



UNIL | Université de Lausanne

Unicentre

CH-1015 Lausanne

<http://serval.unil.ch>

Year : 2020

Identification of molecular mechanisms underlying CIC-DUX4 tumor pathogenesis and aggressiveness

Bakaric Arnaud

Bakaric Arnaud, 2020, Identification of molecular mechanisms underlying CIC-DUX4 tumor pathogenesis and aggressiveness

Originally published at : Thesis, University of Lausanne

Posted at the University of Lausanne Open Archive <http://serval.unil.ch>

Document URN : urn:nbn:ch:serval-BIB_02AE016283687

Droits d'auteur

L'Université de Lausanne attire expressément l'attention des utilisateurs sur le fait que tous les documents publiés dans l'Archive SERVAL sont protégés par le droit d'auteur, conformément à la loi fédérale sur le droit d'auteur et les droits voisins (LDA). A ce titre, il est indispensable d'obtenir le consentement préalable de l'auteur et/ou de l'éditeur avant toute utilisation d'une oeuvre ou d'une partie d'une oeuvre ne relevant pas d'une utilisation à des fins personnelles au sens de la LDA (art. 19, al. 1 lettre a). A défaut, tout contrevenant s'expose aux sanctions prévues par cette loi. Nous déclinons toute responsabilité en la matière.

Copyright

The University of Lausanne expressly draws the attention of users to the fact that all documents published in the SERVAL Archive are protected by copyright in accordance with federal law on copyright and similar rights (LDA). Accordingly it is indispensable to obtain prior consent from the author and/or publisher before any use of a work or part of a work for purposes other than personal use within the meaning of LDA (art. 19, para. 1 letter a). Failure to do so will expose offenders to the sanctions laid down by this law. We accept no liability in this respect.



UNIL | Université de Lausanne

Faculté de biologie
et de médecine

**Institut Universitaire de Pathologie (IPA)
Département « formation et recherche » (CHUV)**

**Identification of molecular mechanisms underlying CIC-DUX4 tumor
pathogenesis and aggressiveness**

Thèse de doctorat en médecine et en sciences

MD-PhD

Présentée à la

Faculté de biologie et de médecine
de l'Université de Lausanne

par

Arnaud BAKARIC

Médecin diplômé de la Confédération Helvétique

Jury

Prof. Pedro MARQUES-VIDAL, président et répondant MD-PhD

Prof. Nicolo RIGGI, directeur de thèse

Prof. Ivan STAMENKOVIC, co-directeur de thèse

Prof. Tatiana PETROVA, experte

Prof. Freddy RADTKE, expert

Prof. Beat SCHAEFER, expert

Lausanne 2020



UNIL | Université de Lausanne

Ecole doctorale

Ecole Doctorale

Doctorat MD-PhD

Imprimatur

Vu le rapport présenté par le jury d'examen, composé de

Président·e	Monsieur Prof. Lucia	Mazzolai
Directeur·trice de thèse	Monsieur Prof. Nicolò	Riggi
Répondant·e	Madame Prof. Lucia	Mazzolai
Expert·e-s	Madame Prof. Tatiana	Petrova
	Monsieur Prof. Freddy	Radtke
	Monsieur Prof. Beat	Schaefer

le Conseil de Faculté autorise l'impression de la thèse de

Monsieur Arnaud BAKARIC

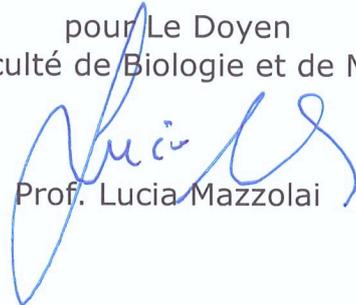
Médecin diplômé de la Confédération helvétique

intitulée

Identification des mécanismes moléculaires responsables de la pathogénèse et de l'agressivité des tumeurs CIC-DUX4

Lausanne, le 9 octobre 2020

pour Le Doyen
de la Faculté de Biologie et de Médecine


Prof. Lucia Mazzolai

Identification of molecular mechanisms underlying CIC-DUX4 tumor pathogenesis and aggressiveness

Table of contents

Table of contents	3
List of figures	5
List of abbreviations	6
Résumé	8
Background and significance	10
Introduction	12
<i>Adult cancer burden and cancer biology</i>	12
<i>Pediatric cancer specificities: the role of epigenetics</i>	13
<i>Epigenetic dysregulation induced by chromosomal translocation</i>	16
<i>Soft tissue sarcoma</i>	16
<i>Undifferentiated soft tissue sarcoma</i>	18
<i>CIC-rearranged sarcomas</i>	19
<i>CIC-DUX4 sarcoma</i>	20
<i>CIC biology</i>	21
<i>CIC discovery in Drosophila</i>	21
<i>Human CIC structure</i>	23
<i>CIC mediated target gene repression</i>	24
<i>CIC biological functions</i>	25
<i>CIC and cancer</i>	26
<i>DUX4 biology</i>	28
<i>DUX4 structure</i>	28
<i>DUX4 transactivation activity</i>	29
<i>DUX4 biological function in physiology and disease</i>	30
<i>CIC-DUX4 biology</i>	31
Hypothesis and aims of the study	34
Results	36
<i>Whole genome chromatin profiling of four frozen CDS primary samples reveals active chromatin features at CIC-DUX4 direct target genes' promoters and enhancers</i>	36
<i>CDS1, CDS2 and CDS3 patient-derived cell lines express CIC-DUX4 protein, have a distinct gene expression profile compared to Ewing sarcoma and induce SBRCT development in vivo</i>	37
<i>The CIC-DUX4 protein is localized in the nucleus and cytoplasm of primary CDS tumors and cell lines, generating heterogenous subpopulations of tumor cells</i>	39

<i>KO serum-complemented cell culture conditions allow for optimal ChIP-seq profiling of CDS cell lines, which show a chromatin landscape similar to the one obtained in primary CDS tumors</i>	40
<i>Induction of CIC-DUX4 expression in human cells leads to increased expression of CIC-DUX4 target genes depending on recipient cell's permissiveness</i>	43
<i>The CIC-DUX4 and p300 protein-protein interaction is mainly localized in the nucleus</i>	46
<i>Inhibition of p300 enzymatic activity in CDS1, CDS2 and CDS3 cell lines strongly reduces cell proliferation and tumor growth, and specifically downregulates CIC-DUX4 target genes</i>	47
<i>p300 is crucial for CIC-DUX4 but not endogenous CIC or DUX4 proteins stability</i>	48
<i>In vivo p300 inhibition strongly reduces CDS tumor growth</i>	49
Discussion and future directions	51
Materials and methods	63
<i>Cell culture</i>	63
<i>Primary tumor samples</i>	63
<i>Lentiviral infections</i>	63
<i>RNA extraction, cDNA synthesis and quantitative real-time PCR</i>	64
<i>Preparation of protein cell lysate and Western Blot</i>	65
<i>Immunofluorescence</i>	65
<i>Proximity ligation assay (PLA)</i>	66
<i>Chromatin Immunoprecipitation</i>	66
<i>In vitro p300 inhibition</i>	67
<i>In vivo tumorigenic assay</i>	67
<i>DUX4 immunohistochemistry</i>	67
<i>In vivo A-485 treatment</i>	68
Appendix 1 : LIN28B underlies the pathogenesis of a subclass of Ewing sarcoma	69
Published paper	72
Acknowledgments	73
References	76

List of figures

Fig. 1 Sarcoma classification diagram	17
Fig. 2 <i>Drosophila</i> CIC and MAPK pathway	22
Fig. 3 Human CIC gene and protein structures	23
Fig. 4 CIC regulation network	25
Fig. 5 CIC mutation and MAPK inhibition	27
Fig. 6 DUX4 gene and protein structures	28
Fig. 7 DUX4 transactivation effect	29
Fig. 8 CIC-DUX4 gene and protein structures	31
Fig. 9 Potential CIC-DUX4 mechanism of action	34
Fig. 10 CIC-DUX4 primary tumor chromatin profile	36
Fig. 11 CDS cell lines characterization	38
Fig. 12 CIC-DUX4 inter and intra cellular heterogeneity	40
Fig. 13 CIC-DUX4 chromatin binding	41
Fig. 14 CDS1 ChIP-seq profiling	43
Fig. 15 CIC-DUX4 induction	45
Fig. 16 CIC-DUX4 and p300 interaction	46
Fig. 17 <i>in vitro</i> p300 inhibition	48
Fig. 18 p300 inhibition effect on CIC, DUX4 and CIC-DUX4 proteins	49
Fig. 19 <i>in vivo</i> A-485 treatment of CDS2-derived tumors	50

List of abbreviations

2HG : 2 HydroxyGlutarate

ALL : Acute Lymphoblastic Leukemia

C-D OE : CIC-DUX4 overexpression

CDS : CIC-DUX4 Sarcoma

CIC : Capicua

CRS : CIC Rearranged Sarcoma

DUX4 : Double Homeobox 4

ES : Ewing Sarcoma

FISH : Fluorescence In Situ Hybridization

FSHD : FacioScapuloHumeral muscular Dystrophy

GCO : Global Cancer Observatory

HAT : Histone Acetyl Transferase

IARC : International Agency for Research on Cancer

IF : Immunofluorescence

MAPK : Mitogen Activated Protein Kinase

NTAS : Non-Translocation Associated Sarcoma

ODG : OligoDendroGlioma

PDX : Patient Derived Xenograft

PEA3 : Polyoma Enhancer Activator 3

RTK : Receptor Tyrosin Kinase

SBRCT : Small Blue Round Cell Tumor

STS : Soft Tissue Sarcoma

TAS : Translocation Associated Sarcoma

TF : Transcription Factor

URCS : Undifferentiated Round Cell Sarcoma

USTS : Undifferentiated Soft Tissue Sarcoma

WHO : World Health Organization

WT : Wild Type

Résumé

Les sarcomes des tissus mous regroupent des tumeurs hétérogènes et agressives qui représentent moins de 1% de toutes les tumeurs malignes de l'adulte et 10% des tumeurs pédiatriques. C'est seulement récemment que les analyses moléculaires et la génétique ont été intégrés dans la routine diagnostique et inclus dans la classification des tumeurs des tissus mous et des os publiée en 2013 par l'Organisation Mondiale de la Santé (OMS) (1)(2). Malgré ces progrès et le développement constant de nouvelles technologies, approximativement 20% de tous les sarcomes des tissus mous restent impossibles à classer à cause de leurs caractéristiques biologiques et cliniques, et sont donc qualifiés de « sarcomes des tissus mous non différenciés ».

Durant les dernières décennies, l'étude de la génétique des sarcomes a permis la distinction de deux sous-types biologiques : les sarcomes associés à des translocations chromosomiques, et les sarcomes non-associés à des translocations. Alors que certaines fusions génétiques résultants de translocations chromosomiques spécifiques ont été étudiées de manière approfondie, comme *EWS-FLI1* dans le sarcome de Ewing, d'autres, comme *CIC-DUX4* ou *BCOR-CCNB3*, sont encore méconnues et sont observées dans des sarcomes indifférenciés qui présentent, à l'analyse histologique, de petites cellules bleues et rondes. Bien que ces protéines de fusion jouent probablement un rôle crucial dans le développement des sarcomes, leurs propriétés biologiques sont toujours mal comprises (3).

Le sarcome *CIC-DUX4* (SCD) représente la majorité des tumeurs à petites cellules bleues qui ne présentent pas le réarrangement du gène *EWSR1* qui caractérise le sarcome de Ewing (3). Toutefois, étant donné leur ressemblance au sarcome de Ewing, ces tumeurs ont longtemps été qualifiées de tumeurs « Ewing sarcoma-like », et traitées en suivant la même approche multidisciplinaire utilisée pour le sarcome de Ewing malgré l'absence de réponse clinique et leur pronostic défavorable. D'ailleurs, il a été démontré que le SCD présente des caractéristiques cliniques et moléculaires spécifiques, suggérant que ce sarcome constitue une entité biologique distincte (4). L'analyse récente du transcriptome des SCD a révélé une signature d'expression génique unique renforçant l'hypothèse que le développement du SCD dépend probablement de l'expression de gènes spécifiques, induisant

la perturbation de certaines voies de signalisation moléculaires bien définies. Cette observation a des implications directes sur le développement de thérapies ciblées et sur la prise en charge de patients atteints de SCD (4)(5). Cependant, malgré cette avancée majeure, la rareté des cas de SCD et le manque de modèles expérimentaux *in vitro* et *in vivo* ont retardé notre compréhension des effets biologiques de CIC-DUX4, ainsi que l'identification des voies de signalisation moléculaires responsables du développement du SCD et de sa progression.

Dans cette étude, nous avons généré des nouveaux modèles expérimentaux pour étudier la biologie du SCD *in vitro* et *in vivo*. Nous avons combiné l'étude génomique non-biaisée de nos modèles avec des expériences fonctionnelles afin d'explorer les événements épigénétiques qui participent au développement et à la progression du SCD. Afin d'évaluer l'utilité clinique de notre travail, nous avons combiné les résultats obtenus à partir de nos modèles expérimentaux avec les profils épigénétiques que nous avons généré à partir de tumeurs primaires congelées.

Grâce à cette approche, nous avons identifié le recrutement de p300 par CIC-DUX4 au niveau de la chromatine de ses gènes cibles dans les lignées cellulaires de SCD, ainsi que leur dépendance à l'activité de cet enzyme de remodelage de la chromatine. L'inhibition pharmacologique de p300 a fortement ralenti la prolifération des cellules de SCD *in vitro* et cela même à des doses plus faibles par rapport à d'autres modèles tumoraux. La croissance tumorale *in vivo* a aussi été drastiquement réduite suite à l'inhibition de p300. De plus, nous avons constaté que les niveaux protéiques de CIC-DUX4, responsable du développement du SCD, ainsi que ceux de ses gènes cibles sont fortement réduits après l'inhibition de p300. L'inhibition pharmacologique de p300 semble donc être une approche thérapeutique prometteuse pour le traitement de ces tumeurs agressives.

Background and significance

Soft tissue sarcoma is a group of highly heterogeneous and aggressive tumors that accounts for less than 1% of all malignant tumors in adults, but represents 10% of all pediatric malignancies. Only recently, molecular analyses and genetics have been integrated in the diagnosis and classification of soft tissue sarcomas, which led to the publication of the 2013 World Health Organization (WHO) classification of tumors of soft tissue and bone (1)(2). Despite this progress and the constant development of new technologies, approximately 20% of all soft tissue sarcomas remain unclassifiable because of their lack of specific clinical and biological features, and have been referred as undifferentiated/unclassified soft tissue sarcomas (USTSs).

The study of sarcoma biology during the last decades has led to the distinction of two main subtypes : translocation-associated and non-translocation-associated sarcomas. Whereas some genetic fusions resulting from specific chromosomal translocations have been extensively studied and associated with specific tumor entities others, such as *CIC-DUX4* or *BCOR-CCNB3*, are typically observed in subsets of USTS harboring a “small blue round cell tumor (SBRCT)” histological phenotype that show no specific patterns of differentiation. Although these fusion proteins are believed to play crucial role in sarcomagenesis, their biological properties are still poorly understood (3).

CIC-DUX4 sarcoma (CDS) represents the vast majority of *EWSR1*-negative SBRCTs (3). Given its histological resemblance to Ewing sarcoma (EwS), it has often been referred as “Ewing sarcoma-like” tumor, and treated following the same multimodal approach used for typical Ewing sarcoma, despite the lack of clinical response and its poor prognosis. Indeed, increasing evidence shows that CDS displays specific clinical and molecular features which support the idea that CDS represents a distinct biological entity (4). Recent transcriptomic analysis of CDS revealed a unique gene expression signature, highlighting the notion that CDS development and maintenance most probably rely on distinct and specific gene expression and signaling pathway perturbations. This has direct implications for the development of targeted therapies and better CDS patient management (4)(5). Despite this major advance, the rarity of CDS cases and the lack of *in vitro* and *in vivo* models have

hampered the proper elucidation of CIC-DUX4 biological effects and the identification of the molecular pathways sustaining CDS development and progression.

In this study, we generated new experimental models to study CDS biology both *in vitro* and *in vivo*. We combined unbiased genome wide approaches with functional studies to investigate epigenetic events that participate in CDS establishment and maintenance. To assess the clinical relevance of our work, we combined the results we obtained from established CDS cellular models with genome wide gene expression and epigenetics data derived from a set of frozen primary CDS tissues. Interestingly, we uncovered a strong dependency of CDS on p300 function, identified the direct recruitment of p300 by the translocation at its direct binding sites, and showed that p300 pharmacological inhibition hindered CDS cell proliferation *in vitro* even at a very low dose. Moreover, inhibition of p300 induced the downregulation of CIC-DUX4 fusion protein, which is believed to be the main driver of CDS development, and of its downstream target genes' transcripts, revealing a potential new attractive therapeutic target for this aggressive malignancy.

Introduction

Adult cancer burden and cancer biology

Based on the information displayed by the Global Cancer Observatory (GCO), an interactive platform which integrates data on cancer epidemiology from the International Agency for Research on Cancer (IARC) projects, there were more than 4.2 millions new cancer cases in 2018 in Europe with nearly 2 millions of cancer-related deaths in the same continent for the same year. The cancer types responsible for the majority of cancer-related deaths worldwide for both sexes are lung cancer (18.4%), colorectal cancer (9.2%), stomach cancer (8.2%), liver cancer (8.2%) and breast cancer (6.6%). Despite the great advances achieved in cancer research and therapy over the last decades, the global burden of this disease is increasing worldwide and represents one of the biggest challenges for future medicine. There is an urgent need to identify new therapeutic approaches since cancer still dramatically impacts patients' quality of life and survival. Moreover, cancer has huge economic implications on healthcare systems. Indeed, cancer related European union's healthcare spending reached 57.3 billion euros in 2017 based on available information on the Cancer Atlas website.

If the frequency of some cancer types can be dramatically reduced by appropriate prevention strategies, the majority of cancers develop from an unknown or/and unpreventable event. The only chance to treat the latter category is through early diagnosis and use of complementary therapeutical approaches. For decades and still nowadays, the vast majority of cancers were treated following a multimodal approach including surgery, chemotherapy and radiotherapy. Even if these multimodal therapies could significantly improve the prognosis of several cancer types, they don't specifically target the tumor cells, but also affects normal cells, thus leading to the apparition of side effects which represent an important limitation for the complete cure of cancer. Moreover, cancer cells have shown a strong ability to elude current therapeutic regimes, and develop resistances responsible for tumor relapses. The failures of this multimodal approach led to an increasing effort in both fundamental and clinical research to better understand cancer biology, with the intent to develop therapeutic tools that specifically target cancer cells.

Cancer can arise from virtually any tissue that compose the human body thus representing a very heterogeneous group of diseases. Nevertheless, at the biological level, all cancers often share similar features that were described in two key reviews published in 2000 (6) and 2011 (7) by Prof. Douglas Hanahan and Prof. Robert Weinberg, two major contributors of cancer research. In these two reviews, the authors describe cancer as a complex and heterogeneous group of cells that include both transformed tumorigenic cells and normal cells from the tumor microenvironment. They describe the major features of cancer cells such as resistance to cell death signals, maintenance of a proliferative signaling and the capacity to invade surrounding tissues and to metastasize. In the 2011 review, new emerging hallmarks of cancer were added to the previous list of cancer features including the capacity to escape the immune system and the ability to take advantage of the plasticity and the redundancy of cell metabolism.

In these reviews, the acquisition of the defining features of cancer cells is explained mainly by the succession of genetic mutations and alterations in a normal precursor cell leading to transformation and tumorigenesis. Thus, cancer is believed to arise from clonal expansion of cells with an overrepresentation of the clones that developed a growth and survival advantage due to newly acquired somatic mutations.

Although this is often true in the majority of adult tumors, that acquire somatic mutations over the span of several years, pediatric cancers typically show a very low mutational load (excepted for pediatric tumors that carry germline mutations of genes implicated in proper DNA damage repair). Indeed, mutation frequency in pediatric cancer is 14 times lower compared to adult cancers (8), and there is strong evidence that adult and children cancer pathogenesis are, most of the time, two distinct processes.

Pediatric cancer specificities: the role of epigenetics

In developed countries, cancer represents the second cause of disease-related morbidity and mortality in children after accidental deaths (9). According to GCO available data, there were 25094 new cases of cancer in Europe in 2018 in the population aged from 0 to 19 years old. The most represented childhood cancer types are leukemias, brain and nervous system cancers and Hodgkin lymphomas.

Children cancer represents 1% of all cancer diagnosed each year and its overall five-year survival is 80% in high income countries (10). Despite the fact that this encouraging number shows the progress in diagnosis and treatment of childhood cancer, it should be considered with caution. First, because there is a lack of strong and reliable data from low income countries where children cancer is more prevalent and where there are less resources available for cancer treatment (10). This also explains why overall survival of childhood cancer varies drastically from one region to the other (11). Second, as for adult cancer, pediatric cancer is a very heterogeneous group of diseases that includes cancers with good overall survival like retinoblastoma and others with poor prognosis like bone sarcoma.

This last observation raises a fundamental question about pediatric cancer : what can explain the heterogeneity of childhood cancer types and the variability of their prognosis since there is a lack of recurrent somatic mutations (8)? Indeed, even some highly aggressive children cancers don't harbor any driver mutation or perturbed signaling pathways (12). One of the first elements that helped understanding the molecular pathogenesis of pediatric cancer is the heterogeneity of other genetic alterations than point mutations, such as copy number alterations, enhancer hijacking events, gene fusions and other structural alterations. This observation underlined the fact that, despite their low mutational burden, pediatric cancer is not a group of diseases with a simple genetic landscape and appears to be much more complex than expected (12). A second interesting point is raised by the fact that the vast majority of highly mutated genes across multiple pediatric cancers are linked to epigenetic modifications (8). These two observations suggest that pediatric cancer heterogeneity and aggressiveness are probably related to the dysregulation of gene expression and genome structure due to disturbed epigenetic pathways rather than by the presence of driver somatic mutations and accumulation of mutational "hits"(9).

Despite the difficulties in obtaining a consensual definition of epigenetics, it's widely accepted that this term refers to "the study of changes in gene function that are mitotically and/or meiotically heritable and that do not entail a change in DNA sequence" (13). Since the beginning of the modern era of epigenetic research which started in 1996 (14), numerous mechanisms that resulted in epigenetic modifications were discovered. DNA methylation, histone post-translational modifications

and histone variants are three well known examples among several others. Epigenetic regulation of gene expression is linked with all physiological processes from development to adulthood since it's crucial to define and maintain cells' identity and functional capacities. Given the central role of epigenetics in physiology, it's easy to understand that the dysregulation of epigenetic mechanisms can lead to cellular dysfunction and transformation. Indeed, it has been shown that all epigenetic mechanisms can be implicated in cancer development and maintenance. For example, cancer is known to show global hypomethylation of DNA with local hypermethylation at tumor suppressor genes' promoters which results in their repression (15). Concerning histones modifications, abnormal histone acetylation by histone acetyl transferases (HATs) such as CBP/p300 has been observed in neoplastic transformation (15). Although global epigenetic dysregulation has been observed in adult cancer, the perturbation of these mechanisms is more prevalent in pediatric cancers and is highly susceptible to play a causal role in the development of this disease (8).

Recent development of pediatric cancer treatment was achieved by using the same strategy as for adult tumors. The aim is to pinpoint specific aberrantly activated molecules such as tyrosin kinases or perturbed pathways like PI3K pathway and target them with small molecule inhibitors. If this approach showed some rate of success in very specific settings such as fusion-activated kinases its action is most of the time very limited due to the lack of targetable events harbored by pediatric cancer (16). Alternatively, due to its recent success in adult cancer treatment, immunotherapy has also been tested on pediatric tumors but, again, due to the relatively low accumulation of genetic mutations and thus neoantigen formation, childhood cancer demonstrates low immunogenicity. Another limitation to the modulation of the immune system to treat pediatric cancer is the relative immaturity of children's immune system which can strongly impact the therapeutic response to this approach. Nevertheless, the development of immunotherapeutical approaches such as checkpoint inhibitors and more recently, CAR-T cells technology, is currently in the clinical testing pipelines in order to better understand the pediatric cancer types that can benefit for this kind of treatment (16).

Epigenetic dysregulation induced by chromosomal translocation

Epigenetic dysregulation can occur through multiple mechanisms. Among these mechanisms, gene fusion resulting from chromosomal translocation was rapidly thought to be able to globally perturbate genes' expression of cells. The first discovery of a gene fusion was done in Burkitt lymphoma in which the chromosomal translocation t(8;14) induces the juxtaposition of the *IgH* promoter with the *MYC* oncogene, resulting in the constitutive expression of *MYC* and its downstream target genes (17). Interestingly, after this first discovery, gene fusions were shown to be able to encode for chimeric proteins capable of aberrant activities. One of the most famous aberrant chimeric protein is BCR-ABL1, which is encoded by the translocation t(9;22) and displays aberrant tyrosine kinase activity in chronic myeloid leukemia (17). Although chimeric proteins generated by specific chromosomal translocations may display specific molecular functions like tyrosine kinase activities, one of the most powerful consequence of gene fusions is the generation of aberrant transcription factors capable of changing the whole epigenetic landscape of precursor cells by perturbing their chromatin architecture (17). A prototypical example of this scenario is the generation of the EWS-FLI1 aberrant transcription factor by the t(11;22) chromosomal translocation which leads to Ewing sarcoma development.

The discovery of chromosomal translocations has helped understanding the biology of multiple tumor types, and provided new diagnostic and classification tools. Some cancer types were shown to frequently harbor gene fusion, such as soft tissue sarcoma, in which 20% of the cases show recurrent chromosomal translocations.

Soft tissue sarcoma

Soft tissue sarcoma (STS) is a group of rare malignant tumors accounting for less than 1% of all cancer types in adults, but 10% in the pediatric population. STS can occur in virtually all anatomical sites and include more than 100 histological entities that display consequent variation in prognosis and clinical history. This heterogeneity renders the classification and diagnosis of specific subtypes of STS extremely difficult, but recent molecular analyses have facilitated this process, and helped characterizing and discovering new STS entities. In 2013, the World Health Organization published

the most recent classification for the tumors of soft tissues and bone, which integrates clinical aspects, macroscopic and microscopic observations with molecular analyses and genetics (1)(18). A schematic presentation of the most recent sarcoma classification is shown in Figure 1, offering an overview of the complexity of sarcoma subtypes. This very simplified picture of sarcoma classification diagram derived from 2013 WHO classification is only made to understand the challenging task of integrating several features such as morphology or molecular aspects when trying to obtain a specific sarcoma diagnosis.

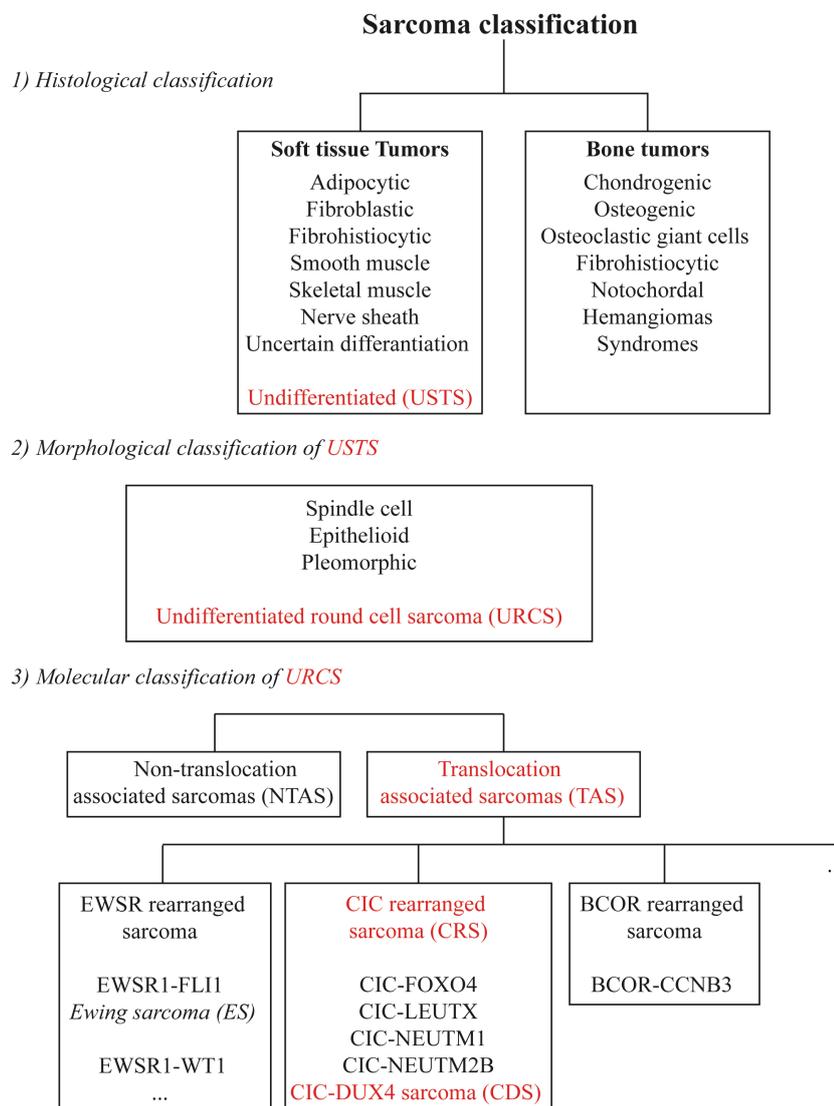


Fig. 1 Sarcoma classification diagram

Schematic overview of sarcoma classification. Obtained and modified from the WHO 2013 classification of tumors of soft tissues and bone. Red = subgroups related to/including CDS

Focusing on the molecular aspects of STS we can distinguish between two broad groups of tumors : non-translocation associated sarcoma (NTAS), and translocation associated sarcoma (TAS), which represent a non-negligible fraction of all STS as mentioned in the previous chapter (19). TAS occur more often in the pediatric population, and include both tyrosine kinase genes fusions like *TPM3/4-ALK* in inflammatory myofibroblastic tumor, and transcription factor genes fusions like *EWSR1-FLI1* in Ewing sarcoma.

