tandem and correspond to noncoding genes mainly involved in ribosomal functions.

6. The MED13L and RERE binding in these CpG-enriched regions is mediated by EWSR1-FLI1 and the enrichment is specific to Ewing sarcoma.

7. Depletion of both MED13L and RERE deregulates the same subsets of genes involved in alternative splicing regulation in Ewing sarcoma.

8. MED13L and RERE depletion promotes the inclusion of long exons in genes related to chromatin organization and histone modifications through indirect regulation of spliceosome genes. EWSR1-FLI1 downregulation causes a broader effect on alternative splicing.

9. Ewing Sarcoma cell lines are more sensitive than other tumor cell lines to the splicing inhibitor Pladienolide B and MED13L and RERE depletion reduces this sensitivity.

10. The combination of Pladienolide B and the CDK8 inhibitor BRD6989 has a synergistic effect on the Ewing Sarcoma cell line A673, suggesting a potential new therapeutic approach.

BIBLIOGRAPHY

BIBLIOGRAPHY

1. World Health Organization. (2024) *Cancer*. Retrieved from https://www.who.int/health-topics/cancer#tab=tab_1

2. Mora, J. (2012). What is a pediatric tumor? *Clinical Oncology in Adolescents and Young Adults*, *2012*(2), 7-15.

3. World Health Organization. (2021) *Childhood cancer*. Retrieved from https://www.who.int/news-room/fact-sheets/detail/cancer-in-children

4. National Cancer Institute. (2024) *Rare Cancers of Childhood Treatment (PDQ®)–Health Professional Version*. Retrieved from https://www.cancer.gov/types/childhood-cancers/hp/rare-childhood-cancers-pdq#_105

5. SIOP Europe. (2024) *Rare Diseases*. Retrieved from https://siope.eu/activities/european-advocacy/rare-diseases/

6. Behjati, S., Gilbertson, R. J., & Pfister, S. M. (2021). Maturation block in childhood cancer. *Cancer Discovery, 11*(3), 542–544.

7. Karlberg, J. (1989). A Biologically-Oriented Mathematical Model (ICP) for Human Growth. *Acta Paediatrica Scandinavica (Supplement), 350*, 70-94.

8. Greaves, M.F., & Alexander, F.E. (1993) An infectious etiology for common acute lymphoblastic leukemia in childhood? *Leukemia,* 7(3), 349–360.

9. Scotting, P. J., Walker, D. A., & Perilongo, G. (2005). Childhood solid tumours: a developmental disorder. *Nature Reviews Cancer*, *5*(6), 481-8.

10. Rivera, M. N., & Haber, D. A. (2005). Wilms' tumour: Connecting tumorigenesis and organ development in the kidney. *Nature Reviews Cancer*, *5*(9), 699–712.

11. De Preter, K., Vandesompele, J., Heimann, P., Yigit, N., Beckman, S., Schramm, A., Eggert, A., Stallings, R. L., Benoit, Y., Renard, M., de

Paepe, A., Laureys, G., Påhlman, S., & Speleman, F. (2006). Human fetal neuroblast and neuroblastoma transcriptome analysis confirms neuroblast origin and highlights neuroblastoma candidate genes. *Genome Biology, 7*(9), 401.

12. Gibson, P., Tong, Y., Robinson, G., Thompson, M. C., Currle, D. S., Eden, C., Kranenburg, T. A., Hogg, T., Poppleton, H., Martin, J., Finkelstein, D., Pounds, S., Weiss, A., Patay, Z., Scoggins, M., Ogg, R., Pei, Y., Yang, Z. J., Brun, S., Lee, Y., … Gilbertson, R. J. (2010). Subtypes of medulloblastoma have distinct developmental origins. *Nature*, *468*(7327), 1095–1099.
13. Amatruda J. F. (2021). Modeling the developmental origins of pediatric cancer to improve patient outcomes. *Disease models & mechanisms*, *14*(2).
14. Filbin, M., & Monje, M. (2019). Developmental origins and emerging therapeutic opportunities for childhood cancer. *Nature medicine*, *25*(3), 367–376.
15. Kandoth, C., McLellan, M. D., Vandin, F., Ye, K., Niu, B., Lu, C., Xie, M., Zhang, Q., McMichael, J. F., Wyczalkowski, M. A., Leiserson, M. D. M., Miller, C. A., Welch, J. S., Walter, M. J., Wendl, M. C., Ley, T. J., Wilson, R. K., Raphael, B. J., & Ding, L. (2013). Mutational landscape and significance across 12 major cancer types. *Nature*, *502*(7471), 333–339.
16. Alexandrov, L. B., Kim, J., Haradhvala, N. J., Huang, M. N., Tian Ng, A. W., Wu, Y., Boot, A., Covington, K. R., Gordenin, D. A., Bergstrom, E. N., Islam, S. M. A., Lopez-Bigas, N., Klimczak, L. J., McPherson, J. R., Morganella, S., Sabarinathan, R., Wheeler, D. A., Mustonen, V., PCAWG Mutational Signatures Working Group, Getz, G., … PCAWG Consortium (2020). The repertoire of mutational signatures in human cancer. *Nature*, *578*(7793), 94–101.
17. Monje, M. (2018). Open questions: why are babies rarely born with cancer? *BMC biology*, *16*(1), 129.

18. Gröbner, S. N., Worst, B. C., Weischenfeldt, J., Buchhalter, I., Kleinheinz, K., Rudneva, V. A., Johann, P. D., Balasubramanian, G. P., Segura-Wang, M., Brabetz, S., Bender, S., Hutter, B., Sturm, D., Pfaff, E., Hübschmann, D., Zipprich, G., Heinold, M., Eils, J., Lawerenz, C., Erkek, S., … ICGC MMML-Seq Project (2018). The landscape of genomic alterations across childhood cancers. *Nature*, *559*(7714), E10.

19. Sinkala, M. (2023). Mutational landscape of cancer-driver genes across human cancers. *Scientific reports*, *13*(1), 12742.