Despite this effort to characterize every specific soft tissue sarcoma subtype, there is still a heterogeneous group of soft tissue sarcomas showing no identifiable line of differentiation referred as undifferentiated soft tissue sarcoma in last WHO classification.

Undifferentiated soft tissue sarcoma

USTS accounts for 10% of all soft tissue sarcoma and is still considered as an exclusion diagnosis, since there is no pathognomonic feature of USTS (20). The gross pathology observation of USTS shows no specific features, whereas the histological analysis reveals variable tissue patterns including epithelioid, round cell and spindle cell patterns, associated with a lack of specific immunohistochemical diagnostic markers. From a clinical standpoint, USTS often occurs in children, adolescents and young adults (1). Like other soft tissue sarcoma, USTS can develop at any anatomical location but has shown to be more frequent in the trunk and the extremities. All USTSs are considered as high-grade tumors but the prognosis can vary depending on the USTS subtype and the patient clinical features. For example, pleomorphic USTS in adults has more than 80% five-year survival rate, but epithelioid USTS ten-year survival is only roughly 40% (1). Considering the pediatric population, the Children's Oncology Group published a report including a cohort of 32 undifferentiated sarcoma patients with median age of 13.6 years in which the five-year overall survival after multimodal therapy was 83%. However, as mentioned in this report, these results should be taken with extra care since USTS contains several entities showing diverse pathogenesis, molecular signatures and above all clinical outcomes going from favorable to ominous (21). Since the publication of the 2013 WHO classification, there was an emergence of newly characterized undifferentiated sarcoma subsets especially in sarcomas harboring a round cell morphology (1).

Indeed, if alveolar rhabdomyosarcoma, Ewing sarcoma and synovial sarcoma are three examples of well-known small blue round cell tumors, there is growing evidence that SBRCTs harboring other genetic events such as *CIC* or *BCOR* rearrangements should be also considered as independent clinical and molecular entities (22)(23).

CIC-rearranged sarcomas

Until now, CIC rearranged sarcomas (CRSs) were included in the group of undifferentiated sarcomas with round cell phenotype (undifferentiated round cell sarcoma or URCS), because of their lack of specific features, and their typical SBRCT cytomorphology. For this reason, they have been referred as “Ewing sarcoma-like” tumors, along with BCOR rearranged sarcoma, and treated following the same therapeutic protocol used for canonical Ewing sarcoma, although they respond poorly and display a much worse prognosis (22). Indeed, clinical features and history of CRS patients are quite different from typical Ewing sarcoma patients. According to a recent study on the largest cohort of CRS patients which included 115 cases, 86% of the tumors developed in soft tissues (*vs.* 15% for Ewing sarcoma), mostly in the trunk and the extremities, but less than 3% in the bones (*vs.* 85% for Ewing sarcoma) (22). Under the microscope, CRS and Ewing sarcoma show almost the same small blue round cell histological appearance, with the tendency of CRS to often resemble to atypical Ewing sarcoma, showing sometimes lobular growth pattern, myxoid stroma and high frequency of geographic necrosis. Immunohistochemical analysis of CRS by using standard antibody panel for URCS differential diagnosis shows diffuse CD99 positive staining only in 24% of the cases, with strong WT1 signal in 75% of all cases (22), whereas Ewing sarcoma displays a strong CD99 staining and negativity for WT1 in nearly all the cases (4). Regarding clinical history of CRS patients, the prognosis is much worse compared to Ewing sarcoma with a metastatic rate of 53%, high risk of recurrence and an overall five-year survival rate of 43% when treated with the same multimodal approach including surgery, chemotherapy and radiotherapy (22). This ominous clinical reality underlies the need to study CRS at the molecular level in order to develop targeted therapies and specific treatment approaches.

Among CIC rearranged sarcomas, including different gene fusions such as *CIC-FOXO4*, *CIC-LEUTX*, *CIC-NUTM1* and *CIC-NUTM2B*, the *CIC-DUX4* fusion gene resulting from the chromosomal translocations t(4;19)(q35;q13) or t(10;19)(q26.3;q13) represent the most frequent genetic event observed in EWSR1-negative SBRCTs (22).

CIC-DUX4 sarcoma

If clinical history of CRS has been well documented in a study including a reasonable number of patients (22), clinical data for CIC-DUX4 sarcoma patients is scarce due to the low prevalence of this tumor. However, as observed in a review study published by Haidar *et al.* on 44 case of CIC-DUX4 sarcoma reported in the literature (24), this translocation related sarcoma seems to be a very aggressive subtype of CRS. Indeed, the average lifespan after diagnosis of the 44 CDS cases reported was 15.4 months with only 6 cases showing prolonged survival (22-48 months). This aggressiveness is undoubtedly related to the fact that the metastatic rate is high (59%) and that there is no current specific treatment modality for CIC-DUX4 sarcoma tumors (24). The patients clinical characteristics of this study show that the mean of the age of all patients is 27, and that there is a slight tendency of CDS to occur most frequently in females (M:F is 1:1.31). The most common affected sites by the primary tumor are the limbs while lungs are the most common sites of metastasis. In one of the studies on 22 CIC-DUX4 sarcoma cases included in Haidar *et al.*'s review, the authors reported a very high tendency (91%) of this tumor to develop in soft tissue (25). The histological findings show the presence of solid sheets of small round/oval cells. However, significant heterogeneity in nuclear shape and size was reported which differ from the monomorphic appearance of the cells in other Ewing family tumors. Moreover, a recent study reported high intra-tumor heterogeneity of cell and nuclear morphologies in CIC-DUX4 sarcoma when observed with electronic microscopy (26). Whether this intra-tumor heterogeneity is related to distinct biological features and relevant for CIC-DUX4 sarcoma treatment still needs to be investigated. In addition, areas of myxoid and edematous stromal changes, as well as geographic necrosis are frequently observed (4). The immunohistochemical findings reveal that CD99 can be positive or negative in CIC-DUX4 tumors, WT1 is positive in the vast majority of tumors and shows a nuclear and cytoplasmic staining pattern (4). The WT1 positivity

is another reason to believe that CDS is a separate entity from EwS, because WT1 is always negative in the latter. ETV4 immunostaining appears to be a valid marker for the diagnostic of CRS (27), but DUX4 immunostaining seems to be the most sensitive and specific marker for CIC-DUX4 diagnosis (28). Currently, the definitive diagnosis of CIC-DUX4 tumor is made by fluorescence in situ hybridization (FISH) or RT-PCR for the gene fusion (4). Once the diagnosis is established there is currently no guidelines for the treatment of CDS. Thus, CDS patients are treated with the same regimen used for EwS or other sarcoma types, but respond very poorly with a strong tendency to relapse and to metastasize (24). Understanding CDS biology will then help developing more accurate diagnostic tools and targeted therapies to treat this extremely aggressive disease.

The balanced t(4;19) translocation was initially reported in a case report published by Richkind *et al.* in 1996 (29), but the resulting fusion gene was firstly described by Kawamura-Saito *et al.* in 2006 (30). In this seminal work the authors showed that the fusion gene included *CIC* on chromosome 19 and *DUX4* on chromosome 4, and that the resulting protein behaved as an aberrant transcription factor capable of transforming NIH 3T3 fibroblasts by inducing the expression of oncogenes such as the *PEA3* family genes. Due to the unique genomic features of the *DUX4* locus explained below, in a minority of cases *CIC* can also be fused to the *DUX4L* gene located on chromosome 10, due to a t(10;19) chromosomal translocation. This genetic event also gives rise to CRS, but only 13 of such cases have been reported to date (31). The strong transforming properties of the CIC-DUX4 protein is directly linked to the specific biological features of both fusion partners, CIC and DUX4.

CIC biology

CIC discovery in Drosophila

Capicua or *CIC* is an evolutionary conserved gene that has been first described in *Drosophila Melanogaster*, where it was reported to be important for the development of the head and tail of the embryo. The effect of CIC in fly development is mainly due to its repressive activity on receptor tyrosine kinase (RTK) expression (32). *CIC* encodes a HMG-box transcription factor that acts on the mitogen activated protein kinase (MAPK) pathway downstream of Torso and EGFR (33). The CIC HMG-box DNA binding domain recognizes the octameric sequence T(G/C)AATG(A/G)A in

promoters and enhancers of *Drosophila* homologue MAPK target genes (32). In addition to the HMG-box domain, a C-terminal motif C1 was also reported to be conserved and essential for CIC repressive activity.

Mechanistically, when a signal activates the RTK pathway, CIC becomes phosphorylated on its C-terminal C2 motif by Rolled (*Drosophila* homolog for human ERK) leading to its degradation or cytoplasmic delocalization depending on the initial upstream MAPK trigger signal, Torso or EGFR respectively (Fig. 2) (33). More recently, it has been observed that CIC targets are activated before downregulation and degradation of CIC but directly after rapid dissociation of CIC from the DNA which occur only 5 minutes after ERK activation (34). In addition to Torso and EGFR signaling, other mechanisms have shown to stabilize or protect CIC protein from degradation, as for example through the COP9 signalosome (35).

CIC has shown to be part of several additional biological pathways in *Drosophila* other than embryonic patterning, including wing development (36), neuroblast differentiation (33) and follicle cells fate determination (33).

Of great importance, CIC also needs to be repressed to permit the proliferation of larval structures during *Drosophila* development, thus acting as a growth inhibitor. Indeed, proliferation-related genes such as *Cyclin E* or *String/Cdc 25* are repressed by CIC in absence of EGFR/Ras activation (37). Interestingly, mutations of *CIC* led to increased proliferation independently of upstream RTK stimulation bypassing the need of an upstream signal for proper MAPK pathway activation (33). This is thought to occur via de-repression of proliferation-related genes that are normally bound by wild type (wt) CIC in absence of MAPK pathway activation (33).

The MAPK activity sensor mechanism is modulated by several other biological pathways such as the Hippo pathway in order to precisely fine-tune the response to EGFR/MAPK at crucial timepoints of

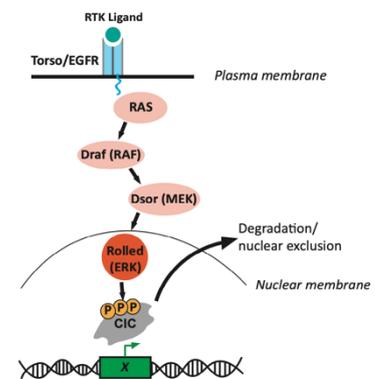


Fig. 2 Drosophila CIC and MAPK pathway

When a ligand binds to Torso/EGFR, the MAPK pathway gets activated leading to Rolled (ERK) mediated phosphorylation of CIC inducing its rapid dissociation from its target DNA and its degradation/nuclear exclusion. RTK, receptor tyrosine kinase; X, any MAPK target gene.

Drosophila development (38). Another layer of complexity is also added by the fact that, depending on which co-factor CIC is interacting with, it can bind to different target genes (39). Of great interest most of the major CIC features discovered in *Drosophila* are also conserved across multiple species including mammals and humans (32).

Human CIC structure

The human *CIC* gene is composed of 20 exons and is located on chromosome 19q13.2. It encodes for two protein isoforms via alternative transcription start sites : CIC-L (2517 a.a) and CIC-S (1608 a.a), which differ for their N-terminal region of unknown function (Fig. 3A,B) (40).

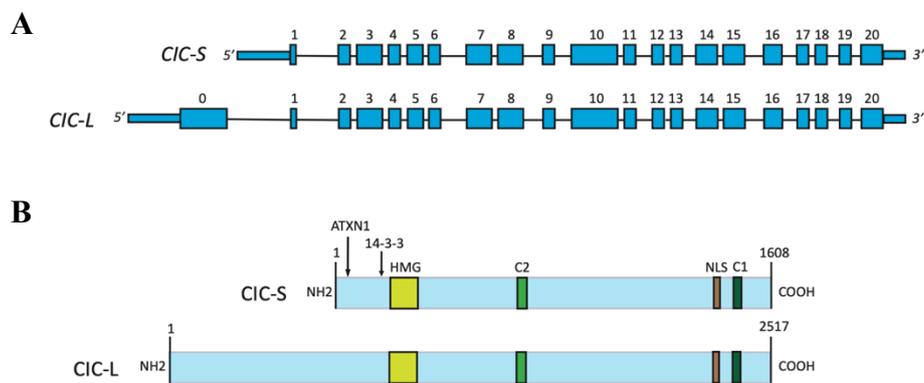


Fig. 3 Human CIC gene and protein structures

A. Structures of *CIC-S* and *CIC-L* genes. B. Structures of *CIC-S* and *CIC-L* proteins. C1, C-terminal domain 1; C2, C-terminal domain 2; *CIC-S*, Capicua short isoform; *CIC-L*, Capicua long isoform; HMG, high mobility group DNA binding domain; NLS, nuclear localization signal. Arrows are showing partners' binding sites.

The human *CIC* protein has retained the N-terminal HMG box DNA binding domain, the repressive C1 motif and the MAPK docking C2 motifs present on *Drosophila* *CIC* protein. In addition to these domains, human *CIC* contains a binding domain for ATXN1 and ATXN1L, two factors that play a role in *CIC* protein stabilization through dimer and tetramer formation and help *CIC* to play its repressive activity (37). Recently, it has been discovered that *CIC* and ATXN1/ATXN1L are reciprocally supporting each other's protein stabilization. Moreover, ATXN1 is required for the repressive activity of *CIC* on some but not all of the MAPK targets, and modulates *CIC* activity depending on the cellular context acting either as a co-repressor or a co-activator (41). Similar to its function in *Drosophila*, mammalian *CIC* binds to a specific DNA motif present in MAPK target genes' promoters and enhancers, repressing their transcription. In a similar way, *CIC* is directly

regulated by ERK or its substrate p90RSK through phosphorylation of its serine residues S1382 and S1409 or S173 respectively, which impairs CIC activity by, respectively, preventing its entry into the nucleus and impairing its DNA binding ability (33)(42). In addition to the MAPK pathway and ATXN1 interaction, the CIC protein has been also shown to be targeted and negatively regulated by several micro-RNAs, including miR-93/106/375 in prostate adenocarcinoma (43), miR-1307 in ovarian carcinoma (44) and miR-106b in renal carcinoma (45).

CIC mediated target gene repression

The mechanism exploited by CIC to repress its direct target genes was recently reported, and depends on the ability of CIC to recruit co-factors in order to build a transcriptional repressive complex. By analyzing CIC binding partners by mass spectrometry S. Weissmann *et al.* identified the SIN3A members HDAC1 and HDAC2, two well-known histone deacetylating enzymes. This group has shown that CIC recruitment at specific target genes induces histones deacetylation through HDAC activity, and thus inhibition of MAPK target genes' transcription (46). More recently, it has been suggested that, during neurogenesis, CIC achieves target gene repression through concomitant recruitment of SIN3A and several components of the SWI/SNF complex (47). Indeed, co-enrichment of CIC, BRG1, SIN3A and HDAC2 occurs at specific CIC target sites (*ETV4*, *ETV5* and *VGF* promoters) and depletion of CIC reduced the occupancy of the other partners at those sites (47). Of interest, CIC colocalization with the SWI/SNF complex has been observed during mitosis around chromosomes, suggesting a role for CIC in proper maintenance of chromosome integrity during mitosis (48). Outside the nucleus, CIC-S interacts with ACLY at the mitochondrial membrane and modulates its enzymatic activity which consists in the conversion of citrate in acetyl-coA (49).

This complex regulatory network (Fig. 4) is constantly tightly modulated by integrating intracellular and extracellular signals, which reflect the importance of this protein in a wide range of biological processes.

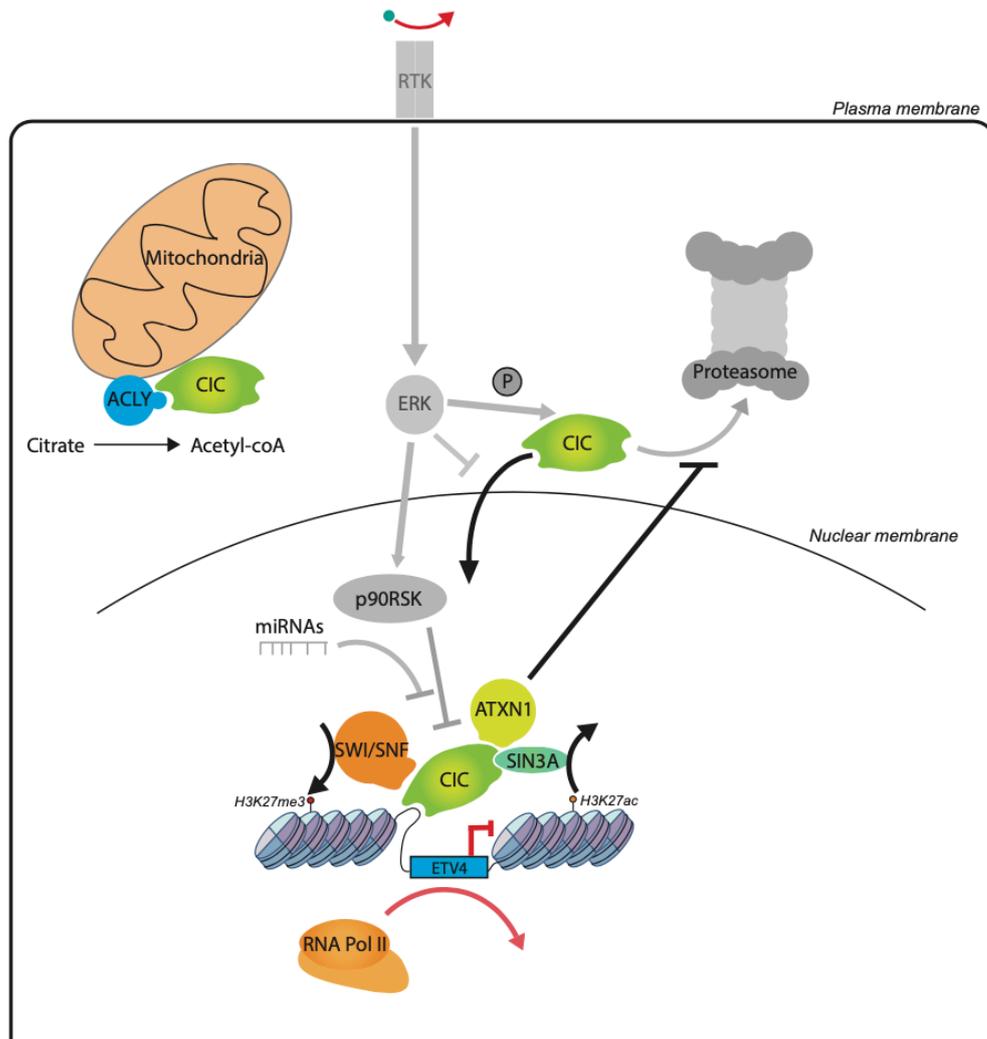


Fig. 4 CIC regulation network

In absence of MAPK pathway stimulation, the tumor suppressor CIC recognizes its target gene regulating elements through its N-terminal HMG box domain. Its repressive activity is mediated through the recruitment of two major complexes : the SIN3A complex containing histone deacetylating enzymes that remove acetyl groups from acetylated lysine 27 of histone 3 tail (H3K27ac), a mark associated with active genes' regulatory elements and the SWI/SNF complex which deposits H3K27me3 repressive mark at target genes promoters. ATXN1 stabilizes CIC. At the mitochondria, CIC modulates ACLY to generate acetyl-Coa. Color = active pathway. Greyscale = inactive pathway. H3K27ac, acetylation of lysine 27 of histone 3 tail ; H3K27me3, methylation of lysine 27 of histone 3 tail ; miRNAs, micro-RNAs 93, 106 and 1307 ; RTK, Receptor Tyrosin Kinase.

CIC biological functions

CIC functional studies in mice have shown that *CIC-L^{-/-}* animals died before weaning due to severe lung developmental defects. Induced expression of a CIC mutant lacking the HMG-box led to the same phenotype in mice, frequently associated to omphalocele (33). Nevertheless, knockdown experiments of *CIC* at specific developmental stages and across different cell types have revealed that *CIC* is expressed in multiple tissues and is implicated in numerous physiological processes. In the brain, its expression has been linked to neuronal and oligodendrocytic cell fate decision and proliferation. Indeed, Rui Yang *et al.* showed that knockout of *CIC* in neural stem cells impaired their

neuronal and oligodendrocytic differentiation program, conferring them with the capacity to proliferate in an EGF-independent manner (50). Intriguingly, CIC has shown to bind to several target genes that are unrelated to the EGFR pathway, suggesting an implication in other biological processes (50). More recently, it has been observed that CIC loss increased glial cells at the expense of neurons through increased proliferation and self-renewal capacities of neural stem cells and higher permissiveness for oligodendroglial lineage mainly because of *ETV5* de-repression (51). In the cerebellum, CIC interaction with the abnormal ATXN1 protein containing expanded CAG triplets is required for Ataxia Type 1 (SCA1) disease manifestation. Deletion of *CIC* in other regions of the brain resulted in different neurobehavioral abnormalities such as hyperactivity, impaired memory and autism spectrum disorders (37). In the murine hematopoietic system, *CIC* expression was shown to maintain a peripheral immune tolerance in order to avoid auto-immunity. Indeed, knockout of *CIC* in hematopoietic cells induced proliferation of follicular helper T lymphocytes, which stimulated germinal center B-cells and the infiltration of immune cells into healthy tissues (52). In addition to this, *CIC* is important for the proper alveolarization of the lung by maintaining low levels of MMP9 during lung development and it has been shown to participate to efficient liver homeostasis (37).

CIC and cancer

Given the crucial role played by *CIC* in regulating MAPK activity and cellular proliferation, its activity as a tumor suppressor has been extensively demonstrated in multiple studies. The close relation between *CIC* mutations and cancer development is mainly linked to the fact that *CIC* inactivation results in the de-repression of the MAPK target genes, particularly the *PEA3* family oncogenes *ETV1*, 4 and 5 (53). Among the most compelling examples of the direct link between *CIC* mutations and cancer development there is oligodendroglioma (ODG). This tumor often harbors 1p/19q codeletion and is associated with better prognosis than higher grade astrocytomas. Both bi-allelic mutations of *CIC* or loss of its expression have been reported in ODG (40). Recent studies have shown that *CIC* is mutated in the majority of oligodendrogliomas along with *IDH* and *FUBP1* mutations. In this tumor, the majority of *CIC* missense mutations arise in *CIC* HMG-box and C1 domains which impairs its DNA binding and repressive capacities (40). Interestingly, *in vitro* studies

demonstrated that concomitant mutations of *IDH* and *CIC* induced a higher production of the oncometabolite 2 hydroxyglutarate (2HG) through reduction of ACLY levels showing that, due to its multiple biological roles, *CIC* loss of function can promote tumor formation through different pathways (49). *Capicua* elevated expression in OGD was linked to both favorable or poor prognosis depending on the tumor genetic background. *IDH* mutated with 1p/19q co-deletion OGDs have a better prognosis when *CIC* is highly expressed whereas OGDs with *IDH* mutation but intact 1p/19q tend to have a poorer overall survival time when *CIC* is overexpressed suggesting a pro-oncogenic role of *Capicua* only in a particular genetic context (54). In Glioblastoma, *CIC* protein is constantly degraded through the ubiquitin-proteasome pathway due to a strong activation of ERK. Interestingly, in this tumor *CIC* stabilization has been shown to potentialize ERK inhibitors (55). Given its gatekeeper role in major oncogenic pathways and its almost ubiquitous expression across all human tissues, *CIC* loss of function has been implicated in cancers arising from very diverse tissue types, including T-cell lymphoblastic lymphoma (56), renal carcinoma (45) and breast cancer (57). *CIC* loss of expression was also shown to play important roles for cancer invasion and metastasis, as in hepatocellular carcinoma where *CIC* low expression is linked to an locally invasive phenotype through the ETV4-MMP1 axis (58), or in lung metastatization (59).

Of clinical interest, *CIC* mutations in cancer are also responsible for the resistance toward MAPK inhibition by MEK and RAF inhibitors, due to its downstream effect on the MAPK pathway. Since mutated *CIC* is not able to bind to its target genes' regulatory elements, the inhibition of upstream MAPK does not result in efficient *CIC* mediated target gene repression. Thus, *PEA3* family genes' expression levels remain high despite upstream MAPK pathway inhibition (Fig. 5). In hepatocellular carcinoma, *CIC* downregulation has been discovered in Sorafenib resistant tumors which correlated with

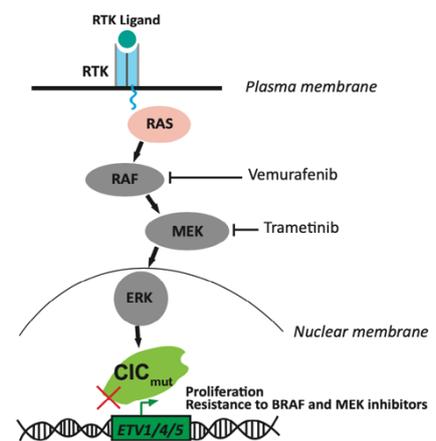


Fig. 5 CIC mutation and MAPK inhibition

Mutated *CIC* is not able to repress its target genes even in the absence of ERK mediated phosphorylation due to the pharmacological inhibition of upstream RAF or MEK. RTK, Receptor Tyrosin Kinase.

a low overall survival. By switching the Sorafenib treatment to Rigorafenib, a more potent ERK inhibitor, Sorafenib-resistant hepatocellular carcinoma cell proliferation was impaired *in vivo* (60).

The loss of CIC repressive function is thus strongly implicated in the pathogenesis of multiple cancer types and, in some cases, really sustains crucial tumorigenesis steps such as invasion and metastatization. Moreover, CIC status could play important role as a prognostic and predictive marker for proper patient categorization and treatment.

Finally, in CDS pathogenesis, the effects of CIC's loss of function are very likely to be intensified due to the fusion of CIC with the DUX4 transactivator domain, creating an aberrant protein which actively induces the expression of oncogenes such as *ETV1*, *ETV4* and *ETV5*.

DUX4 biology

DUX4 structure

Double homeobox 4 gene (*DUX4*) is a retrotransposed intronless gene located within the 3.3kb D4Z4 microsatellite repeated units in the subtelomeric region of chromosome 4 (61). Each repeated sequence contains one copy of the *DUX4* gene and in healthy individuals the number of D4Z4 repeats typically varies between 8 to 100 units (62)

(Fig. 6A).

It is of note that a homologous *DUX4* gene (*DUX4c*) is located at 42 kb from the *DUX4* locus towards the centromere of chromosome 4. Another layer of complexity is added by the fact that a nearly identical D4Z4-like repeated region has been discovered in chromosome 10 and contains the coding sequence for the *DUX4* homologue *DUX4L* (63). *DUX4* gene encodes for two *DUX4* isoforms due to alternative splicing. *DUX4* full-length isoform (*DUX4-FL*) is a pioneer transcription factor that contains 424 amino acids including two N-terminal DNA-binding homeodomains which bind to a TAATCTAATCA sequence, a disordered mid-region and a C-terminal region containing a

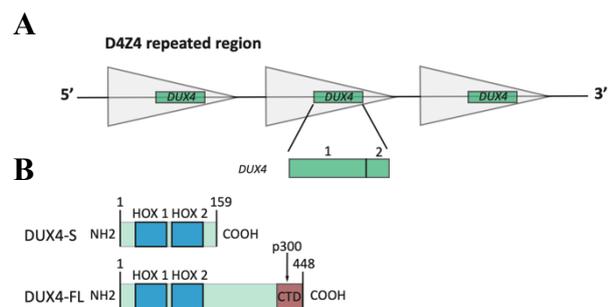


Fig. 6 DUX4 gene and protein structures

A. *DUX4* is an intronless gene that contains two coding exons (1 and 2) located within the D4Z4 repeated region. **B.** Alternative splicing of *DUX4* mRNA leads to the establishment of *DUX4-S* and *DUX4-FL* proteins that differ in their C-terminal regions. CTD; C-terminal domain. The arrow is showing p300 binding site.

transcription activating domain including a CBP/p300 binding domain. The DUX4 short isoform (DUX4-S) contains only the N-terminal 159 amino acids and has no transactivation activity since it lacks the C-terminal domains, thus its function remains unknown (Fig. 6B).

DUX4 transactivation activity

The DUX4 transactivation capacity is believed to rely on its C-terminal 98 amino acids which contain a putative p300 interaction domain, based on the ability of this fragment to induce complete upregulation of DUX4 target genes (64). More specifically, the 20 last C-terminal amino acids of DUX4 seem to play a major role in target gene expression activation (65). The central role of p300 recruitment has been reinforced by the fact that p300 chemical inhibition leads to a decrease in DUX4-mediated target gene expression and a reversion of hyperacetylation induced by DUX4 dependent p300 recruitment (66). DUX4 has also shown to downregulate the expression of some target genes but this is probably not mediated through its C-terminal domain and still needs to be further investigated (64). Of great interest, DUX4 has the capacity to bind to DNase sensitive (accessible) and non-sensitive (inaccessible) chromatin, thus acting as a “pioneer transcription factor” capable of inducing the opening of previously inaccessible chromatin regions. When bound to its target sites, DUX4 displaces histone H3 and induces acetylation of lysine 27 of histone 3 (H3K27ac) through p300 recruitment.

Following this step, epigenetic changes such as methylation of lysine 4 of histone 3 (H3K4me3) occur, DUX4 target genes' promoters get activated and transcription starts (64). Moreover, in order to have an enhanced and sustained transcription of DUX4 target genes, DUX4 induces H3.X and H3.Y histone variants expression and their incorporation at its direct targets (Fig. 7) (67).

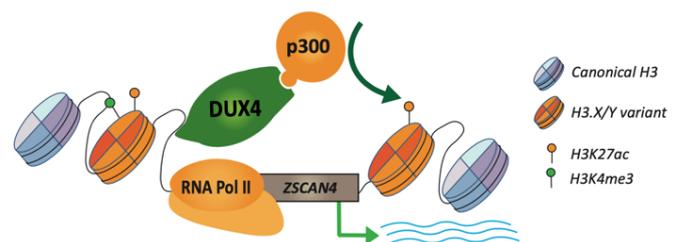


Fig. 7 DUX4 transactivation effect

The pioneer transcription factor DUX4 binds to DNA through its N-terminal domain and recruits p300 that catalyzes the deposition of acetyl groups on histones tails amino acids like H3K27. The target gene's promoter gets activated, associated with H3K4me3 and H3K27ac histone marks and transcription occurs. The incorporation of H3.X/Y histone variants potentiates the transcriptional activation of DUX4 target genes. H3K27ac, acetylation of lysine 27 of histone 3 tail. H3K4me3, three times methylation of lysine 4 of histone 3 tail.

DUX4 biological function in physiology and disease

In human embryonal development, *DUX4* is expressed only during early embryogenesis but gets silenced in most of the differentiated tissues excepted in testis. It is believed to play a role in the initiation of transcription during zygotic genome activation since its mRNA is present at oocyte stage and its target genes transcripts are generated from the 2-cell stage. During this process, *DUX4* activates the transcription of specific target genes and retrotransposed elements that are specifically expressed during the 2-cell stage (68). The mammalian apparent LTR-retrotransposons bound by *DUX4* create novel promoters for genes and non-coding transcripts that are re-activated in testis and in myoblasts in a specific pathologic condition called facioscapulohumeral muscular dystrophy (FSHD) (69). In this disease, both genetic and epigenetic events lead to the abnormal expression of *DUX4* protein in myoblasts which induces cell toxicity, inhibition of myogenesis and induction of cell death through expression of specific target genes from diverse pathways including WNT/ β -catenin, TNF- α and JNK signaling pathways (70). The stabilization of *MYC* transcript by *DUX4* has also been proposed as a major mechanism of apoptosis induction in FSHD (71). Myoblasts showing an abnormal expression of *DUX4* are also more sensitive to oxidative stress, a major feature in FSHD pathogenesis (72). Moreover, *DUX4* expression is associated with accumulation of inflammatory cells and fibroadipogenic progenitors that are more prone to differentiate into adipocytes and fibroblasts instead of replacing the damaged muscle fibers (73).

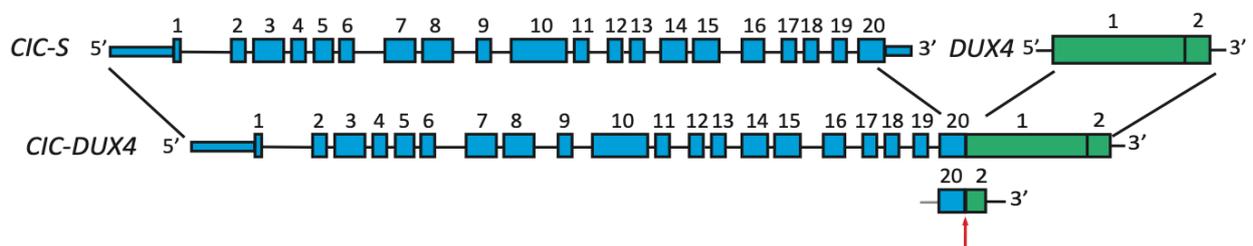
In cancer, *DUX4* has shown to have both tumor suppressor and pro-tumorigenic activities. In colon cancer, its induction mediates the inhibition of CDK1 activity leading to a reduced proliferation of colon cancer cells (74). *DUX4* is also implicated in a subtype of acute lymphoblastic leukemia (ALL) where its N-terminal part is fused with IGH, placing its expression regulation under control of the *IGH* enhancer. Following that, *DUX4*-IGH activates the transcription of an aberrant isoform of ERG which promotes leukemogenesis (75). More generally, a recent study has shown that several solid cancers harbor *DUX4* re-expression and consequently escape immune surveillance through suppression of interferon- λ -mediated induction of MHC class I-dependent antigen presentation (76). In addition to its immunomodulatory functions, *DUX4* has shown to induce the migration of

mesenchymal stem cells through CXCR4-SDF1 suggesting a possible role in additional steps of cancer development like invasion and metastasis (77). From the clinical point of view, DUX4 expression, as well as of other DUX family transcription factors, may also play a role in cancer categorization and classification since these factors are expressed in an important subset of cancers and promote tumorigenesis by re-establishing a zygotic gene activation signature (78).

CIC-DUX4 biology

The mechanism by which *CIC-DUX4* translocation induces CDS development is supposedly by exploiting the biological properties of both fusion partners. Structurally, the *CIC-DUX4* fusion gene includes all 20 exons of *CIC-S* at its 5' end, and the first two or only the second *DUX4* exon at the 3' end. Thus, *CIC-DUX4* encodes for an aberrant protein containing the CIC-S N-terminal HMG box DNA-binding domain and the C2/C1 domains (79) fused to the C-terminal last 100 amino acids of the DUX4 protein, harboring its transactivation domain and p300 binding site (64) (Fig. 8).

A



B

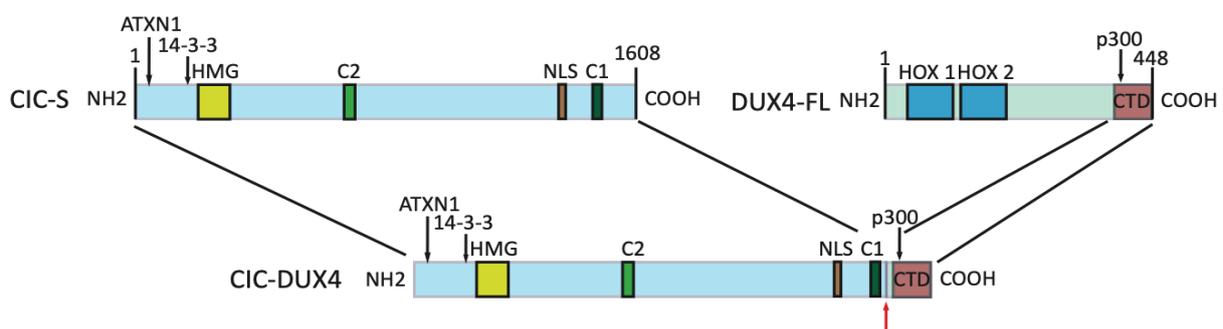


Fig. 8 CIC-DUX4 gene and protein structures

A. *CIC-DUX4* gene results from the chromosomal translocation t(4;19) and includes all 20 exons of *CIC* and the first two or only the second exon of *DUX4*. **B.** The *CIC-DUX4* protein is generated by fusion of the main part of *CIC* protein, including its DNA binding domain, C1/C2 domains and its nuclear localization domain, with *DUX4* C-terminal domain including its p300 binding domain. Red arrows indicate the breakpoint. CTD, C-Terminal Domain, NLS, Nuclear Localization Signal.

Even if the functional domains contributed by the two fusion partners are currently known, there is a lack of mechanistic insight linking *CIC-DUX4* with CDS sarcomagenesis. By inducing the expression of a plasmid construct containing *CIC-DUX4* in NIH 3T3 and U2OS cells, M. Kawamura-

Saito *et al.* showed that the expression of the majority of CIC target genes and especially *PEA3* family genes, is strongly induced by CIC-DUX4. By further analyzing the binding sites of CIC-DUX4, the authors identified the octameric 5'-TGAATGAA-3' CIC binding sequence in the promoters of all three *PEA3* family genes showing that the fusion protein retains the capacity to bind CIC binding sites (30). Few years later, T. Yoshimoto *et al.* developed an *ex vivo* mouse model that recapitulated the CIC-DUX4 tumor histology and showed an induction of CIC-DUX4 target genes. By microarray analysis they demonstrated that mouse CDS and mouse Ewing sarcoma display different gene expression profiles, providing additional evidences that the two tumors represent distinct sarcoma subtypes. Gene set enrichment analysis revealed that the mouse CDS gene expression profile is enriched for extracellular matrix organization genes and cyclin D1 signature which is consistent with the presence of abundant ECM in human CDS and the aggressiveness of the tumor. By further analysis, the authors identified Cyclin D2 and MUC5AC as potential biomarkers for CDS sarcoma diagnosis and showed that the cyclin-dependent kinases 4 and 6 inhibitor palbociclib and alkylating agent trabectedin could represent interesting therapeutic approaches for CDS treatment (80). Interestingly, ETV4 expression has been shown to be important for invasion and metastatization but not really for the primary tumor growth which seems to rely on other mechanisms such as activation of the CCNE-CDK2 axis (81). Another report demonstrated that inhibition of CDK2 by dinaciclib showed promising results *in vitro* and *in vivo*, in contrast to previous observations, no significant effect was reported upon palbociclib treatment (80)(81). Due to the lack of additional somatic driver mutations in CDS and the paucity of functional studies the development of additional targeted therapies against these tumors has not improved over the last years (82). However, it is to note that frequent chromosome 8 polysomy has been observed in a small cohort study, and associated with *MYC* amplification, but further studies are required to evaluate this feature as a potential diagnostic tool or therapeutic opportunity for CDS treatment (83). Although chromosome 8 gain was confirmed in another study, *MYC* amplification occurred only focally in one sample suggesting that this alteration is more likely to be subclonal (82). Frequent chromosome 1p loss has also been reported in a small CDS cohort but its relevance in CIC-DUX4 sarcoma pathogenesis still need to be further

investigated (82). To better understand CDS biology, primary and patient derived xenografts (PDX) cell lines have been successfully established over the past five years, giving hope for targeted therapy development in this extremely aggressive malignancy (30)(84)(85)(86).

Hypothesis and aims of the study

In this work, we aim to investigate the biological function of the CIC-DUX4 fusion protein by leveraging our collaboration with several investigators who made seminal scientific discoveries in the field of undifferentiated soft tissue sarcoma. Given the current knowledge in the field of translocation-associated sarcomas, we hypothesize that the CIC-DUX4 protein may act as an aberrant transcription factor (TF) by dysregulating the expression of several target genes that are crucial for cell transformation and sarcomagenesis, such as *PEA3* family genes. Since CDS harbors no additional driver mutations, we speculated that its pathogenesis may be driven by epigenetic changes directly or indirectly induced by the CIC-DUX4 fusion protein. To increase the translation impact of our work, we focused particularly on the capacity of the aberrant fusion protein to recruit co-factors and chromatin regulators at specific pro-tumorigenic target genes' regulatory elements. Since the p300 interaction domain is retained in the DUX4 fragment fused to CIC, we hypothesize that it may play an important role in the induction of the specific transcriptional program of CIC-DUX4 (Fig. 9) (64).

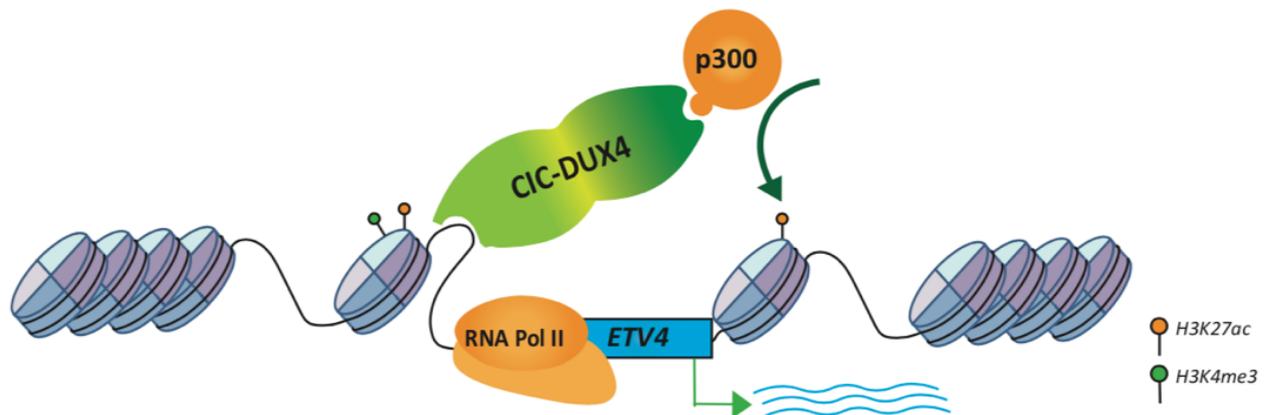


Fig. 9 Potential CIC-DUX4 mechanism of action

CIC-DUX4 binds to DNA through CIC N-terminal DNA binding domain. Its C-terminal part contains the DUX4 transactivation domain that binds to p300 which acetylates the lysin 27 of H3 histone tail (H3K27ac). This acetylation process, the opening of the chromatin and the methylation of the lysin 4 of H3 histone (H3K4me3) facilitate the recruitment of RNA polymerase II at specific CIC-DUX4 target genes and transcription occurs.

Aim 1. To determine epigenetic landscape of primary CDS. In order to define CIC-DUX4 target genes, we used genome-wide epigenetic profiling by chromatin immunoprecipitation followed by sequencing (ChIP-seq) of four frozen CDS samples kindly provided by Prof. C. Antonescu (Memorial Sloan-Kettering Cancer Center, New York), and integrated these data with whole-genome expression

analysis from the same tumors, in order to define the global expression and epigenetic landscape driven by the translocation.

Aim 2. To characterize primary derived CDS cell lines in order to study CIC-DUX4 biology.

We characterized primary-derived and PDX-derived CDS cell lines for their CIC-DUX4 and direct target genes expression. We studied CIC-DUX4 subcellular localization and obtained promising results for their epigenetic landscape assessment by CHIP-seq. We demonstrated that those CDS cell lines are also relevant models for *in vivo* CDS study.

Aim 3. To develop an inducible CIC-DUX4 expression system in a permissive human cell line.

We studied the permissiveness of hpMSCs and 293T cells for CIC-DUX4 expression and assessed its consequences on target genes expression and chromatin profile reprogramming. We demonstrated that induction of CIC-DUX4 expression in 293T cells represents a promising model to investigate the translocation binding sites and associated epigenetic remodeling events.

Aim 4. To identify a druggable target for CDS treatment. We uncovered the interaction between p300 and CIC-DUX4 in CDS and demonstrated a high sensitivity of CDS cell lines for p300 inhibition. We showed the effect of p300 inhibition on CIC-DUX4 and its target genes expression and demonstrated the specificity of this approach to treat CDS tumors. Finally, we obtained promising results of *in vivo* p300 inhibition in mice bearing CDS tumors.

Results

Whole genome chromatin profiling of four frozen CDS primary samples reveals active chromatin features at CIC-DUX4 direct target genes' promoters and enhancers

In order to assess the clinical relevance of our hypothesis, we obtained four frozen tissue samples of human CIC-DUX4 sarcomas (CIC-DUX4.1-4), provided by Prof. Cristina Antonescu from the Memorial Sloan-Kettering Cancer Center in New York, which were already characterized by RNA-seq. Since we hypothesize that CIC-DUX4 has a major impact on whole genome chromatin organization and thus on gene expression, we assessed the chromatin landscape of the four tumors by ChIP-seq. We focused mainly on H3K27ac histone mark which is linked to both active enhancers and promoters, and on H3K4me3 histone mark more specific for promoters. Figure 10 shows ChIP-seq tracks of one representative CDS sample revealing that the promoters of well-established CIC-DUX4 target genes such as *ETV1*, *ETV4* and *ETV5* are highly enriched for both H3K27ac and H3K4me3 histone marks. A particular region within *ETV1* gene shows an enrichment for H3K27ac only, which is linked with an active intragenic enhancer. It is to note that other MAPK target genes such as *DUSP4*, *DUSP6* and *SPRY4* also harbor both H3K27ac and H3K4me3 marks at their promoters.

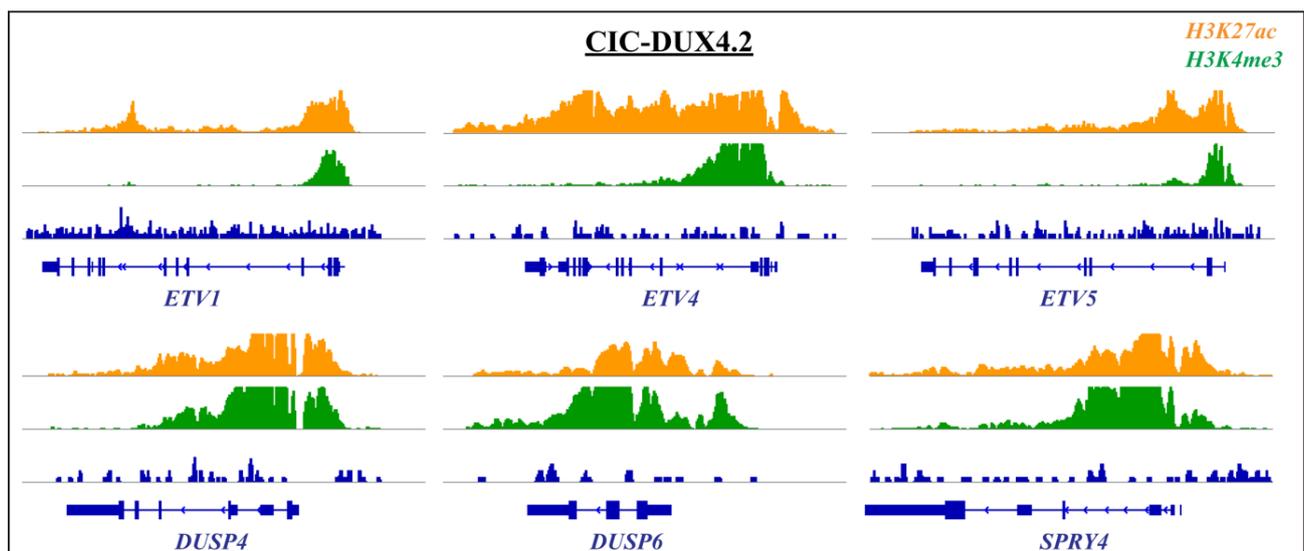


Fig. 10 CIC-DUX4 primary tumor chromatin profile

H3K27ac and H3K4me3 ChIP-seq profiling of one CIC-DUX4 tumor (CIC-DUX4.2) for 6 CIC-DUX4 target genes.

Thus, histones in the proximity of CIC-DUX4 target genes show post translational modifications associated with active transcription. To gain insight on how these genes are regulated in CDS tumors and how they participate to the tumorigenic program established by CIC-DUX4, we are currently defining a complete set of potential targets by integrating ChIP-seq with expression data performed on these four primary CDS samples in order to functionally assess their contribution to CDS pathogenesis using *in vitro* and *in vivo* models.

CDS1, CDS2 and CDS3 patient-derived cell lines express CIC-DUX4 protein, have a distinct gene expression profile compared to Ewing sarcoma and induce SBRCT development *in vivo*

In order to investigate CDS cells biology, we took advantage of three human CDS cell lines. Two cell lines were directly derived from patients' tumors : the first one was established by Dr. M. Kawamura-Saito (30) and is referred as CDS1 hereafter, the second one (CDS3) was established by Dr. M. Yoshimatsu (85). The third cell line (CDS2) was derived from a patient derived xenograft (PDX) and established by Dr. R. Oyama (87).

The CIC-DUX4 fusion protein was detected in all three cell lines as shown by western blot using an antibody directed against the DUX4 protein, which recognizes the C-terminal part of CIC-DUX4 (Fig. 11A). As already reported in the literature, CIC-DUX4 was observed around 250kDa, similarly to 293T cells transfected with a CIC-DUX4 expression vector containing the fusion between CIC-S and DUX4. Intriguingly, a second band of higher molecular weight of was observed in all three cell lines suggesting the possible expression of a longer isoform of CIC-DUX4. To assess the levels of CIC-DUX4 targets transcripts in our cell lines, we performed RT-qPCR for PEA3 family genes in addition to ERG and WT1, which are used in clinical settings for SBRCTs differential diagnosis (Fig. 11B). PEA3 family genes were all highly expressed in CDS cell lines compared to Ewing cell lines. ERG is expressed in 2/3 of CDS cell lines but not in Ewing and WT1 is also very highly expressed in 2/3 CDS cell lines compared to Ewing cell lines, thus validating already reported results on ERG and WT1 expression in CDS and Ewing sarcoma. These results support the fact that CDS and Ewing sarcoma are two different entities with specific gene expression profiles.

Next, we assessed the tumorigenic potential of two cell lines (CDS1 and CDS2), by injecting bilaterally 1, 2 or 5 millions of cells in the subcutaneous layer of the suprascapular region of NOD-SCID gamma KO mice, in order to evaluate the minimum number of cells required for tumor development *in vivo*. Tumor growth was observed at every sites of injection even with the lowest number of injected cells, and mice were sacrificed before tumors reached 1cm³, two to three months after injection. Tumors were extracted, weighed and embedded in paraffin in order to perform standard H&E staining. Of interest, histological analysis revealed the typical small blue round cell cytomorphology organized in sheets and layers of cells reminiscent of human CDS (Fig. 11C).

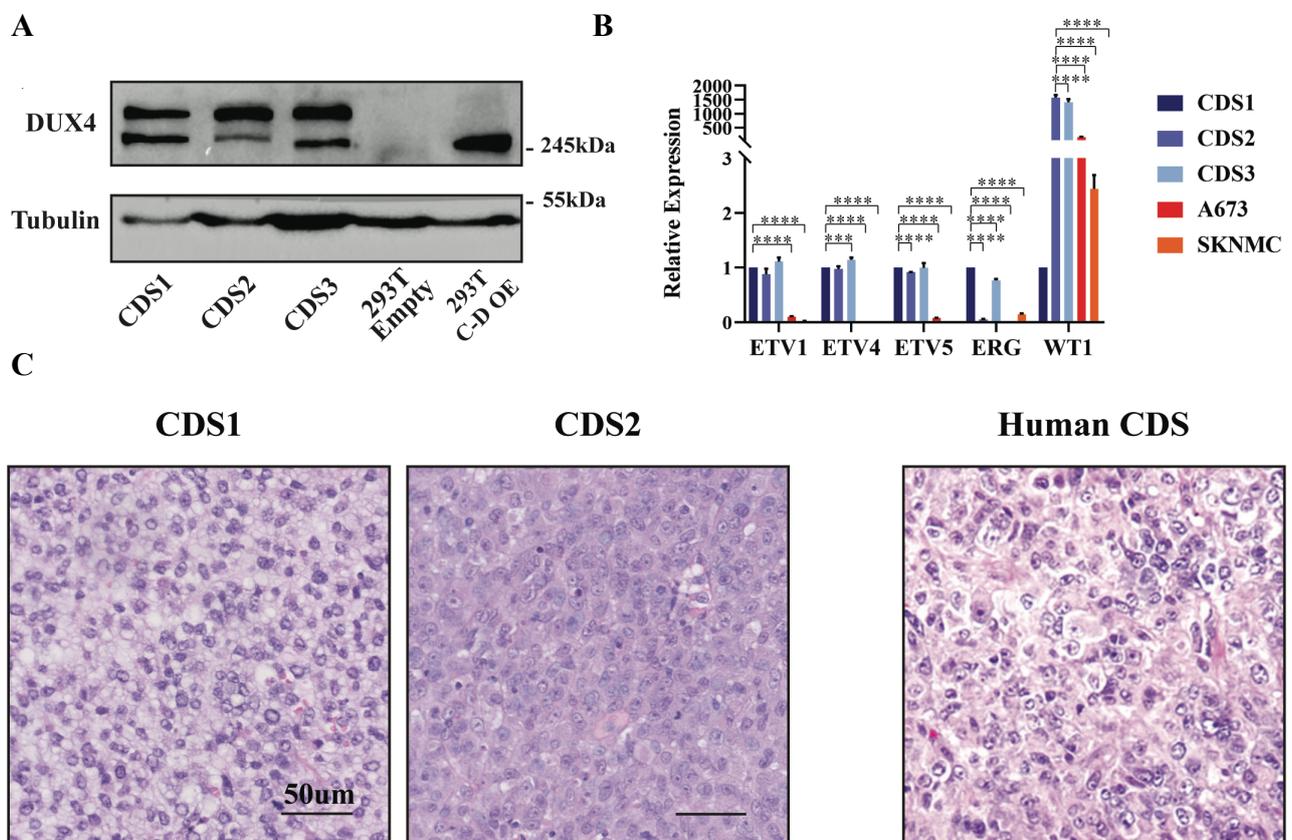


Fig. 11 CDS cell lines characterization

A. Western blot showing the expression of CIC-DUX4 in CDS1, CDS2 and CDS3 cell lines compared to 293T expressing either an empty vector (Empty) or a CIC-DUX4 containing vector (C-D OE). **B.** RT-qPCR analysis showing high PEA3 family genes expression in CDS cell lines and variable ERG and WT1 expression compared to Ewing cell lines. Data are presented as mean \pm SD, with n=3 per group. Unpaired, independent groups of 2 were analyzed by Student's *t* test. **P* = 0.01 to 0.05, ***P* = 0.001 to 0.01, ****P* = 0.0001 to 0.01, *****P* < 0.0001. **C.** H&E staining of CDS1 and CDS2 derived tumors after injection in NSG mice compared to human CDS (modified from "Round cell sarcomas beyond Ewing: emerging entities" by C. Antonescu, 2014, *Histopathology*, 64, 26-37. DOI:10.1111/his.12281).

CDS 1, 2 and 3 cells thus represent relevant models to study CIC-DUX4 sarcoma *in vitro* and *in vivo* since they express CIC-DUX4 fusion protein as well as a panel of its well-established target genes,

and are able to induce the development of sarcomas in mice recapitulating the histological features of their parental human tumors.

The CIC-DUX4 protein is localized in the nucleus and cytoplasm of primary CDS tumors and cell lines, generating heterogeneous subpopulations of tumor cells

Since endogenous CIC protein and CIC-DUX4 share the majority of CIC-S amino-acid sequence, the distinction between the two proteins by an antibody-based approach is technically challenging. On the contrary, the endogenous DUX4 and the fusion protein only share a small portion of their amino-acid sequence, and may therefore represent a better candidate for antibody-based detection techniques. Moreover, the assessment of DUX4 protein expression in our cell lines revealed that CDS2 cells show almost no endogenous DUX4 protein expression whereas CDS1 and CDS3 cells appear to express the endogenous DUX4 at very low level (Fig. 12A). The absence of endogenous DUX4 protein in CDS2 thus allows anti-DUX4 antibodies to specifically detect only the CIC-DUX4 protein. Since CIC-DUX4 is believed to function as a transcription factor and bind to chromatin, we first performed DUX4 immunofluorescence (IF) on CDS1 and CDS2 cell lines to assess CIC-DUX4 nuclear expression. Intriguingly, whereas CIC-DUX4 is expressed at a low level in a fraction of CDS1 (data not shown) and CDS2 cells' nuclei, it also appears to be present in the perinuclear region and diffusely in the cytoplasm (Fig. 12B). To assess if this heterogeneity in CIC-DUX4 expression is due to cell culture conditions, we performed immunohistochemistry (IHC) of CDS cell lines-derived tumors embedded in paraffin as a control. Remarkably, DUX4 IHC showed a very heterogeneous staining with foci of high nuclear CIC-DUX4 expression and other areas which lacked CIC-DUX4 expression or showed only cytoplasmic positive signal. These results demonstrate that CIC-DUX4 subcellular heterogeneity is also retained *in vivo* (Fig. 12C). Finally, to make sure that these observations were also relevant for human CDS tumors, we performed DUX4 IHC on a human CDS tumor and observed that this signal pattern was also observed in primary CDS human sample (Fig. 12C).