20. Wu, G., Broniscer, A., McEachron, T. A., Lu, C., Paugh, B. S., Becksfort, J., Qu, C., Ding, L., Huether, R., Parker, M., Zhang, J., Gajjar, A., Dyer, M. A., Mullighan, C. G., Gilbertson, R. J., Mardis, E. R., Wilson, R. K., Downing, J. R., Ellison, D. W., Zhang, J., … St. Jude Children's Research Hospital–Washington University Pediatric Cancer Genome Project (2012). Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas. *Nature genetics*, *44*(3), 251–253.

21. Bayliss, J., Mukherjee, P., Lu, C., Jain, S. U., Chung, C., Martinez, D., Sabari, B., Margol, A. S., Panwalkar, P., Parolia, A., Pekmezci, M., McEachin, R. C., Cieslik, M., Tamrazi, B., Garcia, B. A., La Rocca, G., Santi, M., Lewis, P. W., Hawkins, C., Melnick, A., … Venneti, S. (2016). Lowered H3K27me3 and DNA hypomethylation define poorly prognostic pediatric posterior fossa ependymomas. *Science translational medicine*, *8*(366), 366ra161.

22. Zhang, J., Walsh, M. F., Wu, G., Edmonson, M. N., Gruber, T. A., Easton, J., Hedges, D., Ma, X., Zhou, X., Yergeau, D. A., Wilkinson, M. R., Vadodaria, B., Chen, X., McGee, R. B., Hines-Dowell, S., Nuccio, R., Quinn, E., Shurtleff, S. A., Rusch, M., Patel, A., … Downing, J. R. (2015). Germline Mutations in Predisposition Genes in Pediatric Cancer. *The New England journal of medicine*, *373*(24), 2336–2346.

23. Taylor, B. S., Barretina, J., Maki, R. G., Antonescu, C. R., Singer, S., & Ladanyi, M. (2011). Advances in sarcoma genomics and new therapeutic targets. *Nature reviews*. Cancer, *11*(8), 541–557.
24. Riggi, N., & Stamenkovic, I. (2007). The Biology of Ewing sarcoma. *Cancer letters*, *254*(1), 1–10.
25. Burningham, Z., Hashibe, M., Spector, L., & Schiffman, J. D. (2012). The epidemiology of sarcoma. *Clinical sarcoma research*, *2*(1), 14.
26. Giannikopoulos, P., & Parham, D. M. (2022). Pediatric Sarcomas: The Next Generation of Molecular Studies. *Cancers, 14*(10), 2515.
27. Anderson, J. L., Denny, C. T., Tap, W. D., & Federman, N. (2012). Pediatric sarcomas: translating molecular pathogenesis of disease to novel therapeutic possibilities. *Pediatric research*, *72*(2), 112–121.
28. Suvà, M. L., Cironi, L., Riggi, N., & Stamenkovic, I. (2007). Sarcomas: genetics, signalling, and cellular origins. Part 2: TET-independent fusion proteins and receptor tyrosine kinase mutations. *The Journal of pathology*, *213*(2), 117–130.
29. Riggi, N., Suvà, M. L., & Stamenkovic, I. (2021). Ewing's Sarcoma. *The New England journal of medicine*, *384*(2), 154–164.
30. Mora, J., Castañeda, A., Perez-Jaume, S., Lopez-Pousa, A., Maradiegue, E., Valverde, C., Martin-Broto, J., Garcia Del Muro, X., Cruz, O., Cruz, J., Martinez-Trufero, J., Maurel, J., Vaz, M. A., de Alava, E., & de Torres, C. (2017). GEIS-21: a multicentric phase II study of intensive chemotherapy including gemcitabine and docetaxel for the treatment of Ewing sarcoma of children and adults: a report from the Spanish sarcoma group (GEIS). *British journal of cancer*, *117*(6), 767–774.
31. Grünewald, T. G. P., Cidre-Aranaz, F., Surdez, D., Tomazou, E. M., de Álava, E., Kovar, H., Sorensen, P. H., Delattre, O., & Dirksen, U. (2018). Ewing sarcoma. *Nature reviews. Disease primers, 4*(1), 5.
32. Spector, L. G., Hubbard, A. K., Diessner, B. J., Machiela, M. J., Webber, B. R., & Schiffman, J. D. (2021). Comparative international incidence of

Ewing sarcoma 1988 to 2012. *International journal of cancer, 149*(5), 1054–1066.

33. Reed, D. R., Grohar, P., Rubin, E., Binitie, O., Krailo, M., Davis, J., DuBois, S. G., & Janeway, K. A. (2023). Children's Oncology Group's 2023 blueprint for research: Bone tumors. *Pediatric blood & cancer, 70*(Suppl 6), e30583.
34. PDQ Pediatric Treatment Editorial Board. (2002) *Ewing Sarcoma and Undifferentiated Small Round Cell Sarcomas of Bone and Soft Tissue Treatment (PDQ®): Health Professional Version.* Retrieved from: https://www.ncbi.nlm.nih.gov/books/NBK66045/
35. Setty, B. A., Gikandi, A., & DuBois, S. G. (2023). Ewing Sarcoma Drug Therapy: Current Standard of Care and Emerging Agents. *Paediatric drugs, 25*(4), 389–397.
36. Sbaraglia, M., Righi, A., Gambarotti, M., & Dei Tos, A. P. (2020). Ewing sarcoma and Ewing-like tumors. *Virchows Archiv : an international journal of pathology, 476*(1), 109–119.
37. 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., Thorner, A. R., de Torres, C., Lavarino, C., Suñol, M., McKenna, A., Sivachenko, A., Cibulskis, K., Lawrence, M. S., Stojanov, P., Rosenberg, M., … Stegmaier, K. (2014). The genomic landscape of pediatric Ewing sarcoma. *Cancer discovery, 4*(11), 1326–1341.
38. Kovar, H. (2005). Context matters: the hen or egg problem in Ewing's sarcoma. *Seminars in cancer biology, 15*(3), 189–196.
39. Tirode, F., Laud-Duval, K., Prieur, A., Delorme, B., Charbord, P., & Delattre, O. (2007). Mesenchymal stem cell features of Ewing tumors. *Cancer cell, 11*(5), 421–429.
40. Hancock, J. D., & Lessnick, S. L. (2008). A transcriptional profiling meta-analysis reveals a core EWS-FLI gene expression signature. *Cell cycle, 7*(2), 250–256.

41. Etchevers, H. C., Dupin, E., & Le Douarin, N. M. (2019). The diverse neural crest: from embryology to human pathology. *Development, 146*(5)

42. Staege, M. S., Hutter, C., Neumann, I., Foja, S., Hattenhorst, U. E., Hansen, G., Afar, D., & Burdach, S. E. (2004). DNA microarrays reveal relationship of Ewing family tumors to both endothelial and fetal neural crest-derived cells and define novel targets. *Cancer research, 64*(22), 8213–8221.

43. Minas, T. Z., Surdez, D., Javaheri, T., Tanaka, M., Howarth, M., Kang, H. J., Han, J., Han, Z. Y., Sax, B., Kream, B. E., Hong, S. H., Çelik, H., Tirode, F., Tuckermann, J., Toretsky, J. A., Kenner, L., Kovar, H., Lee, S., Sweet-Cordero, E. A., Nakamura, T., … Üren, A. (2017). Combined experience of six independent laboratories attempting to create an Ewing sarcoma mouse model. *Oncotarget, 8*(21), 34141–34163.

44. Leacock, S. W., Basse, A. N., Chandler, G. L., Kirk, A. M., Rakheja, D., & Amatruda, J. F. (2012). A zebrafish transgenic model of Ewing's sarcoma reveals conserved mediators of EWS-FLI1 tumorigenesis. *Disease models & mechanisms, 5*(1), 95–106.

45. Molnar, C., Reina, J., Herrero, A., Heinen, J. P., Méndiz, V., Bonnal, S., Irimia, M., Sánchez-Jiménez, M., Sánchez-Molina, S., Mora, J., & Gonzalez, C. (2022). Human EWS-FLI protein recapitulates in Drosophila the neomorphic functions that induce Ewing sarcoma tumorigenesis. *PNAS nexus, 1*(4), 222.

46. Castillero-Trejo, Y., Eliazer, S., Xiang, L., Richardson, J. A., & Ilaria, R. L. Jr. (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 research, 65*(19), 8698–8705.

47. Riggi, N., Suvà, M. L., Suvà, D., Cironi, L., Provero, P., Tercier, S., Joseph, J. M., Stehle, J. C., Baumer, K., Kindler, V., & Stamenkovic, I. (2008). EWS-FLI-1 expression triggers a Ewing's sarcoma initiation program in

primary human mesenchymal stem cells. *Cancer research, 68*(7), 2176–2185.

48. Riggi, N., Suvà, M. L., De Vito, C., Provero, P., Stehle, J. C., Baumer, K., Cironi, L., Janiszewska, M., Petricevic, T., Suvà, D., Tercier, S., Joseph, J. M., Guillou, L., & Stamenkovic, I. (2010). EWS-FLI-1 modulates miRNA145 and SOX2 expression to initiate mesenchymal stem cell reprogramming toward Ewing sarcoma cancer stem cells. *Genes & development, 24*(9), 916–932.
49. Tanaka, M., Yamazaki, Y., Kanno, Y., Igarashi, K., Aisaki, K., Kanno, J., & Nakamura, T. (2014). Ewing's sarcoma precursors are highly enriched in embryonic osteochondrogenic progenitors. *The Journal of clinical investigation, 124*(7), 3061–3074.
50. Gordon, D. J., Motwani, M., & Pellman, D. (2016). Modeling the initiation of Ewing sarcoma tumorigenesis in differentiating human embryonic stem cells. *Oncogene, 35*(24), 3092–3102.
51. Gopinathan, G., & Diekwisch, T. G. H. (2022). Epigenetics and Early Development. *Journal of developmental biology, 10*(2), 26.
52. Skinner, M. K. (2011). Role of epigenetics in developmental biology and transgenerational inheritance. *Birth defects research. Part C, Embryo today: reviews, 93*(1), 51–55.
53. Cooper GM. (2000) *The Cell: A Molecular Approach.* Retrieved from: https://www.ncbi.nlm.nih.gov/books/NBK9863/
54. Mariño-Ramírez, L., Kann, M. G., Shoemaker, B. A., & Landsman, D. (2005). Histone structure and nucleosome stability. *Expert review of proteomics, 2*(5), 719–729.
55. Fyodorov, D. V., Zhou, B. R., Skoultchi, A. I., & Bai, Y. (2018). Emerging roles of linker histones in regulating chromatin structure and function. *Nature reviews. Molecular cell biology, 19*(3), 192–206.
56. Portillo-Ledesma, S., & Schlick, T. (2024). Regulation of chromatin architecture by protein binding: insights from molecular modeling. *Biophysical reviews, 16*(3), 331–343.

57. Moosavi, A., & Motevalizadeh Ardekani, A. (2016). Role of Epigenetics in Biology and Human Diseases. *Iranian biomedical journal, 20*(5), 246–258.
58. Goldberg, A. D., Allis, C. D., & Bernstein, E. (2007). Epigenetics: a landscape takes shape. *Cell, 128*(4), 635–638.
59. Bonasio, R., Tu, S., & Reinberg, D. (2010). Molecular signals of epigenetic states. *Science, 330*(6004), 612–616.
60. Gibney, E. R., & Nolan, C. M. (2010). Epigenetics and gene expression. *Heredity, 105*(1), 4–13.
61. Chen, P., Li, W., & Li, G. (2021). Structures and Functions of Chromatin Fibers. *Annual review of biophysics*, *50*, 95–116.
62. Klemm, S. L., Shipony, Z., & Greenleaf, W. J. (2019). Chromatin accessibility and the regulatory epigenome. *Nature reviews. Genetics, 20*(4), 207–220.
63. Jin, B., Li, Y., & Robertson, K. D. (2011). DNA methylation: superior or subordinate in the epigenetic hierarchy? *Genes & cancer, 2(*6), 607–617.
64. Deaton, A. M., & Bird, A. (2011). CpG islands and the regulation of transcription. *Genes & development, 25*(10), 1010–1022.
65. Kouzarides, T. (2007). Chromatin modifications and their function. *Cell*, *128*(4), 693–705.
66. Jiang, H. (2020). The complex activities of the SET1/MLL complex core subunits in development and disease. *Biochimica et biophysica acta. Gene regulatory mechanisms*, *1863*(7), 194560.
67. Becker, P. B., & Workman, J. L. (2013). Nucleosome remodeling and epigenetics. *Cold Spring Harbor perspectives in biology*, *5*(9), a017905.
68. Peschansky, V. J., & Wahlestedt, C. (2014). Non-coding RNAs as direct and indirect modulators of epigenetic regulation. *Epigenetics*, *9*(1), 3–12.
69. Cramer, P. (2019). Organization and regulation of gene transcription. *Nature*, *573*(7772), 45–54.