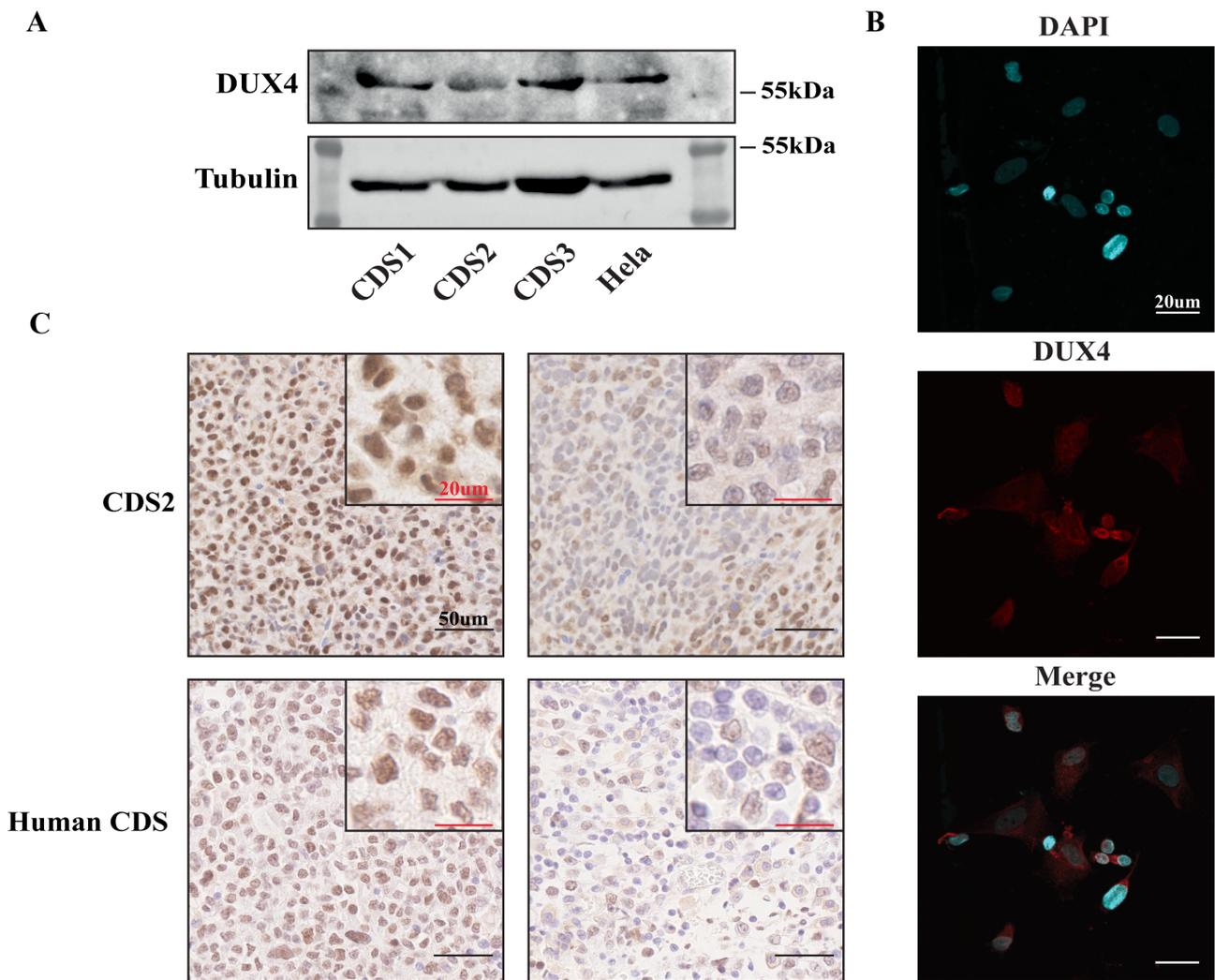


Fig. 12 CIC-DUX4 inter and intra cellular heterogeneity

A. Western blot showing DUX4 endogenous protein expression levels in CDS cell lines compared to HeLa cells. **B.** DUX4 immunofluorescence on CDS2 cell line showing CIC-DUX4 subcellular distribution. Blue = DAPI and red = DUX4. **C.** DUX4 immunohistochemistry in CDS2 derived tumor and human CDS tumor showing *in vivo* CIC-DUX4 signal heterogeneity. Left panels = high nuclear DUX4 signal. Right panels = low nuclear DUX4 signal.

KO serum-complemented cell culture conditions allow for optimal ChIP-seq profiling of CDS cell lines, which show a chromatin landscape similar to the one obtained in primary CDS tumors

The assessment of transcription factor occupancy by ChIP-seq requires a significant number of cells. Given the fact that only a small proportion of CDS cells show CIC-DUX4 nuclear expression, we had to enrich our cell lines for this particular cell subpopulation to successfully assess CIC-DUX4 genomic binding sites. To do so, we first looked for any known mechanism of CIC-DUX4 nucleo-cytoplasmic shuttling or degradation in the literature but only found similar mechanisms for endogenous, non-fused, CIC protein. In several reports, CIC protein has been shown to shuttle from nucleus to cytoplasm in order to be degraded by the ubiquitin-proteasome pathway upon MAPK activation (55). Since CIC protein is almost completely conserved in CIC-DUX4 fusion (79), we

assumed that CIC-DUX4 protein stability could also be dependent on MAPK activation. We thus cultured CDS2 cells in 10% and 5% fetal bovine serum concentration conditions, as well as in 20% KO serum in order to have different growth factors concentrations and thus different activities of the MAPK pathway. We then performed ChIP-seq for DUX4 on CDS2 cells exposed to the varying serum concentrations either in standard adherent conditions or in ultra-low attachment plates, since CDS cells have shown to grow as spheres in low attachment conditions (87). As shown in Figure 13A, CIC-DUX4 binding frequency varies dramatically in different culture conditions, with the highest being observed when cells were grown in KO-complemented medium in standard adherence conditions. Thus, in order to assess CIC-DUX4 transcription factor function, we decided to perform all further experiments in KO conditions including the ChIP-seq profiling of CDS2 cell line (Fig. 13B).

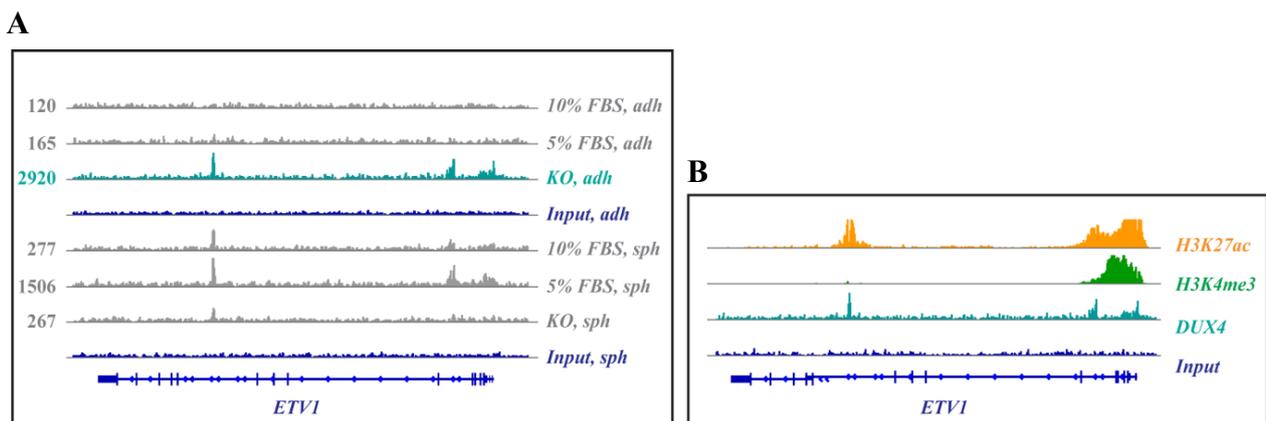


Fig. 13 CIC-DUX4 chromatin binding

A. CIC-DUX4 ChIP-seq in CDS2 cell line exposed to varying FBS concentrations or to KO serum in either standard adherence conditions (adh) or as spheres in low attachment plates (sph) for ETV1 locus. Numbers on the left show CIC-DUX4 genome wide peaks number for every condition. **B.** H3K27ac, H3K4me3 and CIC-DUX4 ChIP-seq profiling of CDS2 cell line in KO culture condition for ETV1 locus.

Before selecting optimal culture conditions for CIC-DUX4 chromatin recruitment, we performed ChIP-seq on CDS1 cell line cultured in standard 10% FBS conditions for H3K27ac, H3K4me3 as well as for DUX4, p300 and CIC. We obtained 781 DUX4 peaks which probably represent a fraction of all the possible binding sites for CIC-DUX4 due to the sub-optimal culture conditions. In comparison, ChIP-seq for DUX4 on CDS2 cells grown in KO-complemented medium showed 2920 DUX4 peaks. However, this experiment showed that CDS1 cells harbor a similar epigenetic profile as CDS primary tumors. Indeed, when CDS1 cells DUX4 ChIP-seq peaks were used as reference for CIC-DUX4 binding sites, and associated with the corresponding histone marks, we observed a

comparable signal pattern between CDS1 cells and CIC-DUX4 primary tumors (Fig. 14A). Interestingly, we also detected a clear colocalization of CIC-DUX4 and p300 signal at specific CIC-DUX4 active target sites, suggesting a possible interplay between those two proteins (Fig. 14B). We also performed CIC ChIP-seq, which showed an overlapping signal with DUX4 and p300. However, this signal is not specific for CIC-DUX4 since it may also recognize the endogenous CIC protein. Additionally, since endogenous DUX4 protein is slightly expressed in CDS1 cell line, we analyzed the chromatin states at genomic loci of known DUX4 target genes to assess the potential relevance of DUX4 expression in our cell lines. The complete absence of active histone marks at well-established DUX4 targets loci (like ZSCAN4, shown in Fig. 14C) in CDS1 cells suggests that the low amount of DUX4 protein expressed in these cells may not be sufficient to induce a significant transcriptional effect, as reported in other models of DUX4 expressing cells. Even if these preliminary results show similarities between the CDS1 cell line and primary tumor epigenetic profiles, this analysis would greatly benefit from the optimization of cell culture conditions, as seen at the beginning of this section. Thus, the same approach is currently being repeated on CDS1 and CDS2 cells cultured in KO-complemented medium.

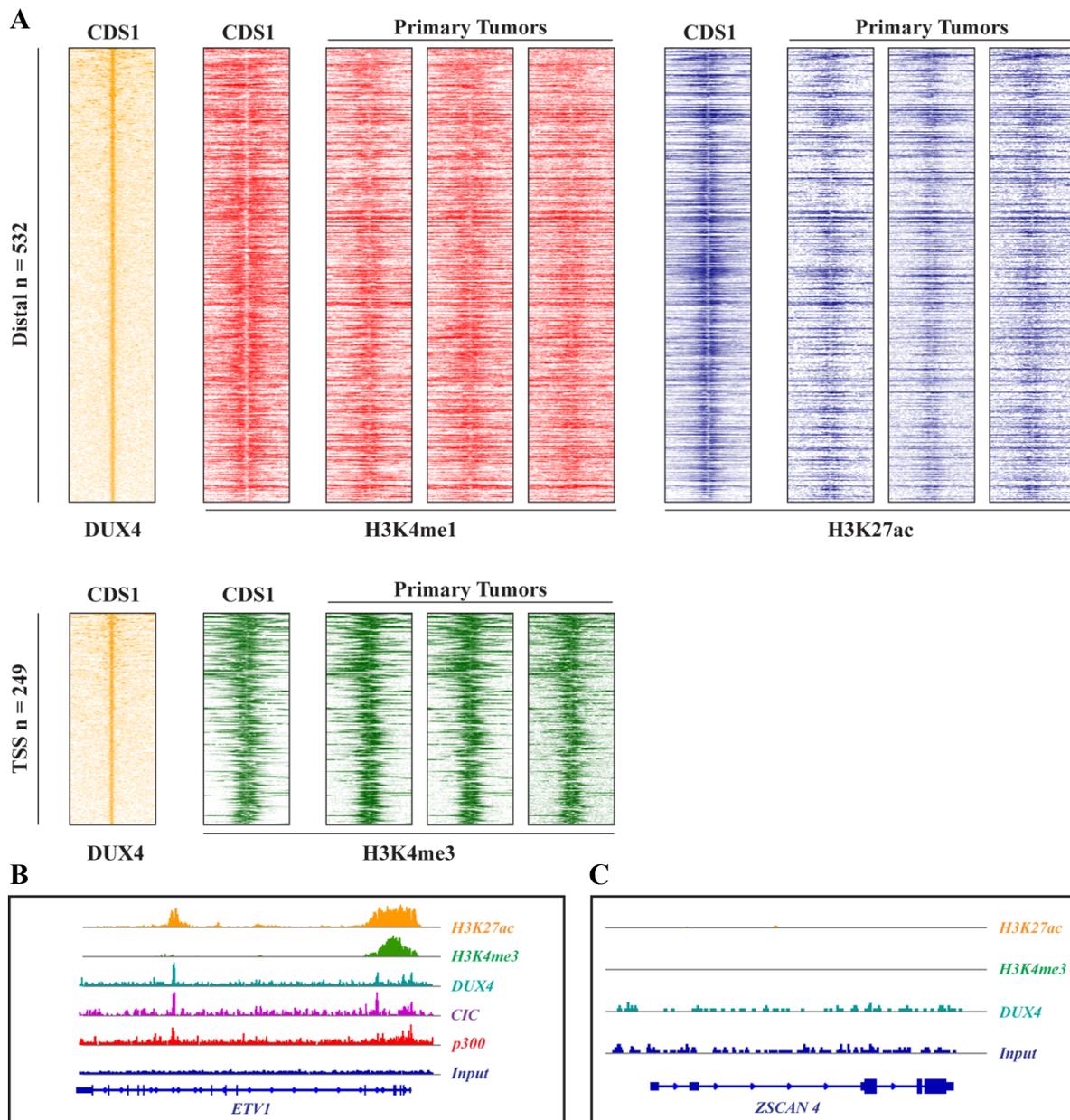


Fig. 14 CDS1 ChIP-seq profiling

A. CDS1 and primary tumors analysis. Heatmaps depict CIC-DUX4, H3K4me1, H3K27ac signal intensities for 532 CIC-DUX4-bound distal regulatory elements ranked by overall signal intensities of H3K4me1 and H3K27ac (upper panels). Heatmaps depict CIC-DUX4 and H3K4me3 signals for 249 CIC-DUX4 peaks overlapping with transcriptional start sites (TSS) ranked by overall intensities of H3K4me3 (lower panels). **B.** H3K27ac, H3K4me3, DUX4, CIC and p300 ChIP-seq profiling of CDS1 cell line in FBS culture condition for ETV1 locus. **C.** H3K27ac, H3K3me3 and DUX4 ChIP-seq profiling of CDS1 for ZSCAN4.

Induction of CIC-DUX4 expression in human cells leads to increased expression of CIC-DUX4 target genes depending on recipient cell's permissiveness

Since human MSCs have been shown to be permissive for specific fusion gene expression such as EWS-FLI1 (Ewing sarcoma) (88), we hypothesized that this cell type may also represent a relevant model to study the initiation of the CIC-DUX4 oncogenic program. We took advantage of a retroviral vector containing a N-terminal FLAG tagged *CIC-DUX4* fusion gene kindly provided by Dr. Nakamura (Japanese Foundation for Cancer Research, Tokyo), that we subcloned in a “Tet-on”

inducible lentiviral vector system (pINDUCER 20). Then, we infected hpMSCs with lentiviruses either containing our *FLAG-CIC-DUX4* construct or the same vector without our gene of interest as a control. After 7 days of induction of CIC-DUX4 expression with doxycycline (100ng/ml), cells were harvested and their CIC-DUX4 expression levels, as well as specific target genes induction were assessed. The fusion protein expression was successfully induced and PEA3 target gene levels were increased in FLAG-CIC-DUX4-infected hpMSCs (Fig. 15A-B, left). However, no changes were observed in terms of histone modifications and no CIC-DUX4 DNA binding signal was observed using ChIP-seq technique (Fig. 15C, left). By performing immunofluorescence (IF) on CIC-DUX4 induced hpMSCs we observed that only a very small fraction of cells was actually expressing the fusion, likely explaining why our ChIP-seq experiment was not successful (Fig. 15D, left). This shows that the hpMSCs primary cell line that we used is poorly permissive to CIC-DUX4 expression in these culture conditions. It is of note that the frequency of CIC-DUX4 expressing hpMSCs didn't change when these cells were cultured in KO serum-complemented medium in absence of growth factors (data not shown).

To obtain a better signal, we then performed the same infection experiment using the 293T cell model, a well-established cell line known to tolerate the expression of oncogenes. CIC-DUX4 protein expression was induced and CIC-DUX4 target genes transcripts were more expressed in CIC-DUX4 infected cells compared to control (Fig. 15A-B, right). Accordingly, 293T cells showed a strong increase in the H3K27ac histone mark at specific genomic loci bound by CIC-DUX4 in the CDS1 cell line (Fig. 15C, right). The ChIP-seq for DUX4, FLAG and p300 are currently being sequenced. We also performed DUX4 IF on 293T cells expressing CIC-DUX4 and identified a higher fraction of 293T cells showing strong DUX4 nuclear positive signal compared to hpMSCs (Fig. 15D, right). Even if in 293T where CIC-DUX4 is overexpressed (C-D OE) this represent only a fraction of all infected cells, it was sufficient to observe global epigenetic changes using ChIP-seq. Based on these results, we showed that different cell types harbor variable permissiveness to CIC-DUX4 expression

and that CIC-DUX4 induction is sufficient to reprogram the epigenome of cells and establish a specific CDS epigenetic signature.

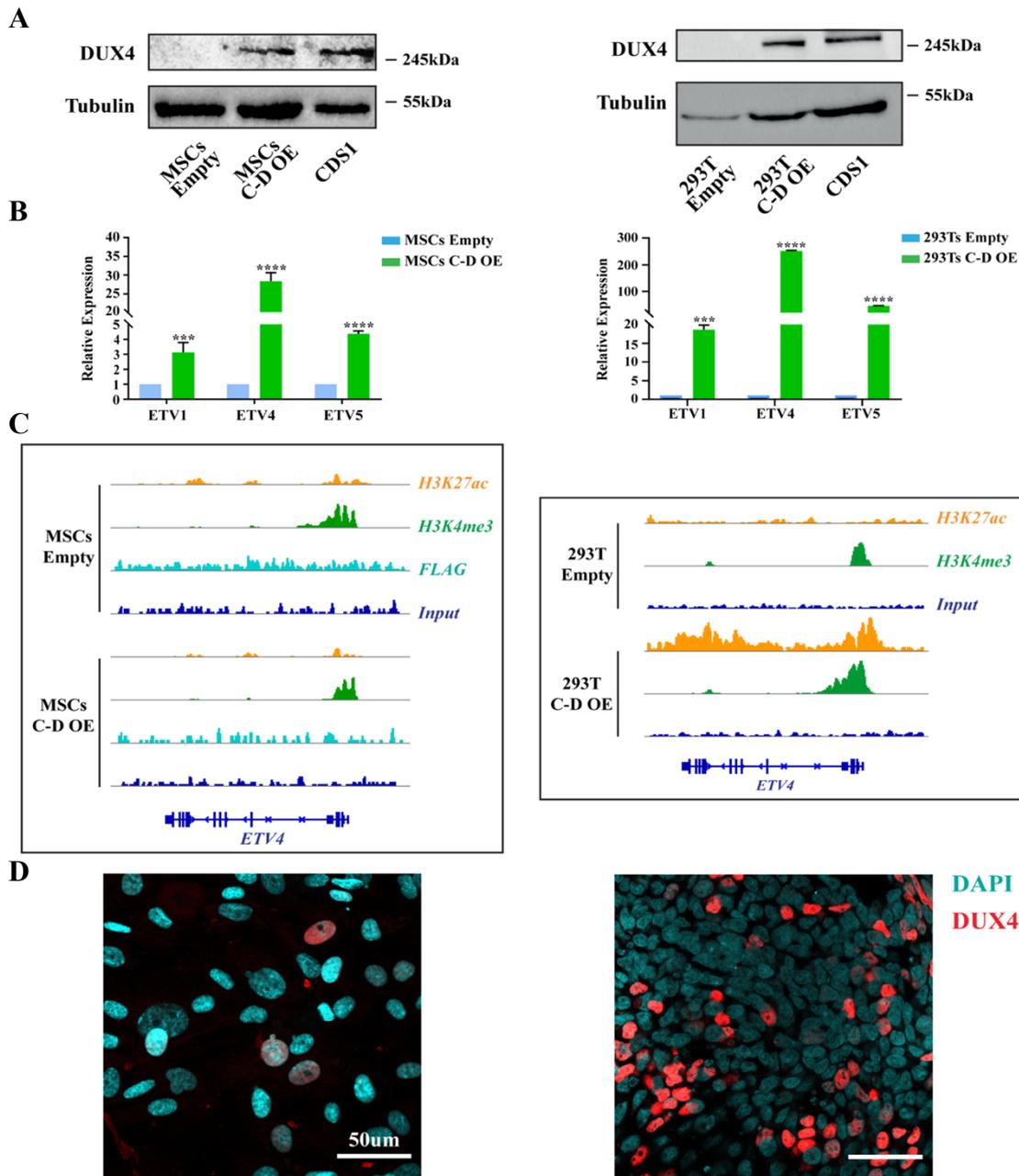


Fig. 15 CIC-DUX4 induction

A. Western blot showing the presence of CIC-DUX4 protein in induced hpMSCs (left) or 293T (right) compared to CDS1 cell line. **B.** RT- qPCR showing the induction of *PEA3* genes expression in control cells infected with an empty vector compared to cells infected with CIC-DUX4 containing vector. Left is showing hpMSCs experiment and right is showing 293T experiment. Data are presented as mean \pm SD, with $n=3$ per group. Unpaired, independent groups of 2 were analyzed by Student's *t* test. * $P = 0.01$ to 0.05 , ** $P = 0.001$ to 0.01 , *** $P = 0.0001$ to 0.01 , **** $P < 0.0001$. **C.** ChIP-seq for H3K27ac, H3K4me3 and Flag showing no differences between control hpMSCs and CIC-DUX4 expressing hpMSCs (left). ChIP-seq for H3K27ac and H3K4me3 showing the apparition of H3K27ac histone mark at *ETV4* promoter in 293T expressing CIC-DUX4. **D.** Immunofluorescence for CIC-DUX4 on either hpMSCs C-D OE (left) or 293T C-D OE (right).

The CIC-DUX4 and p300 protein-protein interaction is mainly localized in the nucleus

The established requirement of p300 recruitment by wild type DUX4 for target genes activation (64) and the ChIP-seq signal colocalization for CIC-DUX4 and p300 in the CDS1 cell line (Fig. 14B) suggest a possible interaction between these two proteins. Thus, after performing IF for p300 and DUX4 in CDS2 cells (Fig. 16A), we assessed the possible CIC-DUX4 and p300 interplay using proximity ligation assay (PLA) (Fig. 16B). PLA is a powerful technique that couples the detection of two proteins of interest with antibodies raised in different species, labeled with nucleotide probes that allow PCR amplification only if they are in close proximity (<40 nm). PLA thus not only demonstrate protein-protein interactions, but also visualizes the subcellular localization where this interaction

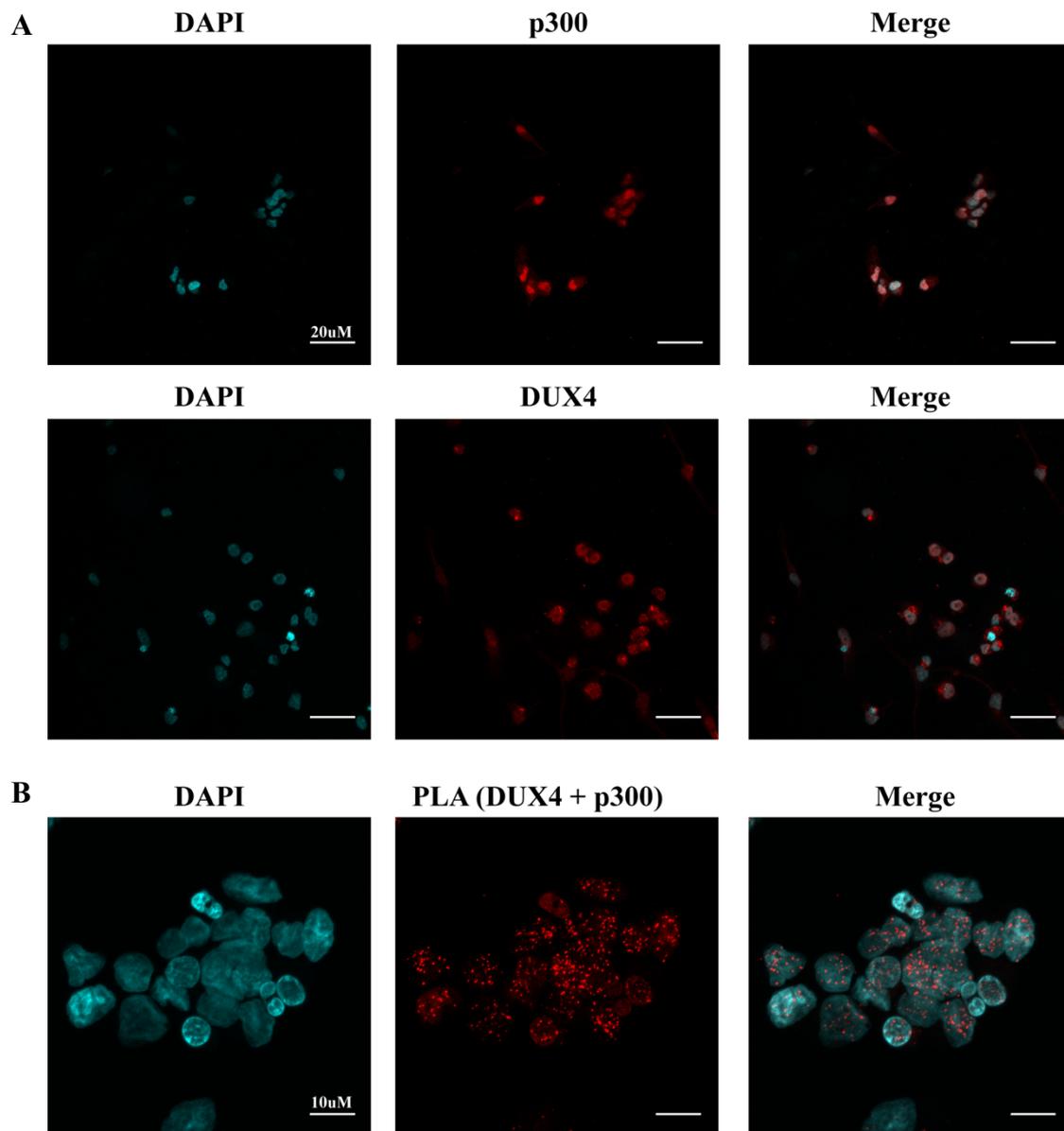


Fig. 16 CIC-DUX4 and p300 interaction

A. Upper panels : immunofluorescence of p300 in CDS2 cell line. Lower panels : immunofluorescence for DUX4 in CDS2 cell line. **B.** PLA using anti-DUX4 and anti-p300 antibodies in CDS2 cell line showing frequent CIC-DUX4 and p300 interactions inside the nuclear space.

occurs. Fig. 16B shows a strong and frequent interaction between CIC-DUX4 and p300 in the CDS2 cell line, which is mainly localized in the nuclei, suggesting that those two proteins are frequently interacting together and binding to the same chromatin loci as suggested by the ChIP-seq tracks from Figure 14B.

Inhibition of p300 enzymatic activity in CDS1, CDS2 and CDS3 cell lines strongly reduces cell proliferation and tumor growth, and specifically downregulates CIC-DUX4 target genes

Since the CIC-DUX4 and p300 interplay may be important for establishing the CDS tumorigenic program, we reasoned that blocking the p300 catalytic activity could strongly impact tumor cells survival. To assess the sensitivity of our CDS cell lines to p300 inhibition, we took advantage of the recently discovered specific inhibitor, A-485, which was used for an *in vitro* screening on 124 cancer cell lines for 3 (74 cell lines), 4 (13 cell lines) and 5 days (37 cell lines) of treatment (89). The authors retained a threshold of 1 μ M below which a cell line was defined as “sensitive” to A-485. We tested our CDS cell lines using the same experimental conditions as in the published study, which allowed us to combine our results with the data previously generated. By drawing a dose response curve and normalizing our results to a positive control treated with the strong chemotherapeutic agent staurosporine, we calculated an absolute effective concentration 50 (EC₅₀) for our three cell lines. CDS2 cells showed the highest sensitivity to A-485 with an EC₅₀ of 330nM after five days of treatment (Fig. 17A), whereas CDS1 and CDS3 cells displayed an EC₅₀ of 987nM and 972nM, respectively. When treated for 5 days and added to the 37 cell lines reported in the published screening, our CDS cell lines are all included in the top 15 sensitive cancer cell lines at rank 15, 13 and 4 for CDS1, CDS3 and CDS2 cells, respectively (Fig. 17B). By performing RT-qPCR on CDS2 and CDS3 cell lines after 5 days of A-485 treatment, we demonstrated that CIC-DUX4 mRNA shows little decrease at higher doses and that ETV1, ETV4 and ETV5 targets are downregulated (Fig. 17C). This effect may underlie the observed sensitivity of CDS lines toward p300 inhibition.