70. Pott, S., & Lieb, J. D. (2015). What are super-enhancers? *Nature genetics, 47*(1), 8–12.

71. Yin, J. W., & Wang, G. (2014). The Mediator complex: a master coordinator of transcription and cell lineage development. *Development, 141*(5), 977–987.

72. Putlyaev, E. V., Ibragimov, A. N., Lebedeva, L. A., Georgiev, P. G., & Shidlovskii, Y. V. (2018). Structure and Functions of the Mediator Complex. *Biochemistry. Biokhimiia*, *83*(4), 423–436.

73. Whyte, W. A., Orlando, D. A., Hnisz, D., Abraham, B. J., Lin, C. Y., Kagey, M. H., Rahl, P. B., Lee, T. I., & Young, R. A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell, 153*(2), 307–319.

74. Hnisz, D., Abraham, B. J., Lee, T. I., Lau, A., Saint-André, V., Sigova, A. A., Hoke, H. A., & Young, R. A. (2013). Super-enhancers in the control of cell identity and disease. *Cell, 155*(4), 934–947.

75. Malik, S., & Roeder, R. G. (2010). The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. *Nature reviews. Genetics*, *11*(11), 761–772.

76. Conaway, R. C., & Conaway, J. W. (2011). Function and regulation of the Mediator complex. *Current opinion in genetics & development*, *21*(2), 225–230.

77. Soutourina J. (2018). Transcription regulation by the Mediator complex. *Nature reviews. Molecular cell biology*, *19*(4), 262–274.

78. Weber, H., & Garabedian, M. J. (2018). The mediator complex in genomic and non-genomic signaling in cancer. *Steroids*, *133*, 8–14.

79. Harper, T. M., & Taatjes, D. J. (2018). The complex structure and function of Mediator. *The Journal of biological chemistry*, *293*(36), 13778–13785.

80. Kagey, M. H., Newman, J. J., Bilodeau, S., Zhan, Y., Orlando, D. A., van Berkum, N. L., Ebmeier, C. C., Goossens, J., Rahl, P. B., Levine, S. S., Taatjes, D. J., Dekker, J., & Young, R. A. (2010). Mediator and cohesin

connect gene expression and chromatin architecture. *Nature*, *467*(7314), 430–435.

81. Papadopoulou, T., Kaymak, A., Sayols, S., & Richly, H. (2016). Dual role of Med12 in PRC1-dependent gene repression and ncRNA-mediated transcriptional activation. *Cell cycle*, *15*(11), 1479–1493.
82. Taatjes, D. J. (2010). The human Mediator complex: a versatile, genome-wide regulator of transcription. *Trends in biochemical sciences*, *35*(6), 315–322.
83. Aranda-Orgilles, B., Saldaña-Meyer, R., Wang, E., Trompouki, E., Fassl, A., Lau, S., Mullenders, J., Rocha, P. P., Raviram, R., Guillamot, M., Sánchez-Díaz, M., Wang, K., Kayembe, C., Zhang, N., Amoasii, L., Choudhuri, A., Skok, J. A., Schober, M., Reinberg, D., Sicinski, P., … Aifantis, I. (2016). MED12 Regulates HSC-Specific Enhancers Independently of Mediator Kinase Activity to Control Hematopoiesis. *Cell stem cell, 19*(6), 784–799.
84. Lessnick, S. L., Braun, B. S., Denny, C. T., & May, W. A. (1995). Multiple domains mediate transformation by the Ewing's sarcoma EWS/FLI-1 fusion gene. *Oncogene, 10*(3), 423–431.
85. 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., Formey, A., Patel, A., Gymrek, M., Thapar, V., Deshpande, V., Ting, D. T., Hornicek, F. J., Nielsen, G. P., Stamenkovic, I., Aryee, M. J., … Rivera, M. N. (2014). EWS-FLI1 utilizes divergent chromatin remodeling mechanisms to directly activate or repress enhancer elements in Ewing sarcoma. *Cancer cell, 26*(5), 668–681.
86. Tomazou, E. M., Sheffield, N. C., Schmidl, C., Schuster, M., Schönegger, A., Datlinger, P., Kubicek, S., Bock, C., & Kovar, H. (2015). Epigenome mapping reveals distinct modes of gene regulation and widespread enhancer reprogramming by the oncogenic fusion protein EWS-FLI1. *Cell reports, 10*(7), 1082–1095.

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

88. Reddy, S. Y., Obika, S., & Bruice, T. C. (2003). Conformations and dynamics of Ets-1 ETS domain-DNA complexes. *Proceedings of the National Academy of Sciences of the United States of America, 100*(26), 15475–15480.

89. Mao, X., Miesfeldt, S., Yang, H., Leiden, J. M., & Thompson, C. B. (1994). The FLI-1 and chimeric EWS-FLI-1 oncoproteins display similar DNA binding specificities. *The Journal of biological chemistry, 269*(27), 18216–18222.

90. 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., & Lessnick, S. L. (2008). Microsatellites as EWS/FLI response elements in Ewing's sarcoma. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(29), 10149–10154.

91. Gangwal, K., & Lessnick, S. L. (2008). Microsatellites are EWS/FLI response elements: genomic "junk" is EWS/FLI's treasure. *Cell cycle, 7*(20), 3127–3132.

92. Gangwal, K., Close, D., Enriquez, C. A., Hill, C. P., & Lessnick, S. L. (2010). Emergent Properties of EWS/FLI Regulation via GGAA Microsatellites in Ewing's Sarcoma. *Genes & cancer*, *1*(2), 177–187.