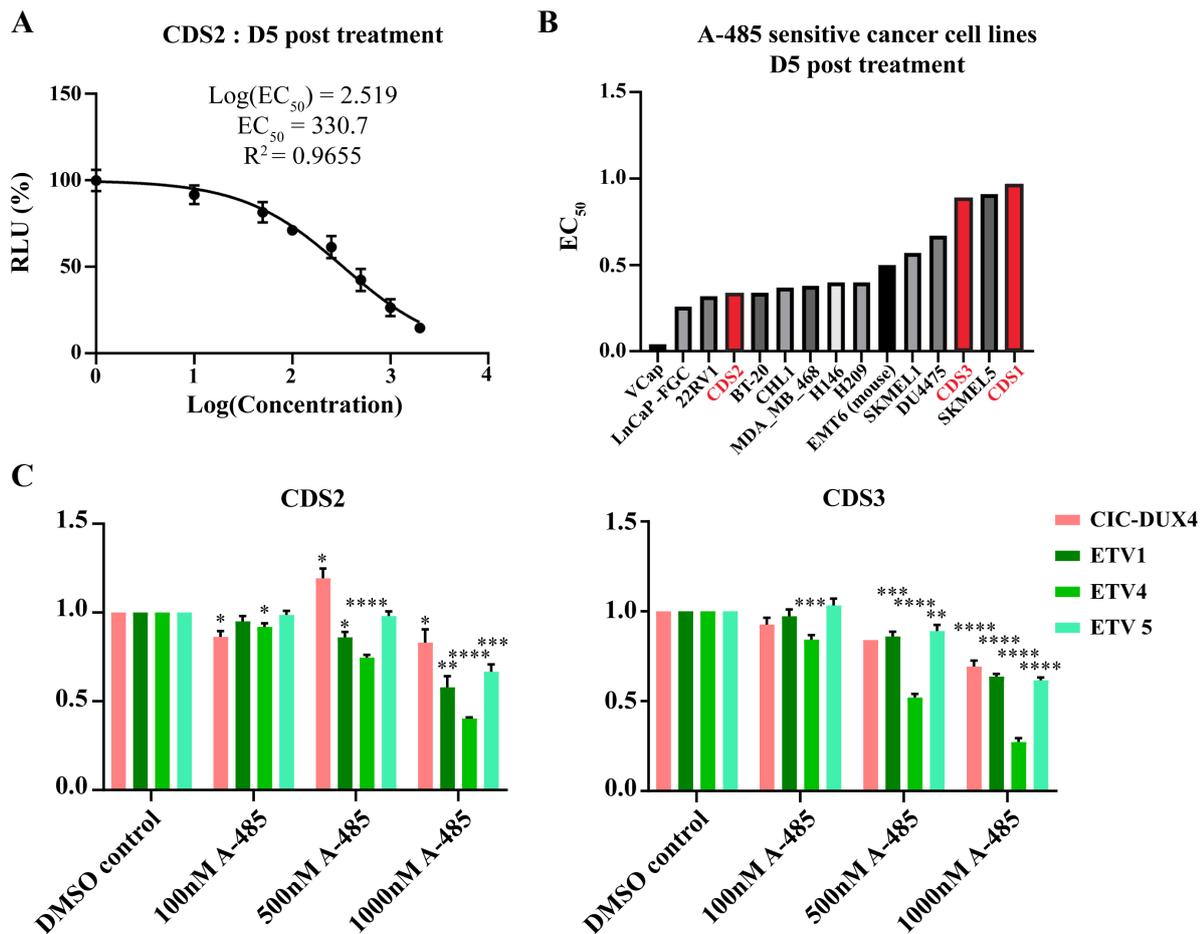


Fig. 17 *in vitro* p300 inhibition

A. Dose response curve of CDS2 after 5 days of treatment with A-485. B. A-485 CDS cell lines sensitivity compared with the most sensitive cancer cell lines from the published study ranked by their EC_{50} after 5 days of A-485 treatment. C. RT-qPCR showing CIC-DUX4 and PEA3 family genes transcriptional levels after 5 days of treatment with increasing A-485 concentrations. Data are presented as mean \pm SD, with $n=3$ per group. Unpaired, independent groups of 2 were analyzed by Student's *t* test. * P = 0.01 to 0.05, ** P = 0.001 to 0.01, *** P = 0.0001 to 0.01, **** P < 0.0001. Each condition is compared to its corresponding DMSO control.

p300 is crucial for CIC-DUX4 but not endogenous CIC or DUX4 proteins stability

Of great interest, although CIC-DUX4 transcript levels were only slightly affected by p300 pharmacological inhibition, CIC-DUX4 protein almost completely disappeared at day 5 after the initiation of A-485 treatment in a dose dependent manner in all three cell lines (CDS2 cells shown in Fig. 18A). This suggests that p300 enzymatic activity may be involved in CIC-DUX4 protein stabilization. Interestingly, neither endogenous CIC nor DUX4 proteins levels seemed to be affected by p300 inhibition, as shown by western blot on A673 and Hela cells treated with $1\mu\text{M}$ A-485 after 5 days (Fig. 18B). Non-CDS cell lines were used for this experiment in order to be sure to observe only CIC and DUX4 endogenous proteins levels. To indirectly assess the specificity of CIC-DUX4 degradation in the CDS2 cell line after A-485 treatment, we incubated the western blot membrane

previously revealed with anti-DUX4 antibody with an anti-CIC antibody which binds to both CIC-DUX4 and endogenous CIC. The resulting stability of CIC signal can only be attributed to endogenous CIC protein since CIC-DUX4 is degraded as shown by the western blot using DUX4 antibody. This shows that endogenous CIC is also stable in A-485 treated CDS cells and that p300 inhibition specifically targets the stability of CIC-DUX4.

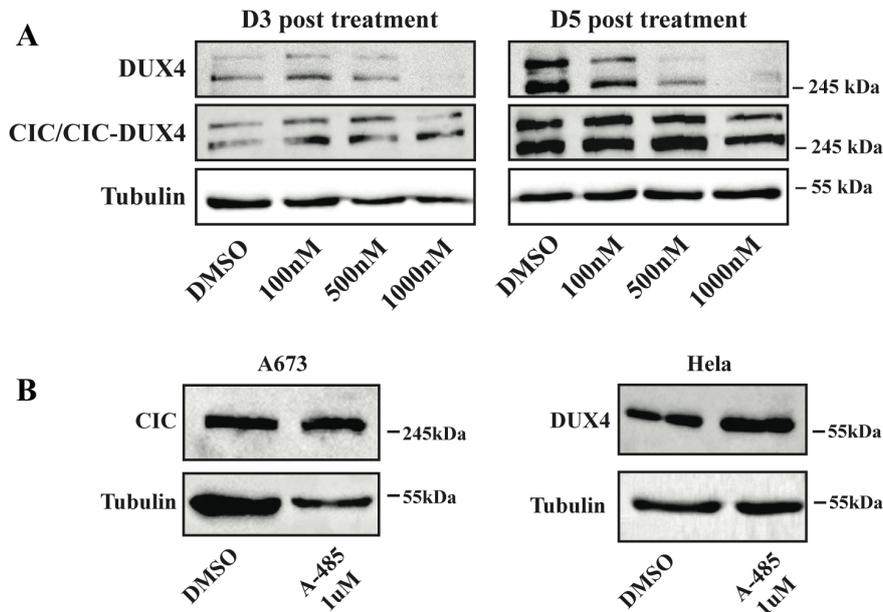


Fig. 18 p300 inhibition effect on CIC, DUX4 and CIC-DUX4 proteins

A. Western blot showing decreasing CIC-DUX4 and stable CIC protein levels in CDS2 cells after treatment with increasing concentrations of A-485 at 3 (left) and 5 (right) days post treatment **B.** Western blot showing the stability of CIC protein after 5 days of treatment of A673 cells with 1µM A-485 (left). Western blot showing the stability of DUX4 protein after 5 days of treatment of HeLa cells with 1µM A-485 (right).

***In vivo* p300 inhibition strongly reduces CDS tumor growth**

In order to validate the results obtained *in vitro* we injected subcutaneously 2 millions CDS2 cells bilaterally in the suprascapular region of 10 NOD-SCID gamma KO mice. 4 weeks later, once small tumors developed at all injection sites, we started intraperitoneal A-485 injections twice daily. 5 mice received 100mg/kg A-485 based on their weight at the first day of injection and 5 mice received the corresponding volume of carrier without A-485. Based on available *in vivo* studies using A-485, our initial plan was to continue the same treatment regimen for 15 days, and generate a survival curve. Unfortunately, we had to stop the experiment after 12 days due to a 10-20% weight loss of the mice receiving the A-485 treatment, and technical issues regarding the length of the needles used for the IP, which resulted in the loss of at least one treated mice. As a consequence, control and treated mice were sacrificed sequentially during 4 days as shown on the timeline of Figure 19A. However, despite the early interruption of the treatment, we observed a strong reduction in both tumor volume and size

compared to control mice (Fig. 19B). Given this promising preliminary result, the experiment will be replicated using 50mg/kg A-485 as initial dose, and adapting it to the weight of each mice throughout the treatment in order to reduce side effects.

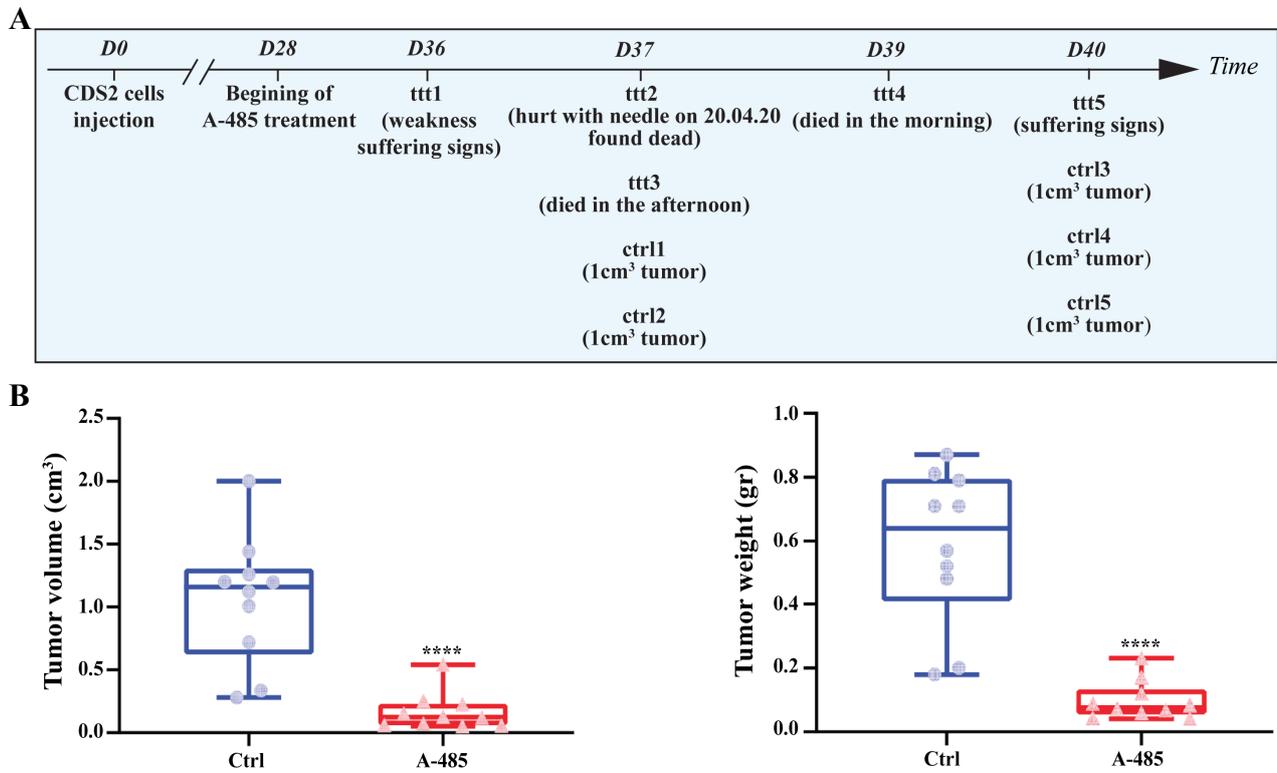


Fig. 19 *in vivo* A-485 treatment of CDS2-derived tumors

A. Timeline showing the date of mice sacrifices. Ttt = treated mouse. Ctrl = control mouse. () = reason of sacrifice **B.** Left panel: tumor volume in cm³ for each tumor collected at the day of sacrifice. Right panel: tumor weight for each tumor collected at the day of sacrifice. Difference between tumor volumes and weights were analyzed by Mann-Whitney *U* test. **P* = 0.01 to 0.05, ***P* = 0.001 to 0.01, ****P* = 0.0001 to 0.01, *****P* < 0.0001.

Discussion and future directions

The undifferentiated sarcoma category encompasses a heterogeneous group of tumors displaying very different clinical and biological properties. Because of their elusive biological properties, a better characterization of these tumors may help improving their clinical management, based on the development of new diagnostic and therapeutic tools tailored on the specific molecular features of these tumors. Among the recent advances in this field, the discovery of the t(4;19)(q35;q13) recurrent translocation in a subtype of small round blue cell sarcomas, and the subsequent characterization of the resulting aberrant CIC-DUX4 fusion protein, represents an excellent opportunity to investigate the specific biological programs responsible for the initiation and maintenance of CIC-DUX4-translocated sarcomas. Given the current lack of therapeutic options for this aggressive disease, there is a tangible need to improve our understanding of CIC-DUX4 biology, and to develop more specific targeted therapies.

In this study, we assessed the genome wide epigenetic and transcriptional changes induced by CIC-DUX4, using primary tissue samples, tumor-derived cell lines and 293T cells in which exogenous CIC-DUX4 expression was induced, in order to decipher the oncogenic program established by the fusion protein. We discovered unexpected CIC-DUX4 biological features including the presence of the fusion protein in both the nuclear and cytoplasmic compartments of CDS cells, as well as the functional interaction between the translocation and the p300 chromatin regulator. The culture of CDS primary and PDX-derived cell lines allowed us to pinpoint a potential therapeutic target and greatly expand our knowledge regarding CIC-DUX4 sarcoma biology.

Although the rarity of CIC-DUX4 sarcoma limits the possibility to build a large experimental cohort of primary tumor samples, we succeeded in obtaining four frozen human primary samples for which gene expression data were provided by Prof. C. Antonescu (Memorial Sloan-Kettering Cancer Center, New York). By pairing these data with our genome wide chromatin analysis obtained by ChIP-seq on the same samples, we will be able to generate a new set of data describing the chromatin

and gene expression networks regulated by CIC-DUX4 in primary tumors. By integrating these results with the same genome wide analysis performed on our CDS cell lines and CIC-DUX4-expressing 293T cell model we will be able to define CIC-DUX4-bound regulatory elements such as promoters and enhancers and link them with gene expression changes induced by the translocation. Furthermore, we are currently testing a set of shRNAs targeting DUX4 in KO serum-complemented culture conditions in order to perform a CIC-DUX4 KD, which will provide us with critical information about the translocation direct target genes. Given the genetic features of DUX4, the feasibility of a CRISPR-Cas9 approach to knock-out DUX4 is technically challenging. The presence of the DUX4 gene as multiple repetitions on several distinct chromosomes does not allow the identification of the specific DUX4 copy fused to CIC, thus preventing its selective targeting by this technology. Further analyses, including the definition of CDS tumor super-enhancers and the discovery of master transcription factors involved in CDS cells' identity maintenance using DNA motif analysis, will provide us with the critical nodes on which the CDS gene expression program is relying. This set of information will be crucial for every future functional experiment on these models, since it will generate a blueprint for further comparisons and analyses. A similar unbiased functional approach was already successfully used for Ewing sarcoma (90), and will help us to define important candidates genes and pathways involved in CDS initiation and maintenance.

The very low prevalence of CIC-DUX4 sarcoma is also impacting the availability of relevant primary cell lines. However, we could benefit from the establishment of patient-derived cell lines, as well as PDX derived cell lines, kindly provided by Dr. M. Kawamura-Saito, Dr. M. Yoshimatsu and Dr. R. Oyama, who generated these models. The tumorigenicity of these cell lines and their histological resemblance to primary CDS tumors suggest that these lines can be used as relevant *in vitro* and *in vivo* experimental models to study CIC-DUX4 sarcoma. We are also characterizing a recently received fourth CDS cell line from Dr. S. Nakai from the orthopaedic surgery department of Osaka, Japan, in a similar way to the other three CDS cell lines already available.

The presence of the CIC-DUX4 sarcoma fusion protein in CDS cell lines, alongside their specific gene expression signature compared to Ewing sarcoma cell lines, indicates that Ewing and CIC-DUX4 sarcoma are two different entities, as already suggested by their distinct prognosis and response to current multimodal therapies (24). Thus, the study of CIC-DUX4 sarcoma requires specific and tailored tools to assess CIC-DUX4 protein function and role in CDS pathogenesis. The only possibility to assess the expression of the CIC-DUX4 protein in CDS cell lines relies on using anti-DUX4 antibodies, since the endogenous DUX4 and CIC-DUX4 proteins display a distinct molecular weight of 45 and 250 kDa, respectively. On the contrary, the currently available CIC antibodies cannot discriminate CIC-DUX4 from CIC, since both proteins display a very similar molecular weight, and share an almost identical amino-acid sequence. However, it would be of great interest to explore the functional relation between the endogenous CIC and CIC-DUX4 proteins, and to understand why a panel of well-established CIC target genes are actively transcribed when both proteins are expressed. This knowledge could potentially help restoring the repressive activity of CIC on its target genes, and deconstruct the translocation-driven oncogenic program.

Western blot analysis on CDS cell lines using anti-DUX4 antibodies intriguingly revealed the presence of one band at 250kDa, already reported in the literature as the CIC-DUX4 protein, and another band of higher molecular weight. Given the fact that CIC can be present as two protein isoforms due to an alternative promoter site, the possibility of both CIC short (250kDa) and CIC long (400kDa) isoforms being fused to the DUX4 C-terminal part is interesting. This event would give rise to the generation of both CIC-DUX4-S and CIC-DUX4-L isoforms, which differ in their N-terminal part. Interestingly, wild type CIC-S and CIC-L proteins have been shown to be preferentially located in the cytoplasm or in the nucleus, respectively. The cytoplasmic CIC-S was shown to be present in close proximity to mitochondria, and its mutation led to an increase in 2HG levels in IDH mutant oligodendrogliomas (49). If these properties are retained by CIC-DUX4-S and CIC-DUX4-L, it is tempting to speculate that nuclear CIC-DUX4-L may mainly act as a transcription factor and cytoplasmic CIC-DUX4-S could probably play a role in cell metabolism, revealing a unique

synergistic mechanism of action never reported in other translocation-associated sarcomas. It is noteworthy that the induction of CIC-DUX4 expression in 293T cells with a vector that only contains the sequence coding for the 250kDa CIC-DUX4 protein isoform was able to generate a nuclear signal by immunofluorescence using an anti-DUX4 antibody, which suggests that the short isoform can also be expressed in the nucleus. Thus, the nuclear and cytoplasmic repartition of CIC-DUX4-S and CIC-DUX4-L proteins probably depends on additional factors than the difference in their N-terminal sequence. Nevertheless, the detailed understanding of both isoform's specific roles in CDS sarcomagenesis would greatly strengthen our knowledge of CIC-DUX4 sarcoma biology. Importantly, it has to be evaluated whether they are both required for CDS development and maintenance or if they play different roles at different stages of CDS pathogenesis. Additional experiments are still ongoing to prove the presence of the two CIC-DUX4 isoforms in CDS tumors, and assess their subcellular localization and function in CIC-DUX4 sarcoma cells. To do so, we are currently designing primers for RT-PCR experiment in CDS cell lines to specifically detect and amplify CIC-DUX4-S and CIC-DUX4-L. The sequence of the resulting amplicons will inform us about the existence of both isoforms in the CDS cell lines. In parallel, we are also developing a plasmid containing the CIC-DUX4-L sequence fused to a V5 tag, in order to perform overexpression experiments. We will take advantage of the fact that our current CIC-DUX4 overexpression construct is fused to a FLAG tag to selectively detect CIC-DUX4-S and CIC-DUX4-L by using anti-FLAG and anti-V5 antibodies, respectively. The sequential or concomitant induction of CIC-DUX4-S and CIC-DUX4-L expression in 293T and CDS cells will show us how both isoforms affect CIC-DUX4 target genes expression and how they are distributed in the subcellular compartments. We will also be able to study the function of each isoform as a transcription factor and as key players of other biological processes. Additionally, the presence of FLAG and V5 tags will also help us to discriminate CIC-DUX4 from the endogenous CIC isoforms.

Under standard adherent cell culture conditions containing 10% FBS, we observed that CIC-DUX4 is mainly expressed in the cytoplasm and less frequently in the nuclear compartment, raising the

question about the role of this translocation as a transcription factor. Interestingly, anti-DUX4 IHC in CDS2-derived mouse xenografts and in primary CDS samples showed a signal for CIC-DUX4 in both the nuclear and cytoplasmic compartments, confirming that this heterogeneity is also present *in vivo* and in primary tumors. These observations raise many fundamental questions about the biological function of CIC-DUX4 in both subcellular compartments. Is cytoplasmic CIC-DUX4 just a reservoir for quick cytoplasm-nuclear shuttling, an intermediate step before degradation in the same manner as for endogenous CIC, or does it serve distinct functions as suggested in the previous paragraph? It is however to note that this observation was not reported in previous studies where anti-DUX4 IHC was performed on CDS primary tumors as a diagnostic tool for CIC-DUX4 detection (28). To make sure that in our anti-DUX4 IHC we detected only CIC-DUX4 and not the endogenous DUX4 protein, we performed the same analysis on CDS2-derived xenografts, since this cell line don't express the endogenous DUX4 protein. Indeed, the low expression of endogenous DUX4 in CDS1 and CDS3 cell lines make them suboptimal models for this kind of approach due to the possible binding of the anti-DUX4 antibody to the endogenous protein. Moreover, the function of endogenous DUX4 in these cells still needs to be investigated in details. In addition, anti-DUX4 IHC needs to be performed in additional human samples to corroborate our results on CIC-DUX4 signal heterogeneity.

The investigation of CIC-DUX4 function as a transcription factor using the ChIP-seq technology is challenging for two main reasons. First of all, this technique requires high total cell numbers (from 5 to 10 millions cells per immunoprecipitation), which are difficult to obtain with CDS cell lines since the majority of these cells don't show nuclear CIC-DUX4 expression under standard culture conditions. Second, since we can rely uniquely on the anti-DUX4 antibody, only the CDS2 cell line represents an appropriate model for the translocation profiling, because of the lack of endogenous DUX4 expression in these cells. To circumvent this problem, based on the fact that wt CIC has been shown to be excluded from the nucleus and then degraded through the ubiquitin proteasome pathway upon MAPK pathway activation, we reasoned that reducing the growth factors concentration in the

culture medium could probably enhance CIC-DUX4 DNA binding frequency and stability. Yet, a tremendous improvement of CIC-DUX4 binding frequency was observed in KO serum under standard adherent conditions but not in low adherence culture conditions with the same KO complemented medium. This suggests that the presence of growth factors is not the only regulator of CIC-DUX4 subcellular localization and that cell adhesion may also influence this phenomenon, but this has to be further demonstrated. The assessment of CIC-DUX4 subcellular localization in CDS cells grown in KO-supplemented cell culture medium by using IF is technically challenging since CDS cells show limited attachment capacities in these conditions. Moreover, in addition to cell culture conditions, we are currently facing technical issues with ChIP-seq characterization of CDS cell lines. Recently, both profiling of transcription factors as well as chromatin regulators failed to produce satisfying results, and DUX4 signal improvement under KO-complemented cell culture conditions couldn't be replicated yet. These issues are most likely independent of CIC-DUX4 subcellular localization since p300 ChIP-seq is also giving sub-optimal results in CDS cell lines as well as in 293T overexpression studies while it should be expressed and bound to DNA even in absence of CIC-DUX4. To be able to rapidly solve this issue, we are currently testing different ChIP conditions, using qPCR as an output in order to reduce the costs and additional time linked with sequencing and bioinformatical analysis. Once we'll be able to obtain satisfying ChIP-qPCR results for DUX4, FLAG and p300 we will perform ChIP-seq analysis using the same antibodies in CDS cell lines and 293T C-D OE model. This step will be crucial for CIC-DUX4 direct target gene assessment.

Despite these difficulties, the ChIP-seq analysis performed on CDS1 cells in standard culture conditions revealed 781 DUX4 binding sites, predominantly located at transcription start sites, and to a lesser extent, at distant regulatory elements. Presence of both CIC and DUX4 signal was observed at promoters of CIC target genes in association with histone marks related to active promoter function. Since CIC usually represses these target genes, and endogenous DUX4 is not reported to bind these sites, the observed signal is most likely due to the presence of the CIC-DUX4 fusion protein. In

addition, these peaks locations are shared with the DUX4 peak signals obtained from CDS2 cell line ChIP-seq performed in KO serum condition. Thus, despite suboptimal cell culture conditions and the presence of endogenous DUX4 protein, CDS1 cells ChIP-seq shows similar CIC-DUX4 binding sites and chromatin modifications at CIC-DUX4 target loci compared to CDS2 cells cultured in KO conditions. However, it cannot be excluded that the presence of both endogenous CIC and DUX4 in the CDS1 cell line may contribute to the ChIP-seq peaks signal obtained using anti-DUX4 and anti-CIC antibodies. In addition, the possibility of CIC-DUX4 binding as a complex with endogenous CIC or DUX4 cannot be assessed here for the same reasons stated above. Yet, the overlap of H3K4me3, H3K27ac ChIP-seq signal in CDS1 cells, CDS2 cells and primary tumors at specific CIC-DUX4 target sites, and the enhancement of CIC-DUX4 binding frequency in KO serum-complemented culture conditions provide a promising framework to further investigate the regulatory network of CDS tumors.

The induction of a FLAG tagged CIC-DUX4 construct expression in human mesenchymal stem cells was performed in order to study newly established CIC-DUX4 regulatory elements, without using neither anti-CIC nor anti-DUX4 antibodies, which would help bypassing the limitations cited in the previous paragraph. hpMSCs were chosen since they showed to be the only permissive cell type for EWS-FLI fusion expression, and thus represent the most likely cell of origin of Ewing sarcoma (91). However, after inducing CIC-DUX4 expression in hpMSCs, we were not able to obtain a satisfying number of CIC-DUX4-expressing hpMSCs, since the fusion protein was expressed only by a very low fraction of these cells. CIC-DUX4 induction was only detectable by RT-qPCR and western blot techniques, but ChIP-seq analysis failed to reveal any changes in hpMSCs chromatin profiles after CIC-DUX4 induction. Concerning CIC-DUX4 target genes, RT-qPCR, which is a highly sensitive technique, was able to detect their expression induction in hpMSCs C-D OE. However, ChIP-seq, for which a consequent number of cells is required, didn't show any deposition of H3K27ac at CIC-DUX4 targets promoters. This discrepancy is most likely explained by the fact that the small fraction of hpMSCs C-D OE cells that are expressing CIC-DUX4 after doxycycline induction was not

sufficient to be detected by ChIP-seq. The ChIP-seq profile of these cells was most probably masked by the majority of cells that didn't express CIC-DUX4, and therefore displayed a profile similar to control cells. Moreover, since CIC-DUX4 target genes already show some basal levels of expression in hpMSCs infected with an empty vector, the H3K27ac induction has to be even more marked to be detected at these loci. This would not be the case in the situation where CIC-DUX4 target genes are transcribed *de novo* in the presence of the fusion protein. The same experiment performed in 293T cells successfully revealed a deposition of H3K27ac histone marks at the promoters of CIC-DUX4 target genes, pointing to their permissiveness for CIC-DUX4 expression and related chromatin remodeling. The finalization of this ChIP-seq analysis using anti-DUX4 and anti-FLAG antibodies will therefore provide us with the *de novo* established CIC-DUX4 binding sites, associated regulatory elements and gene expression changes.

Aberrant transcription factors are known to be hardly targetable using standard pharmacological strategies. Current approaches trying to bypass this limitation include modulation of downstream regulated target genes, and targeting the catalytically active binding partners of transcription factors. Thus, since DUX4 is capable to recruit p300 through its C-terminal last 100 amino acids, which are retained in the CIC-DUX4 fusion, p300 blockade may represent an interesting opportunity to repress CIC-DUX4 activity and target gene induction. The colocalization of p300, CIC and DUX4 ChIP-seq signals in CDS1 cells at CIC-DUX4 target genes promoter suggests that CIC-DUX4 and p300 are both present at those loci. The close proximity of CIC-DUX4 and p300 in CDS cells nuclei confirmed by PLA further strengthen the hypothesis of a dependency of CIC-DUX4 on p300 activity for target gene activation. Importantly, all CDS cell lines tested were sensitive to p300 inhibition *in vitro* and, despite technical difficulties, we also observed a strong reduction of tumor volume and weight after A-485 treatment of CDS-bearing mice *in vivo*. Further analysis of CDS-derived xenografts after p300 inhibition will provide us with additional information about the impact of p300 inhibition on CDS tumor architecture and histology. Frozen fragments of A-485-treated tumors will be assessed for the expression levels of a panel of CIC-DUX4 target genes, and compared to untreated CDS2-derived

xenografts to validate the results we obtained *in vitro*. Despite the development of drug related toxicity in A-485 treated mice, the dramatic effect observed on tumor size and weight suggest that a repeating this experiment using a lower dose would likely be successful in reduce the tumor burden with limited side effects. Since p300 is ubiquitously expressed in human cells, our hope is that the strong dependency of CDS cells on the function of p300 may generate a therapeutic window that will allow us to blunt tumor growth without perturbing the physiological function of p300 in normal cells.