93. Beck, R., Monument, M. J., Watkins, W. S., Smith, R., Boucher, K. M., Schiffman, J. D., Jorde, L. B., Randall, R. L., & Lessnick, S. L. (2012). EWS/FLI-responsive GGAA microsatellites exhibit polymorphic differences between European and African populations. *Cancer genetics*, *205*(6), 304–312.

94. Boulay, G., Sandoval, G. J., Riggi, N., Iyer, S., Buisson, R., Naigles, B., Awad, M. E., Rengarajan, S., Volorio, A., McBride, M. J., Broye, L. C., Zou, L., Stamenkovic, I., Kadoch, C., & Rivera, M. N. (2017). Cancer-

Specific Retargeting of BAF Complexes by a Prion-like Domain. *Cell*, *171*(1), 163–178.e19.

95. Kennedy, A. L., Vallurupalli, M., Chen, L., Crompton, B., Cowley, G., Vazquez, F., Weir, B. A., Tsherniak, A., Parasuraman, S., Kim, S., Alexe, G., & Stegmaier, K. (2015). Functional, chemical genomic, and super-enhancer screening identify sensitivity to cyclin D1/CDK4 pathway inhibition in Ewing sarcoma. *Oncotarget, 6*(30), 30178–30193.
96. Lin, L., Huang, M., Shi, X., Mayakonda, A., Hu, K., Jiang, Y. Y., Guo, X., Chen, L., Pang, B., Doan, N., Said, J. W., Xie, J., Gery, S., Cheng, X., Lin, Z., Li, J., Berman, B. P., Yin, D., Lin, D. C., & Koeffler, H. P. (2019). Super-enhancer-associated MEIS1 promotes transcriptional dysregulation in Ewing sarcoma in co-operation with EWS-FLI1. *Nucleic acids research, 47*(3), 1255–1267.
97. Sheffield, N. C., Pierron, G., Klughammer, J., Datlinger, P., Schönegger, A., Schuster, M., Hadler, J., Surdez, D., Guillemot, D., Lapouble, E., Freneaux, P., Champigneulle, J., Bouvier, R., Walder, D., Ambros, I. M., Hutter, C., Sorz, E., Amaral, A. T., de Álava, E., Schallmoser, K., … Tomazou, E. M. (2017). DNA methylation heterogeneity defines a disease spectrum in Ewing sarcoma. *Nature medicine*, *23*(3), 386–395.
98. Dylla, L., Moore, C., & Jedlicka, P. (2013). MicroRNAs in Ewing Sarcoma. *Frontiers in oncology, 3*, 65.
99. De Vito, C., Riggi, N., Cornaz, S., Suvà, M. L., Baumer, K., Provero, P., & Stamenkovic, I. (2012). A TARBP2-dependent miRNA expression profile underlies cancer stem cell properties and provides candidate therapeutic reagents in Ewing sarcoma. *Cancer cell*, *21*(6), 807–821.
100. Marques Howarth, M., Simpson, D., Ngok, S. P., Nieves, B., Chen, R., Siprashvili, Z., Vaka, D., Breese, M. R., Crompton, B. D., Alexe, G., Hawkins, D. S., Jacobson, D., Brunner, A. L., West, R., Mora, J., Stegmaier, K., Khavari, P., & Sweet-Cordero, E. A. (2014). Long noncoding RNA EWSAT1-mediated gene repression facilitates Ewing

sarcoma oncogenesis. *The Journal of clinical investigation*, *124*(12), 5275–5290.

101. Marasco, L. E., & Kornblihtt, A. R. (2023). The physiology of alternative splicing. *Nature reviews. Molecular cell biology*, *24*(4), 242–254.
102. Smith, C. W., & Valcárcel, J. (2000). Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends in biochemical sciences*, *25*(8), 381–388.
103. Brown, S. J., Stoilov, P., & Xing, Y. (2012). Chromatin and epigenetic regulation of pre-mRNA processing. *Human molecular genetics*, *21*(R1), R90–R96.
104. Wang, Y., Liu, J., Huang, B. O., Xu, Y. M., Li, J., Huang, L. F., Lin, J., Zhang, J., Min, Q. H., Yang, W. M., & Wang, X. Z. (2015). Mechanism of alternative splicing and its regulation. *Biomedical reports*, *3*(2), 152–158.
105. Wright, C. J., Smith, C. W. J., & Jiggins, C. D. (2022). Alternative splicing as a source of phenotypic diversity. *Nature reviews. Genetics, 23*(11), 697–710.
106. Jiang, W., & Chen, L. (2020). Alternative splicing: Human disease and quantitative analysis from high-throughput sequencing. *Computational and structural biotechnology journal, 19*, 183–195.
107. Yang, H., Beutler, B., & Zhang, D. (2022). Emerging roles of spliceosome in cancer and immunity. *Protein & cell, 13*(8), 559–579.
108. Yang, L., Chansky, H. A., & Hickstein, D. D. (2000). EWS.Fli-1 fusion protein interacts with hyperphosphorylated RNA polymerase II and interferes with serine-arginine protein-mediated RNA splicing. *The Journal of biological chemistry, 275*(48), 37612–37618.
109. Zhang, D., Paley, A. J., & Childs, G. (1998). The transcriptional repressor ZFM1 interacts with and modulates the ability of EWS to activate transcription. *The Journal of biological chemistry*, *273*(29), 18086–18091.

110. Knoop, L. L., & Baker, S. J. (2000). The splicing factor U1C represses EWS/FLI-mediated transactivation. *The Journal of biological chemistry*, *275*(32), 24865–24871.

111. Knoop, L. L., & Baker, S. J. (2001). EWS/FLI alters 5'-splice site selection. *The Journal of biological chemistry, 276*(25), 22317–22322.

112. Toretsky, J. A., Erkizan, V., Levenson, A., Abaan, O. D., Parvin, J. D., Cripe, T. P., Rice, A. M., Lee, S. B., & Uren, A. (2006). Oncoprotein EWS-FLI1 activity is enhanced by RNA helicase A. *Cancer research*, *66*(11), 5574–5581.