In addition to pharmacological inhibition, we are currently developing shRNAs against p300 in order to validate the results obtained with A-485 using another technique. It is however to note that, on the contrary to A-485, shRNAs will degrade p300 mRNA thus reducing p300 protein levels which will probably have additional effects on CIC-DUX4 tumorigenic program compared to the inhibition of its catalytic activity. Indeed, p300 protein can potentially stabilize protein complexes containing CIC-DUX4 thus, in this case, p300 catalytic activity may not be crucial but the presence of its protein would maintain the complex stability. Another distinction can be made since the injection of CDS cells expressing shRNA against p300 in immunocompromised mice will help evaluating the ability of these cells to initiate tumor formation in absence of p300, whereas treatment of tumor bearing mice with A-485 is more comparable to a curative approach when the tumor is already established. Additionally, if our hypothesis is confirmed, the comparison between A-485-treated CDS cells and CDS cells expressing either p300 or DUX4 shRNAs should show substantial similarities in terms of phenotype and gene expression changes.

Our *in vitro* experiments demonstrated the reduction of a panel of CIC-DUX4 target genes after A-485 treatment of both CDS2 and CDS3 cell lines, indicating a specific effect of p300 inhibition on CIC-DUX4 downstream gene expression regulation. The reduction of CIC-DUX4 protein expression itself in a dose dependent manner can be partially explained by the effect of p300 inhibition on CIC-DUX4 transcription itself, but the complete degradation of CIC-DUX4 protein at higher A-485 doses reveals a putative post-translational effect of p300 inhibition on CIC-DUX4 protein stability. Indeed, after 5 days of treatment with 1 μ M A-485, CDS2 cells showed a 80% CIC-DUX4 mRNA expression

level compared to the DMSO control condition, while the CIC-DUX4 protein level was strongly reduced in the same condition and timepoint. Interestingly, despite the strong reduction of CIC-DUX4 protein expression after high doses of A-485 treatment, CIC-DUX4 target genes transcription is reduced but still partially maintained. Even if further timepoints of A-485 treatment are needed to assess CIC-DUX4 target genes levels after longer p300 inhibition in CDS cells, it would also be important to understand if endogenous CIC is binding back to its targets after CIC-DUX4 degradation. Indeed, mutation of CIC has shown to be sufficient to induce a de-repression of its target genes allowing for their transcription. In the situation where CIC-DUX4 is degraded after p300 inhibition, endogenous CIC has to actively bind to its target genes to strongly repress their expression. If CIC is not recruited to its target genes after CIC-DUX4 depletion, those targets may show only a partial decrease in their transcriptional level. However in absence of CIC recruitment their expression may not be completely repressed.

Interestingly, since neither the endogenous CIC nor the DUX4 proteins were affected by p300 inhibition, this approach may hold strong promise to selectively target CIC-DUX4 protein stability without perturbing the normal function of the endogenous wild type proteins. Moreover, p300 treatment has shown to reduce western blot signal intensity of both the 250kDa and higher molecular weight bands in a dose dependent manner. Thus, if two CIC-DUX4 isoforms exist and are both important for CDS sarcomagenesis, p300 inhibition will lead to their degradation, perturbing their specific function. Also, since p300 inhibition is directly targeting the CIC-DUX4 protein, CDS bearing a different CIC-DUX4 translocation would also probably benefit from this approach. To test this, we had the chance to receive a fourth CDS cell line called “Kitra cells” from Dr. S. Nakai from the department of orthopaedic surgery of Osaka university, Japan. In Kitra cells, CIC-DUX4 comes from the t(12;19) chromosomal translocation, which involves the DUX4 gene copy located on chromosome 12. By using the same approach as for the other CDS cell lines we will assess the sensitivity to p300 inhibition of this model. Finally, the addition of a MAPK inhibitor to p300 inhibition would probably further help maintaining a stable level of endogenous CIC, and potentially

enhance CIC to return to its genomic target loci, reinforcing the repression of MAPK pro-tumorigenic target genes. This hypothesis warrants further investigation.

Since CIC mutations have been involved in the pathogenesis of multiple tumors, it would be of great interest to apply the knowledge acquired from the study of CIC-DUX4 sarcoma to develop new therapeutic approaches applicable to additional CIC rearranged tumors. Unfortunately, since the effect of inhibiting p300 activity most probably stems from the presence of the C-terminal portion of DUX4 in CIC-DUX4, this approach would probably be ineffective against tumors harboring CIC mutations, or bearing different CIC translocations. Since CIC and CIC-DUX4 contain the same DNA binding domain sequences we can reasonably assume that a substantial number of target genes will be shared between these two proteins. Additionally, if we imagine that CIC-DUX4 couples the effect of derepressing the CIC target genes due to CIC loss of function, with the potent transactivation activity of DUX4, we can imagine CIC-DUX4 functioning as a mutant CIC version displaying extremely strong transactivation capacity. Thus, our current assessment of direct downstream target genes by combining ChIP-seq and RNA-seq techniques on relevant models has the potential to identify a list of perturbed pathways shared by other CIC mutated tumors. The perturbation of these pathways by genetic depletion or pharmacological inhibition studies will provide additional options to impair CIC-DUX4 and CIC-rearranged tumors development.

In conclusion, CIC-DUX4 displays unique biological features in addition to its transcription factor function. Its cytoplasmic biological activity remains to be elucidated and the potential presence of a longer CIC-DUX4 isoform certainly adds another layer of complexity to CDS biology. Despite the technical difficulties we faced during this study, the integrated genome wide epigenetic profiling of primary CDS tumors, as well as experimental models of this malignancy are allowing us to acquire a substantial amount of data that will help improving our knowledge in the field of CIC-DUX4 sarcomas. Other CIC rearranged or mutated diseases could also benefit from this approach, since CIC alterations have been shown to play major roles in several developmental and oncologic settings.

Finally, the discovery of CDS sensitivity to p300 inhibition *in vitro* and *in vivo*, and its strong selective effects on CIC-DUX4 protein stability may represent a promising opportunity for CDS treatment.

Materials and methods

Cell culture

CDS1 cell line was established at the Japanese Foundation for Cancer Research of Tokyo, Japan as described in Dr. M. Kawamura-Saito study from 2006 (30). CDS2 and CDS3 cell lines were established at National Cancer Center Hospital, Tokyo, Japan as described in Dr. R. Oyama study from 2017 (87) and Dr. M. Yoshimatsu study in 2019 (85) respectively.

All CDS cell lines were grown in RPMI medium (Gibco) supplemented with 10% fetal bovine serum (FBS, Gibco) and 1% Penicillin-Streptomycin (PenStrep, Gibco) in adherent cell culture conditions. Then, cells were passed in RPMI medium (Gibco) supplemented with 20% KO serum (Gibco) to perform further experiments. Primary hpMSCs were harvested after surgery from human bone tissue fragments with the approval of the Ethics Committee of the Canton de Vaud. Cells were grown in IMDM (Gibco) supplemented with recombinant human PDGF (Prospec) to a final concentration of 10ng/ml, 10% FBS (Gibco), 1% NEAA (Gibco) and 1% PenStrep (Gibco) in adherent cell culture conditions. HEK293T and Hela cell lines were grown in DMEM (Gibco) supplemented with 10% FBS (Gibco) and 1% PenStrep (Gibco) in adherent cell culture conditions. A673 cells were grown in RPMI medium (Gibco) supplemented with 10% FBS and 1% PenStrep (Gibco).

Primary tumor samples

Four frozen CIC-DUX4 tumor samples were obtained from the Memorial Sloan-Kettering Cancer Center, New York where RNA sequencing was performed for each sample. Frozen samples were kept at -80°C until used for further experiments. Paraffin imbedded tumor tissue from another CDS patient was provided by G. Petur Nielsen from Massachusetts General Hospital pathology department, Boston, USA.

Lentiviral infections

HEK 293T packaging cells were transfected for lentivirus production. After 72h, the supernatant was collected, filtered (0.45µm pores, Milipore) and 100x concentrated using Lenti-X concentrator (Takara) by following manufacturer's recommendations. 100x concentrated lentivirus 200ul aliquots

were snap frozen in liquid nitrogen for storage. CIC-DUX4 induction was obtained using a pIND20 lentiviral plasmid containing the coding sequence for 5' *FLAG-CIC-DUX4* 3' under the control of a tetracycline responsive promoter (Tet-on). Empty pIND20 plasmid was used as control. hpMSCs were plated in standard medium with 10% Tetracycline-free FBS (Gibco) in order to have 1×10^6 hpMSC cells per dish the day of infection. Infection was done by adding 66.7ul of 100x virus aliquot previously warmed up to room temperature with Polybrene (hexadimethrine bromide, Sigma) at a final concentration of $6 \mu\text{gr/ml}$ for 8h before changing cells' medium. Antibiotic selection was initiated 2 days after infection with G418 at a final concentration of $250 \mu\text{gr/ml}$ during 96h. The day of selection removal, hpMSCs were treated with Doxycycline (Doxycycline hyclate, Sigma) at a final concentration of 100ng/ml during 72h then cells were harvested and used for further experiments. HEK293Ts cells were plated in standard medium with 10% Tetracycline-free FBS(Gibco) in order to have 0.5×10^6 cells per dish the day of infection. Infection was done by adding 100ul of 100x virus aliquots previously warmed up to room temperature with Polybrene (hexadimethrine bromide, Sigma) at a final concentration of $6 \mu\text{gr/ml}$ for 8h before changing cells' medium. 24h after, cells were treated with Doxycycline (Doxycycline hyclate, Sigma) at a final concentration of $1 \mu\text{g/ml}$ during 72h then cells were harvested and used for further experiments.

RNA extraction, cDNA synthesis and quantitative real-time PCR

RNA was extracted using the RNeasy RNA isolation kit (Qiagen). For cDNA synthesis, 500ng of RNA were reverse transcribed using the high capacity cDNA reverse transcription kit (Thermofisher scientific). Quantitative real-time PCR was run by QuantStudio 5 Realtime PCR machine (Thermofisher scientific). Each PCR reaction was done in triplicate and fold change was calculated using the comparative ΔCt method. Relative quantifications were normalized to endogenous control *GAPDH* or *TBP*. Primer sequences used for SYBR Green (Thermofisher scientific) quantification of gene expression are listed below:

Gene	Forward primer	Reverse primer
GAPDH	5' GTC TCC TCT GAC TTC AAC AGC G 3'	5' ACC ACC CTG TTG CTG TAG CCA A 3'
TBP	5' CGG CTG TTT AAC TTC GCT TC	5' CAC ACG CCA AGA AAC AGT GA 3'
CIC-DUX4	5' CTC ACC CAG CTC GGA CTC T 3'	5' CCG GGA TGC CTT GCA TCT G 3'
ETV1	5' GCA AGA ACG CTT CCT GGC 3'	5' CCT TCCCGA TAC ATT CCT 3'
ETV4	5' AGG AAC AGA CGG ACT TCG CCT A 3'	5' CTG GGA ATG GTC GCA GAG GTT 3'
ETV5	5' TCA GCA AGT CCC TTT TAT GGT C 3'	5' GCT CTT CAG AAT CGT GAG CCA 3'
ERG	5' CGT GCC AGC AGA TCC TAC G 3'	5' GGT GAG CCT CTG GAA GTC G 3'
WT1	5' GAG ACA TAC AGG TGT GAA 3'	5' GCC AGC TGG AGT TTG GTC A 3'

Preparation of protein cell lysate and Western Blot

Western blot was performed according to standard procedures. Briefly, cells were harvested, centrifuged, washed with PBS and lysed in SDS-RIPA buffer (150mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, 50mM Tris pH8.0 and protease inhibitor cocktail (Roche)). The lysate was incubated on ice for 10 minutes and centrifuged at 4°C for 15 minutes at 13'000 RPM. The supernatant was used for protein sample preparation. 40µg of proteins were loaded per lane for Western blot and samples underwent electrophoresis through an 8% polyacrylamide gel at 125mV for 1h30. Transfer was performed under 250mA for 2h onto Whatman® Protran® BA83 nitrocellulose membrane. The membrane was then blocked in 5% milk in tris-buffered saline containing 0.5% tween 20 (TBST) for 1 hour, washed with TBST and hybridized with primary antibody according to the manufacturer's recommendations. Primary antibodies included : anti-Dux4 antibody-C-terminal (ab171303, Abcam), DUX4 monoclonal antibody (P4H2, Thermofisher), anti-DUX4 antibody (C-2, Santa Cruz), anti-α-Tubulin Mouse monoclonal antibody (DM1A, Merck). Secondary antibodies included: HRP-conjugated goat anti-mouse antibody (GE healthcare) and goat anti-rabbit antibody (Dako). Bands were visualized using Western Bright Sirius (Wiltec AG) detection reagents, according to the manufacturer's instructions.

Immunofluorescence

Cells were seeded on glass coverslips. The next day, cells were fixed with 4% paraformaldehyde for 12 minutes at room temperature followed by permeabilization with a 0.3% Triton X-100, 2% gelatin and 50mM NH₄Cl solution. Cells were then stained with anti-DUX4 monoclonal antibody (P4H2,

Thermofisher), (1/300), at room temperature for 1 hour or with p300 antibody (Cell Signaling), (1/500) at room temperature for 1 hour. Cells were then washed with cold PBS and incubated with Alexa 594-labeled anti-mouse or anti-rabbit (1/1200) secondary antibody for 1 hour at room temperature. Cells were washed again with cold PBS and DAPI staining was performed during 5min. Coverslips were mounted on slides for confocal microscopy examination (Zeiss LSM 710). Pictures were obtained using Fiji software.

Proximity ligation assay (PLA)

Cell seeding, fixation and permeabilization was performed the same way as for immunofluorescence experiments. Specific primary antibodies against DUX4 (P4H2, Thermofisher) and p300 (Cell signaling) were used for the PLA. Duolink II fluorescence kit was purchased from Sigma Aldrich. This kit includes blocking solution, wash buffers, amplification and ligase solution and detection reagents. PLA was performed following manufacturer's recommendations.

Chromatin Immunoprecipitation

ChIP assays were carried out on CDS1, CDS2 cells and frozen primary CDS tissue samples. Approximately $3-10 \times 10^6$ cells or 5-10mg tumor tissue were used per sample per immunoprecipitation as previously described (92). Briefly, formaldehyde fixed cells were lysed and sonicated in order to get chromatin fragments of a size range of 200-600 bases. Solubilized chromatin was immunoprecipitated with antibodies against H3K4me3 (Milipore), H3K27me3 (Cell signaling), H3K27ac (Active Motif), H3K4me1 (Abcam), DUX4 (Abcam), p300 (Cell signaling), FLAG M2 (Sigma) or CIC (Abcam) antibodies. Antibody-chromatin complexes were pulled down with Protein G-Dynabeads (Life Technologies), washed and then eluted. After crosslink reversal, RNase A and Proteinase K treatment, immunoprecipitated DNA was extracted using AMPure XP beads (Beckman Coulter). ChIP DNA was quantified with Qubit (Invitrogen). ChIP DNA samples were used to prepare sequencing libraries using TruSeq ChIP Preparation kit (Illumina). ChIP DNA and input controls were sequenced with the Hi-Seq Illumina Genome Analyzer.

In vitro p300 inhibition

For pharmacological targeting of p300, the A-485 chemical inhibitor (93) was dissolved in DMSO. 5000 CDS1 cells, 5000 CDS2 cells and 15'000 CDS3 cells were plated each well as quadruplicates in a 96-wells' plate and allowed to adhere for 24h in KO serum complemented medium. The A-485 drug was added at a final concentration of 0.01, 0.05, 0.1, 0.25, 0.5, 1 and 2 μm for CDS1 and CDS2 treatment, 5 and 10 μm conditions were added for CDS3 treatment. The according volumes of DMSO were added to control wells and cells were treated with 2 μm in 4 supplementary wells with Staurosporine. The fraction of viable cells was determined using the CellTiter-Glo Luminescent Cell Viability Assay (Promega) according to the manufacturer's recommendations. To generate dose response curves, data were normalized by setting the average value of Staurosporine and DMSO control wells to 0% and 100% viability respectively. The replicate values for each dilution point were averaged and the EC₅₀ values for each compound were generated in Prism GraphPad by fitting normalized data to a sigmoidal curve model of linear regression. D5 timepoint EC₅₀ values of each cell line were reported in the published cell screening study from which the same protocol was followed.

In vivo tumorigenic assay

For *in vivo* tumorigenicity experiments, 1x10⁶ and 5x10⁶ CDS1 and 1x10⁶, 2x10⁶ and 5x10⁶ CDS2 cells were bilaterally injected subcutaneously into one NOD-SCID gamma KO mouse for each condition. Mice were monitored 3 times a week for tumor development and sacrificed before tumor volume was reaching 1 cm³. Tumors were harvested, measured, weighed and fixed in PFA 4%. Tumors were included in paraffine and H&E staining was performed.

DUX4 immunohistochemistry

DUX4 IHC was performed by the EPFL Histology Core Facility using a Ventana automatic tissue processing machine. DUX4 monoclonal antibody (P4H2, Thermofisher) was used to perform the staining.

In vivo A-485 treatment

2×10^6 CDS2 cells were injected bilaterally in the subcutaneous suprascapular region of 10 NOD-SCID gamma KO mice. When small tumors started to be visible (i. e. 3 weeks after cells injection), mice were randomized and A-485 (MedChemExpress) was administrated intraperitoneally at a dose of 100mg/kg twice daily. Mouse weight was followed every three days. In vivo experiment was terminated when mice showed signs of suffering, equal or higher weight loss than 20% or when the tumor reached 1cm^3 .

Appendix 1 : LIN28B underlies the pathogenesis of a subclass of Ewing sarcoma

EWS-FLI1 is a fusion protein resulting, in the vast majority of the cases, from the chromosomal translocation t(11;22), and responsible for Ewing sarcoma pathogenesis. In the last decades, strong effort has been deployed in order to directly target EWS-FLI1 protein function, but given the highly unstable structure of this translocation, this approach didn't meet our expectations. Several promising alternative approaches were then proposed in order to impair the function of this aberrant fusion protein, including the modulation of its direct binding partners' activity and the perturbation of the downstream molecular pathways induced by EWS-FLI1.

In this study, to define key factors that could potentially impair Ewing sarcoma tumor formation and maintenance, we integrated survival data on Ewing sarcoma patient with CRISPR screening gene dependencies and identified LIN28B as a potential candidate. This oncofetal RNA binding protein was found to be expressed in approximately 10% of Ewing sarcoma for which gene expression was publicly available. By using primary derived Ewing sarcoma cell cultured as spheroid in low adherence condition, we demonstrated that Ewing sarcoma cells expressing LIN28B display a higher proliferation rate *in vitro* and tumorigenic potential *in vivo*. Knockdown of LIN28B in primary Ewing sarcoma-derived cultures drastically reduced their spherogenic ability and abrogated *in vivo* tumor formation. On the contrary, induction of exogenous LIN28B expression in primary culture models lacking LIN28B expression enhanced their clonogenic capacity, as well as their tumor formation ability.

The bioinformatic analysis of the gene and microRNA expression profiles of our experimental models revealed that LIN28B-positive Ewing sarcoma cells phenotype may result from the combined properties of LIN28B on let-7 microRNA family maturation and direct target transcripts stabilization. By integrating those data with results obtained from EWS-FLI1 knockdown and overexpression studies, we discovered that LIN28B may directly impact EWS-FLI1 transcript stability. In keeping with this, we observed a strong decrease in EWS-FLI1 mRNA and protein levels after LIN28B knockdown, and showed that LIN28B directly binds to EWS-FLI1 transcripts. The assessment of

nascent RNA levels after LIN28B knockdown revealed that LIN28B is increasing EWS-FLI1 transcript stability. In addition, the study of LIN28B's effect on EWS-FLI in primary human pediatric mesenchymal stem cells showed that LIN28B mediated stabilization of EWS-FLI1 can occur in a permissive non-transformed cellular environment. Moreover, by introducing EWS-FLI1 cDNA that lacks the 3' UTR region in hpMSCs, we showed that LIN28B mediated EWS-FLI1 stabilization is independent of the let-7 microRNA family.

Finally, we took advantage of a LIN28B small molecule inhibitor to impair LIN28B regulated pathways in our models. Treatment of LIN28B⁺ Ewing sarcoma primary spheres with this inhibitor led to a decrease in EWS-FLI1 expression level and a reduction in their clonogenic ability. Finally, the treatment of Ewing sarcoma spheres with the LIN28B inhibitor prior to their injection into immunocompromised mice, resulted in a robust reduction of tumor size in all treated mice compared to untreated controls, confirming the potential therapeutic interest of this approach.

Contribution statement

During my 4 years of MD-PhD thesis, I spent 1 year full time on this project. I worked exclusively on it during the first 6 months and then worked simultaneously on both CIC-DUX4 sarcoma and Ewing sarcoma projects.

I mainly contributed to the following sections of the Ewing sarcoma project :

- *In vitro* characterization of primary Ewing sarcoma-derived cells.
- Generation and characterization of LIN28B knockdown models.
- Generation and characterization of LIN28B overexpression and rescue models.

I contributed to the published paper by participating to the following sections :

Results : Primary LIN28B⁺ and LIN28B⁻ EwS cells have different growth kinetics in vitro and in vivo.

- LIN28B⁺ and LIN28B⁻ primary derived cells culture maintenance.
- Figure S2 D and E generation.

Results : LIN28B is required for LIN28B⁺ EwS self-renewal and tumor initiation

- LIN28B knockdown models establishment and maintenance.
- Conception, planification and performance of the shRNA and CRISPR knockdown experiments.
- Figure 3 B generation.

Results : Exogenous LIN28B Expression increases clonogenicity of and tumor initiation by LIN28B⁺ EwS spheres

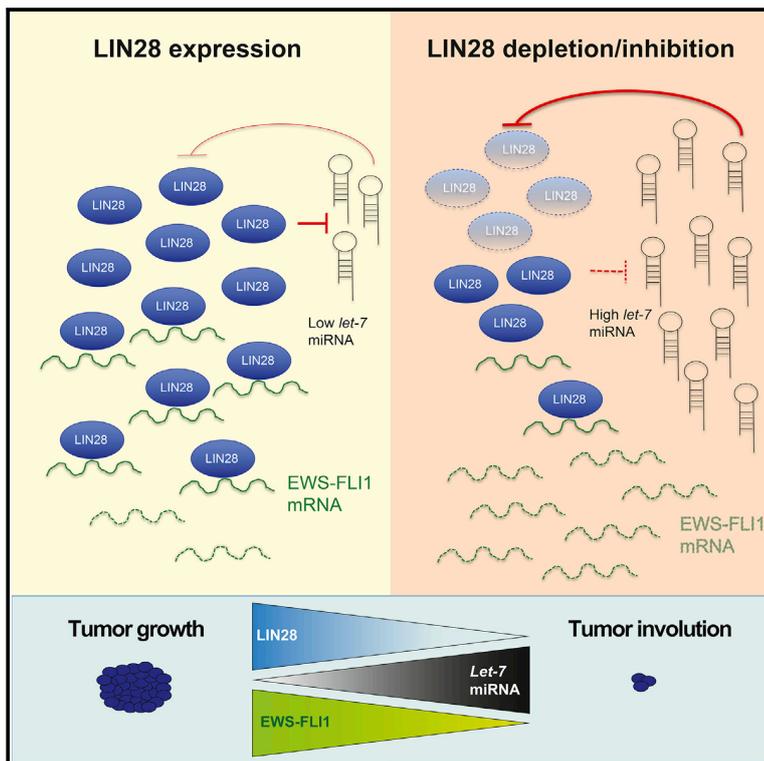
- LIN28B overexpression model and rescue experiment establishment and maintenance.
- Conception, planification and performance of the LIN28B overexpression experiments.
- Figure 4 A generation.

I participated in the generation of the manuscript by providing critical feedback and contributed to the finalization of the manuscript during the reviewing process.

Published paper

LIN28B Underlies the Pathogenesis of a Subclass of Ewing Sarcoma

Graphical Abstract



Authors

Tugba Keskin, Arnaud Bakaric, Patricia Waszyk, ..., Miguel N. Rivera, Nicolò Riggi, Ivan Stamenkovic

Correspondence

ivan.stamenkovic@chuv.ch

In Brief

Keskin et al. show that Lin28B directly binds EWS-FLI-1 and maintains its stability in ~10% of Ewing sarcomas. Lin28B depletion or its chemical inhibition in these tumors results in the loss of EWS-FLI-1 expression and tumor deconstruction. Lin28B may provide an attractive therapeutic target in a subset of Ewing sarcomas.

Highlights

- Lin28B expression identifies a subset of Ewing sarcomas
- Lin28B binds and stabilizes EWS-FLI-1 transcripts in these tumors
- Depletion or inhibition of Lin28B results in the deconstruction of the tumors
- Lin28B may provide a therapeutic target in a subset of Ewing sarcomas



LIN28B Underlies the Pathogenesis of a Subclass of Ewing Sarcoma

Tugba Keskin,^{1,14} Arnaud Bakaric,^{1,14} Patricia Waszyk,^{1,14} Gaylor Boulay,^{2,13,14} Matteo Torsello,^{1,7} Sandrine Cornaz-Buros,^{1,3} Nadja Chevalier,^{1,10} Thibaud Geiser,¹ Patricia Martin,¹ Angela Volorio,^{1,2} Sowmya Iyer,² Anupriya Kulkarni,² Igor Letovanec,⁴ Stéphane Cherix,⁵ Gregory M. Cote,⁶ Edwin Choy,⁶ Antonia Digkila,⁷ Michael Montemurro,⁷ Ivan Chebib,⁸ Petur G. Nielsen,⁸ Angel M. Carcaboso,⁹ Jaime Mora,⁹ Raffaele Renella,¹⁰ Mario L. Suvà,² Carlo Fusco,¹ Paolo Provero,^{11,12} Miguel N. Rivera,^{2,13,15} Nicolò Riggi,^{1,15} and Ivan Stamenkovic^{1,15,16,*}

¹Experimental Pathology Service, Centre Hospitalier Universitaire Vaudois, University of Lausanne, 1011 Lausanne, Switzerland

²Department of Pathology and Center for Cancer Research, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA

³Department of Pediatrics, Centre Hospitalier Universitaire Vaudois, University of Lausanne, 1011 Lausanne, Switzerland

⁴Department of Pathology, Centre Hospitalier Universitaire Vaudois, University of Lausanne, 1011 Lausanne, Switzerland

⁵Department of Orthopaedics and Traumatology, Centre Hospitalier Universitaire Vaudois, University of Lausanne, 1011 Lausanne, Switzerland

⁶Center for Sarcoma and Connective Tissue Oncology, Massachusetts General Hospital, Boston, MA 02114, USA

⁷Department of Oncology, Centre Hospitalier Universitaire Vaudois, University of Lausanne, 1011 Lausanne, Switzerland

⁸Department of Pathology, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA

⁹Department of Pediatric Hematology and Oncology, Hospital Sant Joan de Déu, 08950 Barcelona, Spain

¹⁰Department Woman-Mother-Child, Division of Pediatrics, Centre Hospitalier Universitaire Vaudois, University of Lausanne, 1011 Lausanne, Switzerland

¹¹Center for Translational Genomics and Bioinformatics, San Raffaele Scientific Institute, 20132 Milan, Italy

¹²Department of Molecular Biotechnology and Health Sciences, University of Turin, 10124 Turin, Italy

¹³Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA

¹⁴These authors contributed equally

¹⁵Senior author

¹⁶Lead Contact

*Correspondence: ivan.stamenkovic@chuv.ch

<https://doi.org/10.1016/j.celrep.2019.12.053>

SUMMARY

Ewing sarcoma (EwS) is associated with poor prognosis despite current multimodal therapy. Targeting of EWS-FLI1, the fusion protein responsible for its pathogenesis, and its principal downstream targets has not yet produced satisfactory therapeutic options, fueling the search for alternative approaches. Here, we show that the oncofetal RNA-binding protein LIN28B regulates the stability of *EWS-FLI1* mRNA in ~10% of EwSs. LIN28B depletion in these tumors leads to a decrease in the expression of EWS-FLI1 and its direct transcriptional network, abrogating EwS cell self-renewal and tumorigenicity. Moreover, pharmacological inhibition of LIN28B mimics the effect of LIN28B depletion, suggesting that LIN28B sustains the emergence of a subset of EwS in which it also serves as an effective therapeutic target.