113. Erkizan, H. V., Schneider, J. A., Sajwan, K., Graham, G. T., Griffin, B., Chasovskikh, S., Youbi, S. E., Kallarakal, A., Chruszcz, M., Padmanabhan, R., Casey, J. L., Üren, A., & Toretsky, J. A. (2015). RNA helicase A activity is inhibited by oncogenic transcription factor EWS-FLI1. *Nucleic acids research, 43*(2), 1069–1080.

114. 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., Wu, C. H., Lawlor, E. R., Üren, A., & Toretsky, J. A. (2015). Oncogenic fusion protein EWS-FLI1 is a network hub that regulates alternative splicing. *Proceedings of the National Academy of Sciences of the United States of America, 112*(11), E1307–E1316.

115. Selvanathan, S. P., Graham, G. T., Grego, A. R., Baker, T. M., Hogg, J. R., Simpson, M., Batish, M., Crompton, B., Stegmaier, K., Tomazou, E. M., Kovar, H., Üren, A., & Toretsky, J. A. (2019). EWS-FLI1 modulated alternative splicing of ARID1A reveals novel oncogenic function through the BAF complex. *Nucleic acids research, 47*(18), 9619–9636.

116. Graham, G. T., Selvanathan, S. P., Zöllner, S. K., Stahl, E., Shlien, A., Caplen, N. J., Üren, A., & Toretsky, J. A. (2022). Comprehensive profiling of mRNA splicing indicates that GC content signals altered cassette exon inclusion in Ewing sarcoma. *NAR cancer*, *4*(1), zcab052.

117. Carrillo, J., García-Aragoncillo, E., Azorín, D., Agra, N., Sastre, A., González-Mediero, I., García-Miguel, P., Pestaña, A., Gallego, S., Segura,

D., & Alonso, J. (2007). Cholecystokinin down-regulation by RNA interference impairs Ewing tumor growth. *Clinical cancer research : an official journal of the American Association for Cancer Research*, *13*(8), 2429–2440.

118. Chou, T. C., & Martin, N. (2005) CompuSyn for Drug Combinations: PC Software and User's Guide. *ComboSyn Inc; Paramus*, (NJ).
119. Yu, G., Wang, L. G., & He, Q. Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*, *31*(14), 2382–2383.
120. Bilke, S., Schwentner, R., Yang, F., Kauer, M., Jug, G., Walker, R. L., Davis, S., Zhu, Y. J., Pineda, M., Meltzer, P. S., & Kovar, H. (2013). Oncogenic ETS fusions deregulate E2F3 target genes in Ewing sarcoma and prostate cancer. *Genome research*, *23*(11), 1797–1809.
121. Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M. J., Amin, V., Whitaker, J. W., Schultz, M. D., Ward, L. D., Sarkar, A., Quon, G., Sandstrom, R. S., Eaton, M. L., Wu, Y. C., … Kellis, M. (2015). Integrative analysis of 111 reference human epigenomes. *Nature*, *518*(7539), 317–330.
122. Machanick, P., & Bailey, T. L. (2011). MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics, 2712*, 1696-1697.
123. Nassar, L. R., Barber, G. P., Benet-Pagès, A., Casper, J., Clawson, H., Diekhans, M., Fischer, C., Gonzalez, J. N., Hinrichs, A. S., Lee, B. T., Lee, C. M., Muthuraman, P., Nguy, B., Pereira, T., Nejad, P., Perez, G., Raney, B. J., Schmelter, D., Speir, M. L., Wick, B. D., … Kent, W. J. (2023). The UCSC Genome Browser database: 2023 update. *Nucleic acids research, 51*(D1), D1188–D1195.
124. Robinson, M. D., & Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome biology*, *11*(3), R25.

125. Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140.

126. Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21.

127. Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923–930.

128. Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research*, *43*(7), e47.

129. Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., & Mesirov, J. P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(43), 15545–15550.

130. Yu, G., Wang, L. G., Han, Y., & He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics: a journal of integrative biology*, *16*(5), 284–287.

131. Gohr, A., Mantica, F., Hermoso-Pulido, A., Tapial, J., Márquez, Y., & Irimia, M. (2022). Computational Analysis of Alternative Splicing Using VAST-TOOLS and the VastDB Framework. *Methods in molecular biology*, *2537*, 97–128.

132. Schafer, S., Miao, K., Benson, C. C., Heinig, M., Cook, S. A., & Hubner, N. (2015). Alternative Splicing Signatures in RNA-seq Data: Percent Spliced in (PSI). *Current protocols in human genetics*, *87*, 11.16.1–11.16.14.

133. Arakelian, E., & Kfoury, F. EECS Department, University of California, Berkeley (2016) *Predicting Percent Spliced In (PSI) in Alternative Splicing Using Deep Networks.* Retrieved from

http://www2.eecs.berkeley.edu/Pubs/TechRpts/2016/EECS-2016-126.pdf

134. Wang, S., Huo, X., Yang, Y., Mo, Y., Kollipara, R. K., & Kittler, R. (2023). Ablation of EWS-FLI1 by USP9X inhibition suppresses cancer cell growth in Ewing sarcoma. *Cancer letters*, *552*, 215984.
135. Sekar, R. B., & Periasamy, A. (2003). Fluorescence resonance energy transfer (FRET) microscopy imaging of live cell protein localizations. *The Journal of cell biology*, *160*(5), 629–633.
136. Yamamoto, T., Shimojima, K., Ondo, Y., Shimakawa, S., & Okamoto, N. (2017). MED13L haploinsufficiency syndrome: A de novo frameshift and recurrent intragenic deletions due to parental mosaicism. *American journal of medical genetics. Part A*, *173*(5), 1264–1269.
137. Siavrienė, E., Petraitytė, G., Mikštienė, V., Maldžienė, Ž., Sasnauskienė, A., Žitkutė, V., Ambrozaitytė, L., Rančelis, T., Utkus, A., Kučinskas, V., & Preikšaitienė, E. (2023). Molecular and Functional Characterisation of a Novel Intragenic 12q24.21 Deletion Resulting in *MED13L* Haploinsufficiency Syndrome. *Medicina*, *59*(7), 1225.
138. Zhang, N., Song, Y., Xu, Y., Liu, J., Shen, Y., Zhou, L., Yu, J., & Yang, M. (2020). MED13L integrates Mediator-regulated epigenetic control into lung cancer radiosensitivity. *Theranostics*, *10*(20), 9378–9394.
139. Vilhais-Neto, G. C., Maruhashi, M., Smith, K. T., Vasseur-Cognet, M., Peterson, A. S., Workman, J. L., & Pourquié, O. (2010). Rere controls retinoic acid signalling and somite bilateral symmetry. *Nature*, *463*(7283), 953–957.
140. Wang, H., Gui, H., Rallo, M. S., Xu, Z., & Matise, M. P. (2017). Atrophin protein RERE positively regulates Notch targets in the developing vertebrate spinal cord. *Journal of neurochemistry*, *141*(3), 347–357.
141. Fregeau, B., Kim, B. J., Hernández-García, A., Jordan, V. K., Cho, M. T., Schnur, R. E., Monaghan, K. G., Juusola, J., Rosenfeld, J. A., Bhoj, E., Zackai, E. H., Sacharow, S., Barañano, K., Bosch, D. G. M., de Vries, B. B.