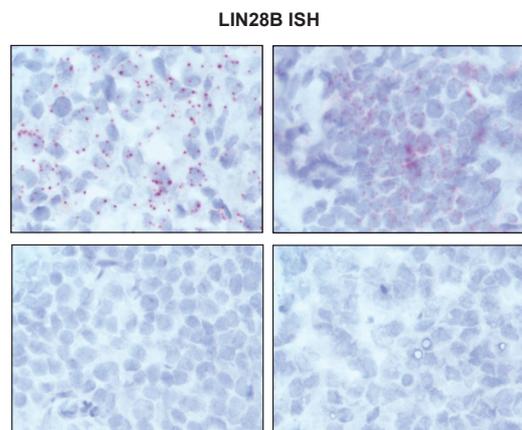
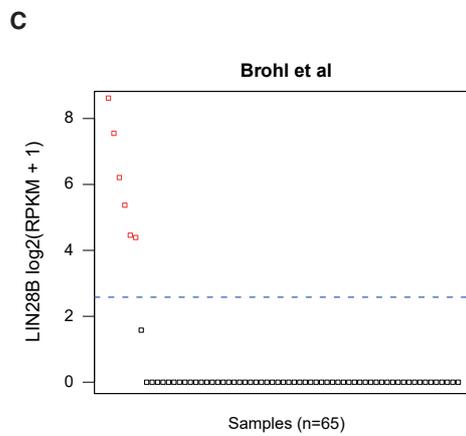
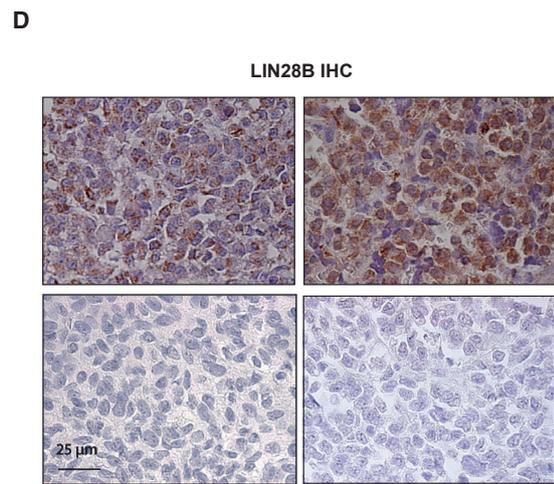
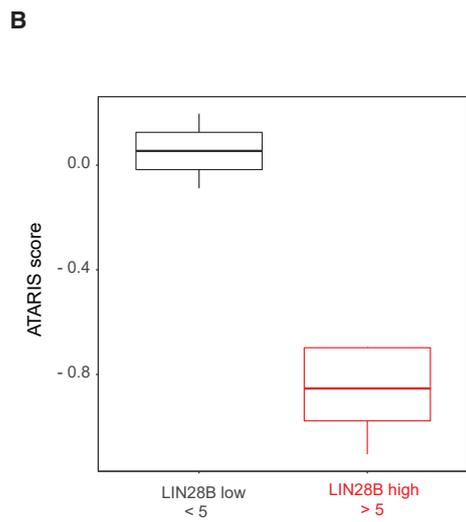
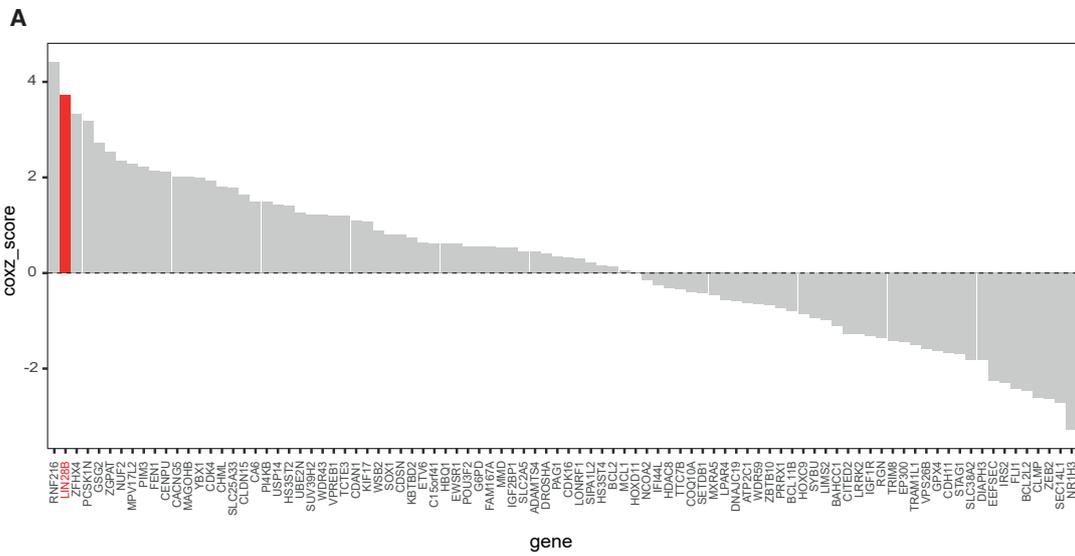
INTRODUCTION

Ewing sarcoma (EwS) is the second most common bone malignancy in children and young adults. Its distinguishing biological

feature is a unique set of reciprocal chromosomal translocations that generate fusions between *EWS* and one of several genes encoding ETS family transcription factors, the most common being *FLI1* (Riggi and Stamenkovic, 2007). The EWS-FLI1 fusion protein underlies EwS pathogenesis and behaves as an aberrant transcription factor that can induce both transcriptional activation and repression. Upon binding to GGAA repeat elements, EWS-FLI1 induces chromatin relaxation and the recruitment of chromatin-remodeling enzymes to activate *de novo* enhancers at genomic regions that are normally devoid of any regulatory function in other cell types (Boulay et al., 2017; Riggi et al., 2014; Tomazou et al., 2015). In contrast, by displacing more active wild-type ETS transcription factors from enhancers containing non-repetitive canonical ETS motifs, EWS-FLI1 causes target gene repression (Riggi et al., 2014). EWS-FLI1 therefore behaves as an oncogenic pioneer factor that can reprogram the regulatory and transcriptional features of permissive primary cells, leading to their malignant transformation.

EwS has a strong tendency toward relapse and dissemination. Recurring and metastatic disease respond poorly even to the most aggressive forms of multimodal therapy that are available and are associated with a high mortality rate (Pishas and Lessnick, 2016). Although the search for effective therapeutic strategies continues to be intense, with some potentially promising leads (Cornaz-Buros et al., 2014; De Vito et al., 2012; Engert et al., 2015; Erkizan et al., 2009; Vormoor and Curtin, 2014),





(legend on next page)

few options are available when the standard first-line combination of surgery and cytotoxic drugs fails. The unstructured features of EWS-FLI1 hinder its direct targeting (Dunker and Uversky, 2010), and therapeutically accessible regulators and downstream effectors of EWS-FLI1 that are vital to EwS growth have thus far been elusive. One relatively unexplored avenue is the search for factors that may complement or synergize with the oncogenic function of EWS-FLI1, and whose targeting may indirectly blunt EwS aggressiveness and improve patient survival. To identify such putative factors, we addressed the effects on patient survival of the top 100 genes observed to affect EwS cell growth in a recently published whole-genome CRISPR screen (Aguirre et al., 2016). Among the top candidates identified by this combinatorial approach, we found the gene that encodes the oncofetal RNA-binding protein (RBP) LIN28B.

Expression of the LIN28 paralogs (LIN28A and LIN28B) occurs during normal embryogenesis, at which time they control the balance between pluripotency and differentiation and regulate the transition from early to late embryonic development (Shinoda et al., 2013; Zhang et al., 2016). By inhibiting let-7 microRNA (miRNA) family biogenesis, LIN28A and LIN28B protect the expression of let-7 target transcripts, which include numerous pluripotency-related genes and oncogenes, thereby promoting normal stem cell maintenance and tumor growth (Madison et al., 2013, 2015; Mayr and Heinemann, 2013; Thornton and Gregory, 2012; Viswanathan and Daley, 2010). However, LIN28A and LIN28B also exert let-7-independent regulatory functions by binding to and influencing the translation of a wide repertoire of mRNAs (Balzeau et al., 2017; Viswanathan and Daley, 2010). LIN28B and, to a lesser extent, LIN28A are aberrantly expressed in a broad range of adult human malignancies associated with poor prognosis (Chatterji et al., 2018; Chatterji and Rustgi, 2018), in which they are often confined to poorly differentiated cell subpopulations that may express cancer stem cell (CSC) features (Balzeau et al., 2017; Carmel-Gross et al., 2016; Viswanathan and Daley, 2010). LIN28B is implicated in the pathogenesis of several primitive pediatric malignancies, including Wilms tumor (Urbach et al., 2014), neuroblastoma (Hennchen et al., 2015; Powers et al., 2016), and primitive neuro-ectodermal brain tumors (Choi et al., 2016; Picard et al., 2012).

Based on these observations, we addressed the potential implication of LIN28B expression in the pathogenesis of EwS. Using patient-derived EwS spheres from primary LIN28B⁺ and LIN28B⁻ tumors, we show that LIN28B-expressing cells not only generate tumor growth more rapidly than their LIN28B⁻ counterparts but are also dependent on LIN28B expression to sustain their self-renewal and tumor-initiating

properties. We demonstrate that LIN28B directly binds EWS-FLI1 transcripts, increasing their stability and ensuring the maintenance of EWS-FLI1 expression in tumor cells. Consequently, LIN28B depletion or pharmacological inhibition in primary EwS cells decreases the expression of both the fusion protein and its direct target genes, leading to the deconstruction of the oncogenic program and the progressive loss of tumorigenic potential *in vitro* and *in vivo*. Our observations identify an EwS subset whose dependence on LIN28B provides a unique opportunity to apply a pharmacological approach toward disrupting its otherwise undruggable oncogenic driving force.

RESULTS

LIN28B Expression Identifies a Subset of EwS

To identify candidate factors whose function may influence EwS evolution and prognosis, we correlated survival in a cohort of 44 EwS patients (Savola et al., 2011) with the expression of the top 100 genes observed to affect EwS cell growth in a recent whole-genome CRISPR library screen of 33 cancer cell lines (Aguirre et al., 2016) (Table S1). In terms of prognostic relevance, LIN28B ranked second only to RNF216, which encodes a RING finger protein involved in the regulation of the nuclear factor κ B (NF- κ B) pathway (Chuang and Ulevitch, 2004; Nakhaei et al., 2009), known to promote EwS cell survival (Javelaud et al., 2000) (Figure 1A). Whereas LIN28B underlies the emergence and evolution of several primitive pediatric malignancies, its role in EwS is unknown, prompting us to interrogate its putative implication in EwS pathogenesis. Patients with tumors expressing LIN28B (n = 3) had the poorest prognosis within the cohort, with a median survival of <2 years, whereas the evolution of patients with tumors lacking LIN28B was less severe and included multi-year survivors (n = 41; Figure S1A). Interrogation of the Cancer Cell Line Encyclopedia database (CCLE, Broad Institute; <https://portals.broadinstitute.org/ccle>) (Barrettina et al., 2012) revealed that 3 of the 9 EwS cell lines used in the CRISPR-Cas9 screen (CADOES1, TC32, and MHES1; data not shown) did not express LIN28B and were consequently unaffected by its targeting by gene editing (Figure 1B). These observations suggest that LIN28B plays a functional role only in a subset of EwS.

We next assessed the frequency of LIN28B⁺ EwS in other datasets. Analysis of publicly available gene expression datasets of primary EwS revealed that in a cohort of 65 patients, 6 bore LIN28B⁺ tumors (Figure 1C) (Brohl et al., 2014). Furthermore, 3 primary EwS were LIN28B⁺ in a microarray dataset of 44 patients (Figure S1A), and assessment of LIN28B expression in our cohort of 40 primary EwSs using immunohistochemistry (IHC)

Figure 1. LIN28B Expression Identifies a Subclass of Aggressive EwSs

(A) The waterfall plot depicts Cox survival Z scores for 92 of the 100 genes that selectively affect EwS cell line growth in the Achilles database (Aguirre et al., 2016). Expression of 8 genes was not recorded in the survival dataset (Savola et al., 2011).

(B) Analytic technique for assessment of RNAi by similarity (ATARiS) score distributions for cell lines that have low (<5) and high (\geq 5) expression levels for LIN28B.

(C) Scatterplot depicting LIN28B gene expression (red squares, positive; black squares, negative) across a cohort of 65 primary EwSs (Brohl et al., 2014).

(D) Immunohistochemical (IHC, upper panels) and RNA *in situ* hybridization (ISH, lower panels) assessment of LIN28B expression in primary EwS. The selected images are representative of the 40 primary EwS samples tested (the same tumor samples are shown in the IHC and ISH panels). Expression of LIN28B was observed in 4 of the 40 tumors. The top and bottom images in each panel are LIN28B⁺ and LIN28B⁻, respectively.

See also Figure S1 and Tables S1 and S2.

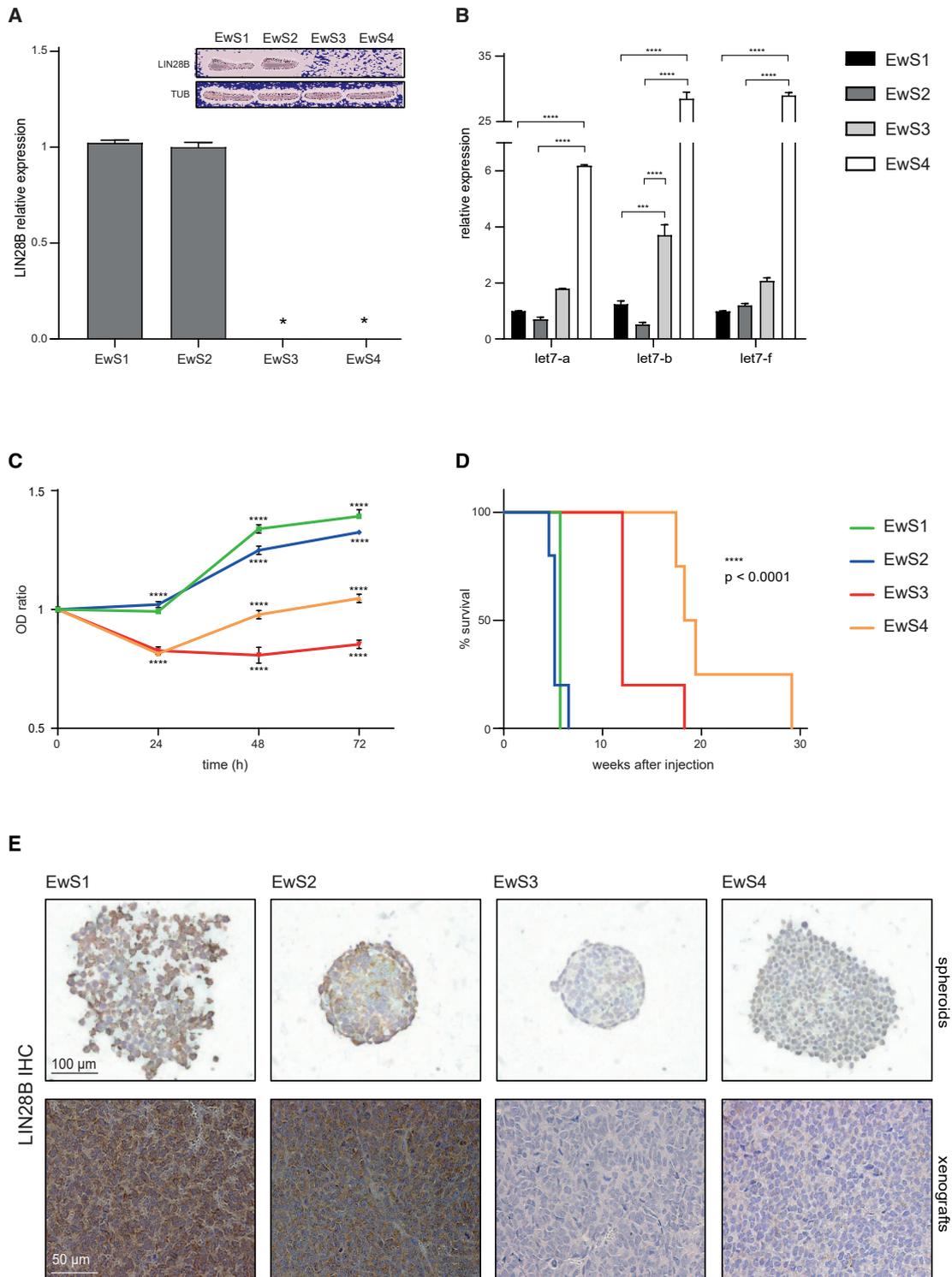


Figure 2. Primary LIN28B⁺ and LIN28B⁻ EwS Cells Display Different Growth Kinetics *In Vitro* and *In Vivo*

(A) qRT-PCR and western blot analyses of LIN28B expression in 4 primary EwS cell cultures (EwS1–EwS4) analyzed. Asterisks represent values that were set to 0 based on late amplification (means \pm SEMs, n = 3 technical replicates).

(B) LIN28B⁺ tumors (EwS1 and EwS2) display reduced expression levels of let-7 family members compared to their LIN28B⁻ counterparts, as assessed by qRT-PCR analysis. Mean values \pm SEMs of 3 technical replicates are shown. The statistical analysis was performed by 2-way ANOVA.

(legend continued on next page)

and *in situ* hybridization (ISH) identified 4 LIN28B⁺ tumors (Figure 1D; data not shown). Examination of these 4 tumors revealed that LIN28B expression was uniform and not confined to cell subpopulations, as is often the case in adult malignancies. Furthermore, there was no gradation of expression among the tumors, which scored either strongly positive or negative. The sum of primary LIN28B⁺ tumors in these 3 datasets amounted to 13 from a total of 149 (8.7%).

The recent association between *STAG2* and *TP53* mutations and poor clinical outcome in EwS patients (Brohl et al., 2014; Crompton et al., 2014; Tirode et al., 2014) prompted us to exclude a possible correlation between these mutations and *LIN28B* expression. Analysis of publicly available whole-exome and transcriptome data from primary EwS (Brohl et al., 2014) failed to reveal a statistically significant correlation between *STAG2/TP53* mutations and *LIN28B* expression (Table S2). LIN28B expression therefore appears to be a marker of a subset of EwS whose biological properties warrant assessment.

Primary LIN28B⁺ and LIN28B⁻ EwS Cells Have Different Growth Kinetics *In Vitro* and *In Vivo*

To address the effect of LIN28B on the behavior of EwS, we generated patient-derived spheroid cultures from 2 LIN28B⁺ and 2 LIN28B⁻ EwS removed at surgery (EwS1–EwS4; Table S3). The 2 primary cultures derived from LIN28B⁺ tumors (EwS1 and EwS2) displayed robust LIN28B mRNA and protein expression, whereas neither LIN28B transcripts nor proteins were detectable in their LIN28B⁻ tumor-derived counterparts (EwS3 and EwS4; Figure 2A). LIN28A was undetectable in any of our samples, excluding the possibility that it may substitute for its paralog in LIN28B⁻ cells (Figure S1C). Consistent with the inhibitory effect of LIN28B on let-7 biogenesis, EwS1 and EwS2 displayed lower levels of mature let-7 than EwS3 and particularly EwS4 (Figures 2B and S1D), which bore a remarkably high expression of let-7 family members.

We then asked whether EwS cell behavior *in vitro* and *in vivo* correlates with LIN28B expression. EwS1 and EwS2 proliferated more rapidly, as assessed by MTS assays *in vitro*, with a roughly 2-fold shorter doubling time than that of EwS3 and EwS4 (Figure 2C). Following injection of 1×10^4 cells from dissociated spheres beneath the kidney capsule of NSG mice, EwS1 and EwS2 cells produced tumors after 6–8 weeks, whereas the median duration required for EwS3 and EwS4 tumor engraftment was 12 and 20 weeks, respectively (Figure 2D). EwS1 and EwS2 sphere-derived tumors maintained LIN28B expression, whereas tumors formed by EwS3 and EwS4 spheres remained LIN28B⁻ (Figure 2E), recapitulating the original primary tumor phenotype. LIN28B expression therefore correlates with rapid EwS cell division and tumor initiation.

LIN28B Is Required for LIN28B⁺ EwS Self-Renewal and Tumor Initiation

To determine whether LIN28B plays a role in LIN28B⁺ EwS cell self-renewal and tumor initiation, we infected EwS1 and EwS2 cells with lentiviral vectors containing either of 2 short hairpin RNAs (shRNAs), shRNA1 and shRNA2, whose sequences target the protein coding region and the 3' UTR of *LIN28B* mRNA, respectively. Expression of either shRNA resulted in comparable LIN28B depletion (Figure 3A, left panel, and Figure 3B). LIN28B depletion virtually abolished sphere formation by EwS1 and EwS2 cells (Figure 3C). Calcein AM staining of LIN28B-depleted cells revealed a <10% decrease in viability (Figure 3D; data not shown), indicating that the observed abrogation of spherogenicity was not due to cell death. Clonogenic assays revealed a dramatic decrease in sphere formation from single EwS1 and EwS2 cells depleted of LIN28B (Figure 3E). Accordingly, 1×10^4 LIN28B-depleted EwS1 and EwS2 cells were unable to form tumors following injection beneath the kidney capsule of NSG mice (Figure 3F). Three mice injected with EwS2 cells expressing LIN28B shRNA2 developed tumors, which were found to express LIN28B at levels superior to those in control tumors (Figure S2C), having most likely originated from cells that had eluded stable LIN28B knockdown. These observations suggest that LIN28B is necessary for the maintenance of self-renewal and tumor initiation by LIN28B⁺ EwS cells.

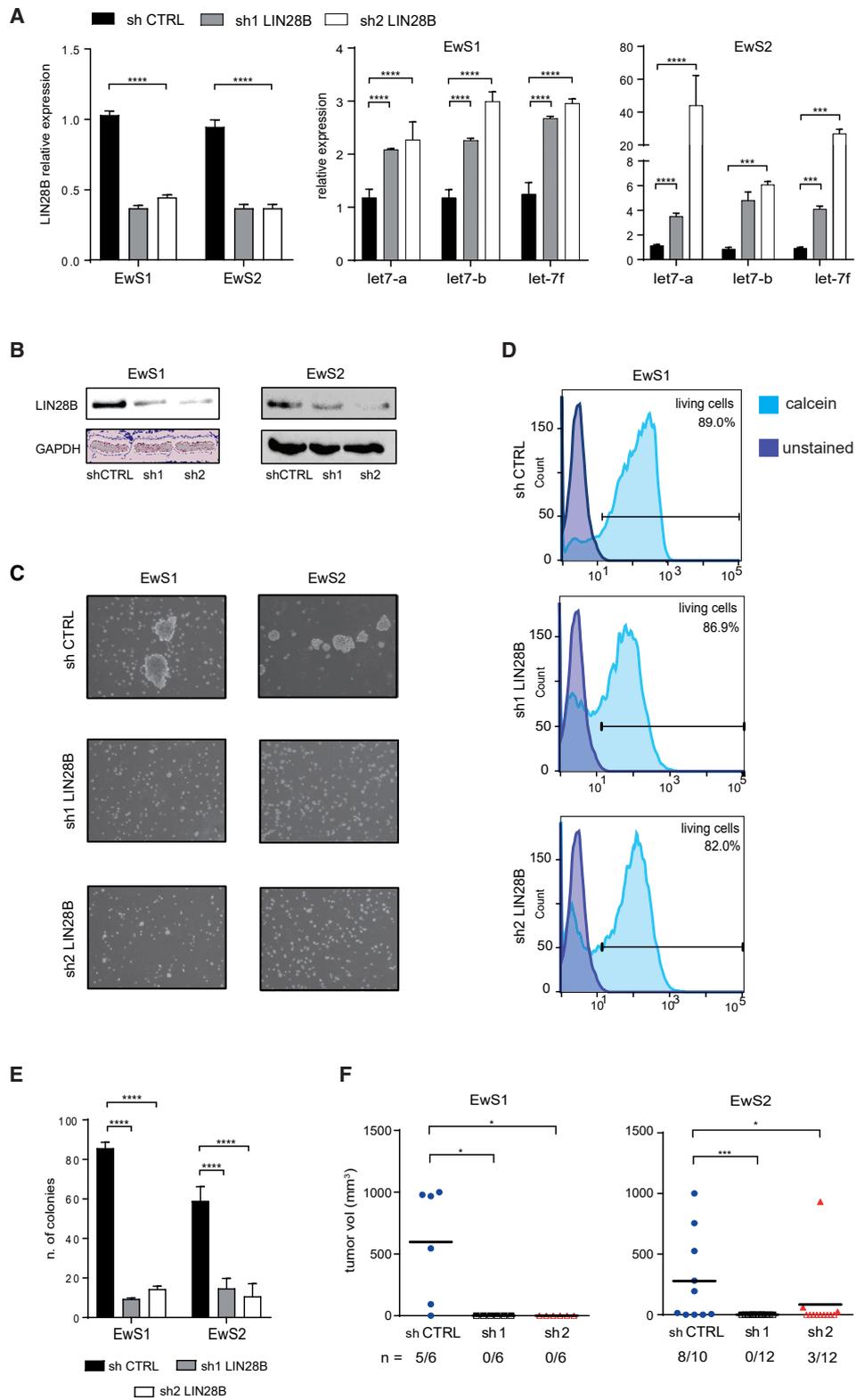
To exclude the possibility that the decreases in clonogenicity and tumor initiation observed using shRNAs were due to off-target effects, we attempted to rescue the LIN28B⁺ EwS phenotype by expressing a *LIN28B* construct that lacks the 3' UTR (*LIN28B*^{Δ3'UTR}) in cells depleted of endogenous LIN28B. Following *LIN28B*^{Δ3'UTR} cDNA introduction into LIN28B⁺ EwS cells, the endogenous transcript was depleted using shRNA2. As expected, shRNA2 affected neither transcript nor protein LIN28B levels in cells expressing exogenous *LIN28B*^{Δ3'UTR} (Figures S2D and S2E). Unlike their counterparts infected with a control vector, cells expressing *LIN28B*^{Δ3'UTR} maintained sphere formation despite the presence of shRNA2, excluding an off-target effect as an explanation for our initial observations (Figure S2F). To provide further support for our findings, we depleted *LIN28B* using CRISPR-Cas9. The expression of a validated single-guide RNA (sgRNA) sequence targeting exon 2 of *LIN28B* (Powers et al., 2016) in EwS1 cells virtually abrogated *LIN28B* mRNA and protein expression (Figures S2G and S2H) and resulted in a robust reduction in clonogenicity (Figure S2I), reproducing the phenotype obtained using shRNAs.

(C) Growth curves of the 4 spherogenic EwS cultures *in vitro* as assessed by MTS assays over 72 h. Four biological replicates were performed, and values were normalized to time point 0 (mean ± SD). Two-way ANOVA was used for statistical analysis.

(D) Survival curves of NSG mice injected with 1×10^4 EwS sphere-derived cells into the left subcapsular kidney compartment. Five mice per group were injected, and the Kaplan-Meier test was used for statistical analysis.

(E) Immunohistochemical assessment of LIN28B expression in the 4 spherogenic EwS cultures (upper panels) and the tumors generated by their injection into NSG mice in (D) (lower panels).

p ≤ 0.001; *p ≤ 0.0001. See also Figure S1 and Table S3.



(legend on next page)

Exogenous LIN28B Expression Increases Clonogenicity of and Tumor Initiation by LIN28B⁻ EwS Spheres

We next asked whether the introduction of *LIN28B* into LIN28B⁻ tumor cells might accelerate their growth and tumor initiation by stably expressing *LIN28B*^{Δ3UTR} in EwS3 cells (Figure 4A, left panels). As expected, LIN28B expression resulted in the suppression of let-7 family maturation and the concomitant induction of the canonical let-7 target gene *HMGA2* (Figure 4A, center and right panels, respectively). *In vitro*, EwS3 cells expressing LIN28B rapidly re-formed spheres following dissociation (Figure 4B) and displayed a 3-fold increase in clonogenicity compared to control cells (Figure 4C). Injection of LIN28B-expressing EwS3 cells under the kidney capsule of NSG mice resulted in detectable tumor growth within 8 weeks in 4 of 6 mice, reminiscent of the behavior of primary LIN28B⁺ cells (Figure 2D). The mean tumor volume at 8 weeks was ~3-fold higher than that of tumors derived from control cells (Figure 4D). Both *LIN28B* and *HMGA2* were elevated in tumors derived from LIN28B-expressing EwS3 cells (Figures 4E and 4F), compared to their empty-vector-infected counterparts. These observations support the results obtained using LIN28B-depleted cells and indicate that the expression of LIN28B alone can augment the self-renewal and tumor-forming capacity of primary EwS cells.

LIN28B Modulation Affects Expression of EWS-FLI1 and Its Target Genes

To further explore the properties of LIN28B⁺ and LIN28B⁻ EwS subtypes and interrogate the mechanism underlying LIN28B⁺ EwS cell dependence on LIN28B expression for self-renewal and tumor initiation, we compared the miRNome and transcriptome of primary LIN28B⁺ and LIN28B⁻ EwS cells using miRNA arrays and RNA sequencing (RNA-seq), respectively. EwS1 and EwS2 cells displayed a similar miRNA expression profile and the same held true for EwS3 and EwS4 cells. Although mature let-7 species were downregulated in EwS1 and EwS2 compared to EwS3 and EwS4, consistent with the inhibition of let-7 maturation by LIN28B, the corresponding miRNA target genes were not significantly overrepresented among the genes that were differentially expressed between LIN28B⁺ and LIN28B⁻ tumors (Figure 5A). Specifically, only 7 predicted let-7 targets were upregulated in LIN28B⁺ tumors (7.757 expected by chance, rendering the enrichment p value non-significant), suggesting that the LIN28B⁺ EwS phenotype may arise not

only from the inhibition of let-7 biogenesis but also from the effects that LIN28B may exert on its direct target transcripts.