A., Lindstrom, K., Schroeder, A., James, P., Kulch, P., Lalani, S. R., … Sherr, E. H. (2016). De Novo Mutations of RERE Cause a Genetic Syndrome with Features that Overlap Those Associated with Proximal 1p36 Deletions. *American journal of human genetics*, *98*(5), 963–970.

142. Jordan, V. K., Fregeau, B., Ge, X., Giordano, J., Wapner, R. J., Balci, T. B., Carter, M. T., Bernat, J. A., Moccia, A. N., Srivastava, A., Martin, D. M., Bielas, S. L., Pappas, J., Svoboda, M. D., Rio, M., Boddaert, N., Cantagrel, V., Lewis, A. M., Scaglia, F., Undiagnosed Diseases Network, … Scott, D. A. (2018). Genotype-phenotype correlations in individuals with pathogenic RERE variants. *Human mutation*, *39*(5), 666–675.
143. Moccia, A., Srivastava, A., Skidmore, J. M., Bernat, J. A., Wheeler, M., Chong, J. X., Nickerson, D., Bamshad, M., Hefner, M. A., Martin, D. M., & Bielas, S. L. (2018). Genetic analysis of CHARGE syndrome identifies overlapping molecular biology. *Genetics in medicine: official journal of the American College of Medical Genetics*, *20*(9), 1022–1029.
144. Kolberg, L., Raudvere, U., Kuzmin, I., Adler, P., Vilo, J., & Peterson, H. (2023). g:Profiler-interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic acids research*, *51*(W1), W207–W212.
145. Liao, D., & Weiner, A. M. (1995). Concerted evolution of the tandemly repeated genes encoding primate U2 small nuclear RNA (the RNU2 locus) does not prevent rapid diversification of the (CT)n.(GA)n microsatellite embedded within the U2 repeat unit. *Genomics*, *30*(3), 583–593.
146. Adane, B., Alexe, G., Seong, B. K. A., Lu, D., Hwang, E. E., Hnisz, D., Lareau, C. A., Ross, L., Lin, S., Dela Cruz, F. S., Richardson, M., Weintraub, A. S., Wang, S., Iniguez, A. B., Dharia, N. V., Conway, A. S., Robichaud, A. L., Tanenbaum, B., Krill-Burger, J. M., Vazquez, F., … Stegmaier, K. (2021). STAG2 loss rewires oncogenic and developmental programs to promote metastasis in Ewing sarcoma. *Cancer cell*, *39*(6), 827–844.e10.

147. Surdez, D., Zaidi, S., Grossetête, S., Laud-Duval, K., Ferre, A. S., Mous, L., Vourc'h, T., Tirode, F., Pierron, G., Raynal, V., Baulande, S., Brunet, E., Hill, V., & Delattre, O. (2021). STAG2 mutations alter CTCF-anchored loop extrusion, reduce cis-regulatory interactions and EWSR1-FLI1 activity in Ewing sarcoma. *Cancer cell*, *39*(6), 810–826.e9.

148. Chen, J., Bardes, E. E., Aronow, B. J., & Jegga, A. G. (2009). ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic acids research*, *37*, W305–W311.

149. Eischer, N., Arnold, M., & Mayer, A. (2023). Emerging roles of BET proteins in transcription and co-transcriptional RNA processing. *Wiley interdisciplinary reviews. RNA*, *14*(1), e1734.

150. Wahl, M. C., & Lührmann, R. (2015). SnapShot: Spliceosome Dynamics I. *Cell*, *161*(6), 1474–e1.

151. 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., O'Hanlon, D., Pan, Q., Ray, D., Zheng, H., Vizeacoumar, F., Datti, A., Magomedova, L., Cummins, C. L., Hughes, T. R., Greenblatt, J. F., … Blencowe, B. J. (2017). Multilayered Control of Alternative Splicing Regulatory Networks by Transcription Factors. *Molecular cell*, *65*(3), 539–553.e7.

152. Murphy, A. J., Li, A. H., Li, P., & Sun, H. (2022). Therapeutic Targeting of Alternative Splicing: A New Frontier in Cancer Treatment. *Frontiers in oncology*, *12*, 868664.

153. Zhang, Q., Di, C., Yan, J., Wang, F., Qu, T., Wang, Y., Chen, Y., Zhang, X., Liu, Y., Yang, H., & Zhang, H. (2019). Inhibition of SF3b1 by pladienolide B evokes cycle arrest, apoptosis induction and p73 splicing in human cervical carcinoma cells. *Artificial cells, nanomedicine, and biotechnology*, *47*(1), 1273–1280.

154. Bashari, A., Siegfried, Z., & Karni, R. (2023). Targeting splicing factors for cancer therapy. *RNA*, *29*(4), 506–515.

155. Yan, Y., Ren, Y., Bao, Y., & Wang, Y. (2023). RNA splicing alterations in lung cancer pathogenesis and therapy. *Cancer pathogenesis and therapy*, *1*(4), 272–283.

156. Pelish, H. E., Liau, B. B., Nitulescu, I. I., Tangpeerachaikul, A., Poss, Z. C., Da Silva, D. H., Caruso, B. T., Arefolov, A., Fadeyi, O., Christie, A. L., Du, K., Banka, D., Schneider, E. V., Jestel, A., Zou, G., Si, C., Ebmeier, C. C., Bronson, R. T., Krivtsov, A. V., Myers, A. G., … Shair, M. D. (2015). Mediator kinase inhibition further activates super-enhancer-associated genes in AML. *Nature*, *526*(7572), 273–276.

157. Syring, I., Klümper, N., Offermann, A., Braun, M., Deng, M., Boehm, D., Queisser, A., von Mässenhausen, A., Brägelmann, J., Vogel, W., Schmidt, D., Majores, M., Schindler, A., Kristiansen, G., Müller, S. C., Ellinger, J., Adler, D., & Perner, S. (2016). Comprehensive analysis of the transcriptional profile of the Mediator complex across human cancer types. *Oncotarget*, *7*(17), 23043–23055.

158. Kuuluvainen, E., Domènech-Moreno, E., Niemelä, E. H., & Mäkelä, T. P. (2018). Depletion of Mediator Kinase Module Subunits Represses Superenhancer-Associated Genes in Colon Cancer Cells. *Molecular and cellular biology*, *38*(11), e00573-17.

159. Franzetti, G. A., Laud-Duval, K., Bellanger, D., Stern, M. H., Sastre-Garau, X., & Delattre, O. (2013). MiR-30a-5p connects EWS-FLI1 and CD99, two major therapeutic targets in Ewing tumor. *Oncogene*, *32*(33), 3915–3921.

160. Schnetz, M. P., Handoko, L., Akhtar-Zaidi, B., Bartels, C. F., Pereira, C. F., Fisher, A. G., Adams, D. J., Flicek, P., Crawford, G. E., Laframboise, T., Tesar, P., Wei, C. L., & Scacheri, P. C. (2010). CHD7 targets active gene enhancer elements to modulate ES cell-specific gene expression. *PLoS genetics*, *6*(7), e1001023.

161. Wang, L., Charroux, B., Kerridge, S., & Tsai, C. C. (2008). Atrophin recruits HDAC1/2 and G9a to modify histone H3K9 and to determine cell fates. *EMBO reports*, *9*(6), 555–562.

162. Wang, L., Rajan, H., Pitman, J. L., McKeown, M., & Tsai, C. C. (2006). Histone deacetylase-associating Atrophin proteins are nuclear receptor corepressors. *Genes & development*, *20*(5), 525–530.

163. Ellegren, H. (2004). Microsatellites: simple sequences with complex evolution. *Nature reviews. Genetics*, *5*(6), 435–445.

164. Malwe, A. S., Mondal, S., Harlapur, P., & Sharma, V. K. (2023). Comparative Analysis of TG/CA Repeats in Sixteen Primate Genomes Reveals the Dynamics and Role of TG/CA Repeats in the Human Genome. *bioRxiv.*
chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.biorxiv.org/content/10.1101/2023.05.09.539941v1.full.pdf

165. Hui, J., Stangl, K., Lane, W. S., & Bindereif, A. (2003). HnRNP L stimulates splicing of the eNOS gene by binding to variable-length CA repeats. *Nature structural biology, 10*(1), 33–37.

166. Amemiya, H. M., Kundaje, A., & Boyle, A. P. (2019). The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Scientific reports*, *9*(1), 9354.

167. Hershey, J. W. B., Sonenberg, N., & Mathews, M. B. (2019). Principles of Translational Control. *Cold Spring Harbor perspectives in biology, 11*(9), a032607

168. Vaklavas, C., Blume, S. W., & Grizzle, W. E. (2017). Translational Dysregulation in Cancer: Molecular Insights and Potential Clinical Applications in Biomarker Development. *Frontiers in Oncology, 7*.

169. Hwang, S. P., & Denicourt, C. (2024). The impact of ribosome biogenesis in cancer: from proliferation to metastasis. *NAR cancer*, *6*(2), zcae017.

170. Williamson, D., Lu, Y. J., Fang, C., Pritchard-Jones, K., & Shipley, J. (2006). Nascent pre-rRNA overexpression correlates with an adverse prognosis in alveolar rhabdomyosarcoma. *Genes, chromosomes & cancer*, *45*(9), 839–845.

171. Zentner, G. E., Hurd, E. A., Schnetz, M. P., Handoko, L., Wang, C., Wang, Z., Wei, C., Tesar, P. J., Hatzoglou, M., Martin, D. M., & Scacheri, P. C. (2010). CHD7 functions in the nucleolus as a positive regulator of ribosomal RNA biogenesis. *Human molecular genetics*, *19*(18), 3491–3501.

172. Ribeiro-Dantas, M. D. C., Oliveira Imparato, D., Dalmolin, M. G. S., de Farias, C. B., Brunetto, A. T., da Cunha Jaeger, M., Roesler, R., Sinigaglia, M., & Siqueira Dalmolin, R. J. (2021). Reverse Engineering of Ewing Sarcoma Regulatory Network Uncovers PAX7 and RUNX3 as Master Regulators Associated with Good Prognosis. *Cancers*, *13*(8), 1860.

173. Salton, M., & Misteli, T. (2016). Small Molecule Modulators of Pre-mRNA Splicing in Cancer Therapy. *Trends in molecular medicine*, *22*(1), 28–37.

174. Grohar, P. J., Kim, S., Rangel Rivera, G. O., Sen, N., Haddock, S., Harlow, M. L., Maloney, N. K., Zhu, J., O'Neill, M., Jones, T. L., Huppi, K., Grandin, M., Gehlhaus, K., Klumpp-Thomas, C. A., Buehler, E., Helman, L. J., Martin, S. E., & Caplen, N. J. (2016). Functional Genomic Screening Reveals Splicing of the EWS-FLI1 Fusion Transcript as a Vulnerability in Ewing Sarcoma. *Cell reports*, *14*(3), 598–610.

175. Lawrence, L. A., Miller, H., & Bishop, A. (2022) Targeting the dysregulation of transcription and splicing in Ewing sarcoma [abstract]. *Proceedings of the AACR Special Conference: Sarcomas; Clinical Cancer Research, 28*(18_Suppl): Abstract nr A012.