LIN28B affects diverse biological processes, including cell metabolism and proliferation, independent of its let-7-regulatory function (Balzeau et al., 2017). To gain deeper insight into the biological effects of LIN28B in EwS, we generated RNA-seq expression profiles of EwS1 and EwS2 cells depleted of LIN28B using shRNA2, and performed functional gene set enrichment analysis (GSEA) pathway analysis. We found a marked overlap between the list of downregulated genes in LIN28B-depleted cells and multiple gene sets related to the transcriptional function of EWS-FLI1 in EwS (Figure 5B, upper left panel, and Figure S3A). The top-scoring gene set identified by this approach was *Riggi_Ewing_Sarcoma_Progenitor_Up*, which contains the list of genes induced by EWS-FLI1 in human pediatric mesenchymal stem cells (hpMSCs) (Riggi et al., 2010). Direct comparison of this gene set to the genes repressed by LIN28B depletion revealed a highly significant overlap in both primary cell cultures (EwS1 $p = 4.8 \times 10^{-16}$, EwS2 $p = 3.4 \times 10^{-37}$; Figure 5B, lower left panel). A set of 32 genes found to be common to the 2 primary cell cultures and the GSEA gene set included numerous known EWS-FLI1 targets (Figure 5B, right panel), suggesting that LIN28B may regulate EWS-FLI1 target gene expression or, more likely, the expression and function of EWS-FLI1 itself.

To address the latter possibility, we depleted EwS1 and EwS2 cells of EWS-FLI1 using a validated shRNA sequence (De Vito et al., 2011) and conducted RNA-seq analysis on these cells and their corresponding control counterparts infected with an shRNA sequence targeting *GFP* mRNA. Comparison of RNA-seq data from EwS1 and EwS2 cells depleted of either LIN28B or EWS-FLI1 revealed a striking overlap between the two expression datasets (p value for common downregulated genes $p = 8.6 \times 10^{-68}$ and 1.4×10^{-183} for EwS1 and EwS2, respectively; Figures 5C and S3B). We then compared the RNA-seq profiles of LIN28B-depleted EwS1 and EwS2 cells with a curated list of 111 EWS-FLI1-activated genes identified by combining the changes in gene expression observed upon EWS-FLI1 knock down in 2 EwS cell lines (A673 and SK-N-MC) and upon the lentiviral introduction of exogenous EWS-FLI1 into hpMSCs (Table S4) (Boulay et al., 2018). Of the 111 genes, only 99 were expressed in our RNA-seq samples and were used in the comparative analysis. Once again, we found a highly significant overlap for upregulated genes between these datasets from both

Figure 3. LIN28B Is Essential for Self-Renewal and Tumorigenic Properties of LIN28B⁺ EwS Cells

(A) qRT-PCR assessment of *LIN28B* transcript (left panel) and mature let-7 (center panel and right panels) levels in EwS1 and EwS2 cells transduced with *LIN28B*-targeting shRNAs (sh1 LIN28B and sh2 LIN28B) compared to those in the same cells transduced with GFP-targeting shRNA controls (sh CTRL). Mean values \pm SEMs of 3 independent experiments are shown. Two-way ANOVA was used for statistical analysis.

(B) Western blot analysis of LIN28B protein levels in EwS1 and EwS2 cells upon shRNA-mediated LIN28B depletion in (A).

(C) Micrographs of EwS1 and EwS2 spherogenic cultures 96 h after lentiviral infection with either control or *LIN28B*-targeting shRNAs.

(D) Fluorescence-activated cell sorting (FACS) analysis of EwS1 sphere viability 96 h post-transduction with either control or *LIN28B*-targeting shRNA vectors.

(E) Clonogenic assay of EwS1 and EwS2 cultures depleted or not depleted of LIN28B (means \pm SDs, $n = 3$ technical replicates). Two-way ANOVA was used for statistical analysis.

(F) *In vivo* tumorigenicity assay after the injection of 1×10^4 control or LIN28B-depleted EwS1 and EwS2 cells. The loss of LIN28B expression results in virtually complete abrogation of the tumorigenic potential of EwS1 and EwS2 cells. The bar indicates the mean tumor volume and n indicates the number of mice that developed tumors out of the total number of mice used in each experiment. Fisher's exact test was used for the statistical analysis of tumor counts only. Full symbols indicate visible tumors; empty symbols indicate undetectable tumor growth.

* $p > 0.05$; *** $p \leq 0.001$; **** $p \leq 0.0001$. See also Figure S2.

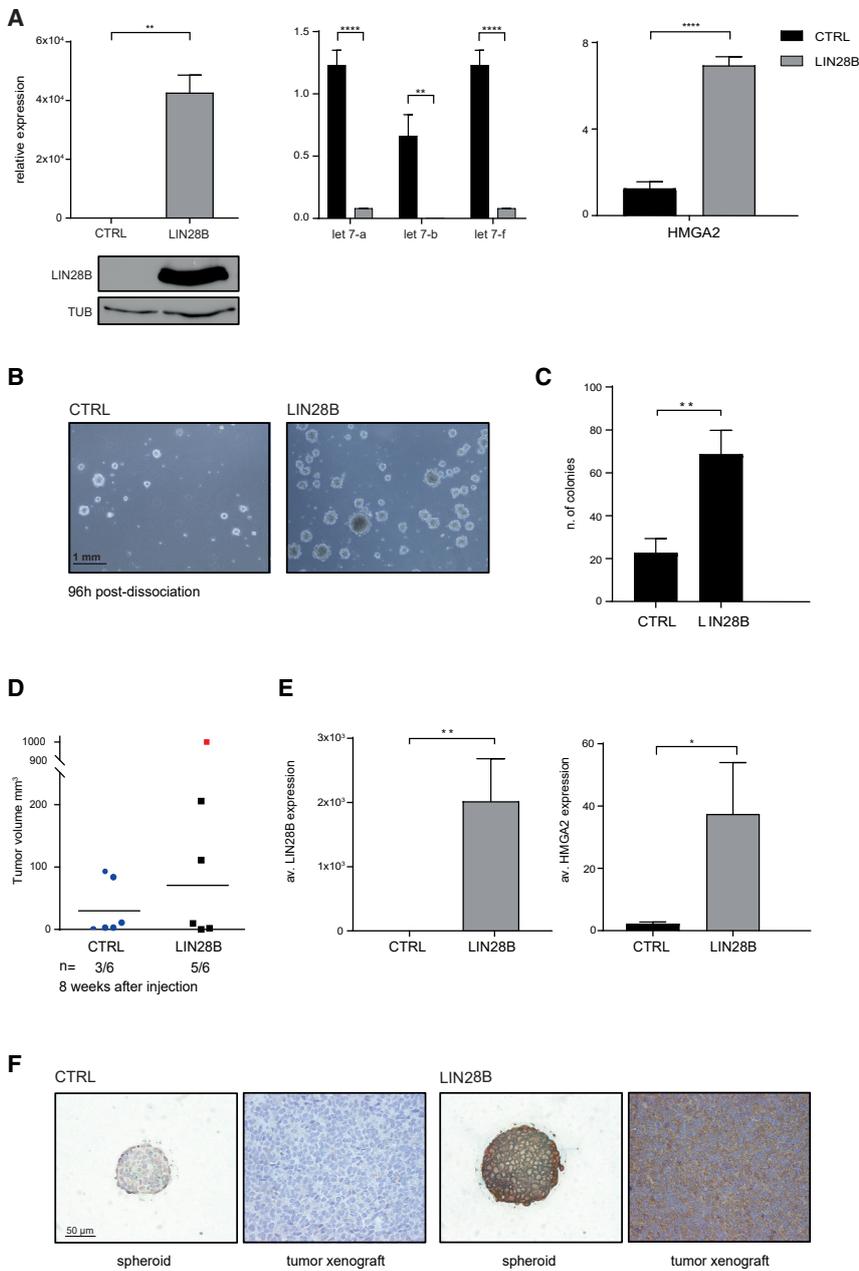


Figure 4. LIN28B Expression Increases LIN28B⁻ EwS Cell Self-Renewal and Tumorigenicity

(A) Left panel: qRT-PCR (top) and western blot (bottom) analyses of LIN28B expression in EwS3 spheres transduced with an empty (CTRL) vector, compared to spheres transduced with a LIN28B-expressing vector (LIN28B), compared to spheres transduced with an empty (CTRL) vector. Mean values \pm SEMs of 3 independent experiments are shown. Center and right panels: qRT-PCR assessment of mature let-7a, b, and f (center) and HMGA2 (right) transcript levels in LIN28B-expressing EwS3 spheres compared to their empty-vector-infected counterparts. The means \pm SEMs of 3 independent experiments are shown.

(B) Micrographs of EwS3 cultures 96 h after dissociation, showing a marked increase in the spherogenic ability of cells transduced with LIN28B.

(C) Clonogenic assays of LIN28B-expressing EwS3 cells compared to their empty-vector-infected counterparts (CTRL, means \pm SDs).

(D) Injection of 1×10^4 EwS3 cells into NSG mice reveals increased tumor initiation by EwS3 cells expressing exogenous LIN28B. The letter "n" indicates the number of mice that developed tumors out of the total number of mice used in each experiment. The bar indicates the mean tumor volume. The outlier (red square) was not included in the calculation of the mean.

(E) Average LIN28B (left) and HMGA2 (right) transcript levels in tumors from LIN28B-expressing EwS3 cells, as assessed by qRT-PCR (means \pm SDs, n = 3 technical replicates).

(F) Immunohistochemical analysis of LIN28B protein expression in control and LIN28B-expressing EwS3 spheroids, as well as in the tumor xenografts derived from injection of the corresponding cell populations into NSG mice. Student's t test was used to perform statistical analyses in (A) and (C) (right panel) and (E) and (G); 2-way ANOVA for analysis in (C) (left panel).

*p < 0.05; **p < 0.001; ***p \leq 0.0001; ****p \leq 0.0001.

primary EwS cell cultures ($p = 4.86 \times 10^{-26}$ and 6.8×10^{-51} for EwS1 and EwS2, respectively; Figures 5C and S3B). The same held true when the analysis was repeated for downregulated genes (Figure S3C). These data strongly support a direct functional relation between LIN28B and EWS-FLI1.

To determine the nature of this putative functional relation, we assessed possible changes in EWS-FLI1 expression upon LIN28B depletion in our two primary cell cultures. In both EwS1 and EwS2 cells, the depletion of LIN28B resulted in a robust decrease in EWS-FLI1 transcript and protein levels (Figures 5D and S3D). The effect on wild-type (WT) EWS was also

significant in EwS1 cells, albeit slightly less marked, whereas it was not significant in EwS2 cells (Figure S3D). In contrast, LIN28B expression increased upon EWS-FLI1 knockdown (Figure S3E), excluding a reciprocal positive feedback loop between the two genes. A comparable decrease in the expression of a panel of direct EWS-FLI1 target genes in response to the depletion of either LIN28B or EWS-FLI1 (Figure 5E, left and center panels) further validated these results. Similar observations were made upon the depletion of LIN28B from EwS1 cells using CRISPR-Cas9 (Figure 5E, right panel).

LIN28B Stabilizes EWS-FLI1 Transcripts

To address the possibility that LIN28B may protect EWS-FLI1 transcripts from let-7-mediated silencing, we attempted to over-express let-7a in EwS1 and EwS2 spheres. However, we were

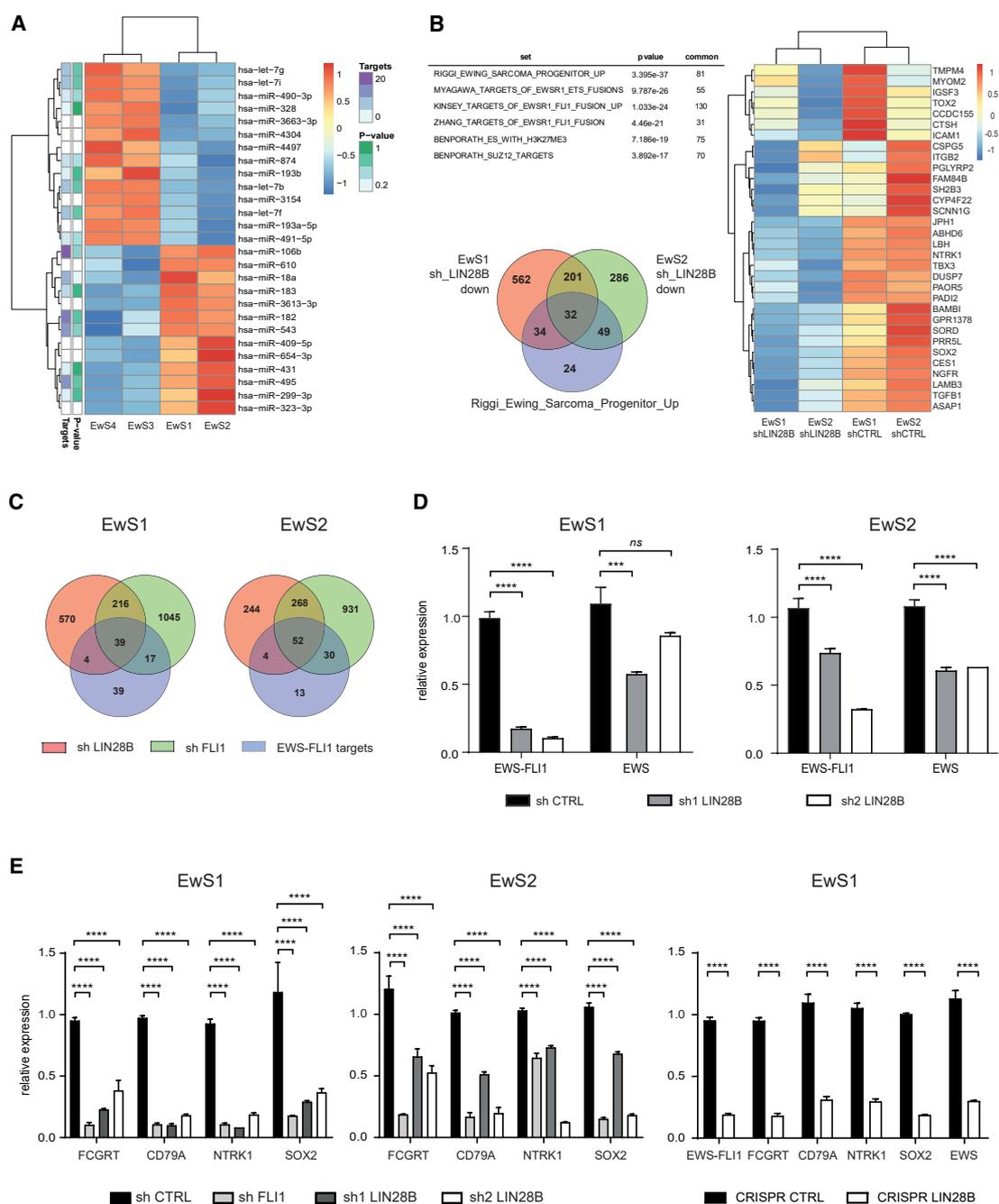


Figure 5. LIN28B Regulates EWS-FLI1 Expression in LIN28B⁺ EwS Cells

(A) Heatmap of Z scores of differentially expressed miRNAs between LIN28B⁺ and LIN28B⁻ EwS cells. The number of discordant predicted targets (down-regulated in response to upregulated miRNAs and vice versa) and the corresponding enrichment p values are shown at left. White cells correspond to poorly conserved miRNA families for which no high-confidence target predictions are available in TargetScan.

(B) Top panel: functional GSEA pathway analysis of EwS2 cells depleted of LIN28B using shRNA2. Bottom panel: Venn diagram depicting the overlap among genes repressed by LIN28B depletion in EwS1 and EwS2 cells and the “*Riggi_Ewing_Sarcoma_Progenitor_Up*” GSEA gene set (bottom). Right panel: gene expression heatmap of the 32 shared genes in the Venn diagram.

(C) Venn diagrams illustrating the overlap between genes repressed by *LIN28B* and *EWS-FLI1* depletion in EwS1 and EwS2 cells, as well as part of a curated list of 99 direct EWS-FLI1 target transcripts.

(D) qRT-PCR assessment of *EWS-FLI1* and WT *EWS* expression in EwS1 and EwS2 cells following shRNA-mediated *LIN28B* depletion. Means \pm SEMs of 3 independent experiments are shown.

(legend continued on next page)

unable to obtain overexpression, most likely because of let-7 toxicity in EwS cells (De Vito et al., 2011). We therefore interrogated the TargetScanHuman 7.1 web tool (Agarwal et al., 2015), but found that *EWS-FLI1* mRNA does not contain any let-7 binding sites. A reasonable assumption then may be that as an RBP targeting a broad repertoire of mRNAs, LIN28B binds *EWS-FLI1* transcripts and preserves their expression. Consistent with this hypothesis, *EWS-FLI1* was enriched in LIN28B RNA immunoprecipitation (RIP) assays on the lysates of both EwS1 and EwS2 cells (Figure 6A). WT *EWS* transcripts were also enriched in LIN28B RIPs (Figure 6A), whereas *FLI1* mRNA is not expressed in EwS cells. To distinguish between the possible effects of LIN28B on *EWS-FLI1* transcription versus stability, we assessed *EWS-FLI1* and its target gene nascent mRNA levels in the presence and absence of LIN28B. Whereas nascent *EWS-FLI1* transcripts were unaffected by LIN28B depletion, all of its selected target gene transcript levels displayed a marked decrease (Figure 6B), refuting LIN28B-mediated control of *EWS-FLI1* transcription and pointing toward the regulation of its stability. To further examine this potential mechanism of action, actinomycin D (ActD)-mediated blockade of transcription was conducted in EwS2 cells in the presence and absence of LIN28B, and *EWS-FLI1* transcript levels were measured by qRT-PCR. Depletion of LIN28B led to the observed decrease in *EWS-FLI1* transcripts, and treatment with ActD resulted in a significant, additional decrease in *EWS-FLI1* expression (Figure 6C, upper panel). To obtain deeper insight into the effect of LIN28B on *EWS-FLI1* stability, we addressed changes in *EWS-FLI1* expression at early time points following ActD administration in the presence or absence of LIN28B. ActD alone revealed that the half-life of *EWS-FLI1* was 5.9 and 5.2 h in EwS1 and EwS2 cells, respectively (Figures 6C and S3F). In the absence of LIN28B, the *EWS-FLI1* transcript half-life was reduced to 2 h in both primary cultures (Figures 6C and S3F). In the LIN28B⁻ EwS3 cells, the half-life of *EWS-FLI1* was 1.7 h, comparable to that in EwS1 and EwS2 cells depleted of LIN28B (Figure S3G). These observations support the notion that by directly binding *EWS-FLI1* transcripts, LIN28B protects them from degradation. Accordingly, the depletion of either LIN28B or *EWS-FLI1* in EwS1 and EwS2 cells produced a comparable decrease in their proliferation (Figure S3H).

The demonstration that LIN28B stabilizes *EWS-FLI1* mRNAs suggests that it may boost *EWS-FLI1* expression. However, the expression of *EWS-FLI1* and that of its direct target genes was not higher in LIN28B⁺ than in LIN28B⁻ EwS cells (Figure S4A), nor did the expression of *EWS-FLI1* increase in EwS3 cells (LIN28B⁻) upon the introduction of *LIN28B*^{43UTR} (Figure S4B). Furthermore, comparison of differentially expressed genes in EwS1 and EwS2 cells depleted of LIN28B to those in EwS3 cells overexpressing LIN28B revealed no significant overlap (Figure S4C), indicating that exogenous LIN28B expression in LIN28B⁻ EwS cells does not recreate the naturally occurring LIN28B⁺ EwS phenotype.

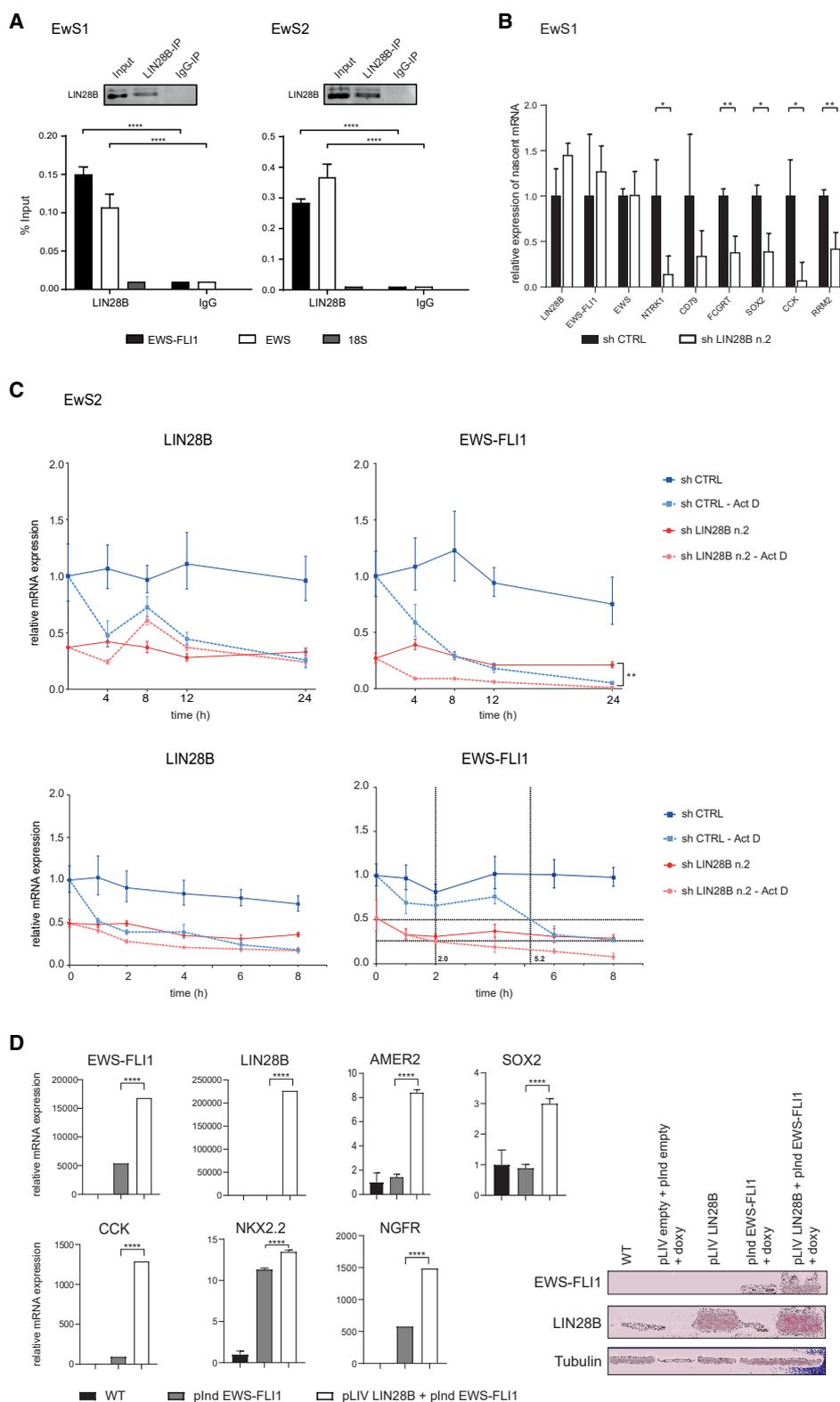
LIN28B⁻ EwS cells may have developed different mechanisms to ensure *EWS-FLI1* stability, rendering LIN28B redundant. Full assessment of the effect of LIN28B may therefore require a cellular environment that is permissive for *EWS-FLI1* expression and function but that does not naturally express the fusion protein. To test this possibility, we engineered primary hpMSCs to conditionally express *EWS-FLI1* cDNA under the control of a doxycycline-inducible promoter and infected them with lentiviruses containing *LIN28B* or an empty control vector. Expression of *EWS-FLI1* was then induced by treating the cells with doxycycline and the expression level of *EWS-FLI1*, as well as that of a panel of its target genes in LIN28B-expressing and LIN28B⁻ hpMSCs, were compared. Using 2 unrelated primary hpMSC batches, we observed ~3-fold higher *EWS-FLI1* expression in cells expressing LIN28B and an even greater difference in its expression at the protein level (Figures 6D and S4D). The augmented *EWS-FLI1* expression in LIN28B-expressing hpMSCs was accompanied by a corresponding increase in the expression of a selection of its target genes, including *AMER2*, *SOX2*, *CCK*, *NKX2.2*, and *NGFR* (Figures 6D and S4D), ranging from slightly less than 1.5-fold for *NKX2.2* to >10-fold for *CCK*. These observations support the LIN28B-mediated stabilization of *EWS-FLI1* transcripts in the appropriate cellular environment. They also confirm that the mechanism of *EWS-FLI1* maintenance by LIN28B is let-7 independent, since the *EWS-FLI1* expression construct used in hpMSCs lacks the 3' UTR. These observations point to a hitherto unrecognized regulatory mechanism of *EWS-FLI1* expression that is exploited by a subclass of EwS.

Treatment with LIN28 Inhibitor 1632 Impairs *In Vitro* Clonogenicity and *In Vivo* Growth of LIN28B⁺ but Not LIN28B⁻ EwS Cells

The observation that LIN28B stabilizes *EWS-FLI1* expression renders its inhibition a potentially attractive means to treat the subset of tumors that express LIN28B. Recently, the small molecule 1632 (*N*-methyl-*N*-[3-(3-methyl[1,2,4]triazolo[4,3-*b*]pyridazin-6-yl)phenyl]acetamide) was shown to block the recognition of let-7 miRNA precursors by the LIN28 paralogs (Roos et al., 2016). Because the let-7 miRNA family directly regulates LIN28 expression, we reasoned that the 1632 inhibitor may provide a pharmacological means to disrupt *EWS-FLI1* expression in LIN28B⁺ EwS, by the repression of LIN28B itself. Treatment with the 1632 inhibitor at 2 different concentrations (50 and 250 μ M) for different durations (4–7 days) elicited different responses in LIN28B⁺ and LIN28B⁻ EwS cells. In EwS1 and EwS2 cells, the compound caused a robust increase in let-7 maturation (Figure 7A, left panel, and Figure S5A; data not shown) and a corresponding decrease in the expression of the LIN28B target gene *HMG2A* (Figure 7A, right panel). In contrast, the expression of immature and mature let-7 forms did not undergo a significant change in EwS3 and EwS4 cells in response to the inhibitor (Figure 7A, left panel; data not shown), and *HMG2A* expression displayed no significant decrease in either

(E) Expression of a panel of known direct *EWS-FLI1* target genes in EwS1 and EwS2 cells following shRNA- (left and center panels) and CRISPR-Cas9-mediated (right panel) *LIN28B* depletion. shRNA- and CRISPR-Cas9-LIN28B-depleted cells were compared to *GFP*-targeting shRNA and CRISPR-Cas9 control guide RNA vector-infected cells, respectively. Means \pm SEMs of 3 independent experiments are shown. Two-way ANOVA was used for statistical analysis.

p \leq 0.001; *p \leq 0.0001; ns, non-significant. See also Figures S3 and S4 and Table S4.



(legend on next page)

primary cell culture (Figure 7A, right panel). Consistent with these observations, functional assays revealed that after 96 h of treatment with the inhibitor at 250 μ M, EwS1 and EwS2 cells lost their ability to re-form spheres following dissociation, whereas sphere formation by EwS3 and EwS4 remained unaffected irrespective of the concentration of the compound or treatment duration (Figures 7B and S5B, upper panel; data not shown). Consistent with our observations on the effect of LIN28B depletion, survival of the 4 primary cultures was not significantly affected by the compound (Figure S5B, lower panel). Clonogenic assays using all 4 primary cell cultures pre-treated with the inhibitor for 96 h and then treated continuously at 250 μ M for 2 weeks showed an unequivocal decrease in sphere-forming capacity by single EwS1 and EwS2 cells, but no change in EwS3 and EwS4 cell clonogenicity (Figure 7C).

We next sought to determine whether the treatment of LIN28B⁺ cells with the 1632 inhibitor results in the expected decrease in EWS-FLI1 expression, as initially hypothesized. In response to the compound at 250 μ M, both EWS and EWS-FLI1 protein expression was abolished in EwS1 cells (Figure S5C). We then performed an RNA-seq analysis on EwS1 spheres that had been cultured in the presence of the inhibitor for 7 days and compared the resulting gene expression profile with the ones obtained following LIN28B and EWS-FLI1 depletion. Once again, we observed a statistically significant overlap between the expression profile of 1632-treated EwS1 cells and the different datasets ($p = 1.32 \times 10^{-34}$, 3.77×10^{-51} , and 4.92×10^{-23} for shLIN28B, shEWS-FLI1, and EWS-FLI1 target genes, respectively; Figures 7D and S5D). These results were validated by the assessment of changes in the expression of LIN28B, EWS-FLI1, and EWS-FLI1 target genes in EwS1 cells treated with the 1632 inhibitor for 7 days (Figure 7E).

As the 1632 compound has not been tested *in vivo*, we could not assess the effect of its continuous infusion on tumors growing in mice. Nevertheless, to determine whether the observed effect on clonogenicity is reflected by the inhibition of tumor growth *in vivo*, we injected 1×10^4 EwS1 and EwS2 cells under the kidney capsule of NSG mice after 96 h of pre-treatment with 250 μ M of the 1632 inhibitor or DMSO. All of the mice were sacrificed 8 weeks post-injection, when tumors in untreated mice had grown to the maximal size allowed by the animal experimentation ethics commission. Despite the suboptimal treatment conditions, we observed a robust decrease in

the size of tumors derived from treated EwS1 and EwS2 cells (Figure 7F). These observations demonstrate the feasibility of suppressing EWS-FLI1 and its oncogenic effects by targeting the LIN28B-dependent mechanism that maintains its expression.

DISCUSSION

Using an unbiased approach to search for factors that associate with EWS-FLI1 in promoting EwS pathogenesis, we identified LIN28B as a regulator of EWS-FLI1 expression and function in a small percentage of EwSs that depend on LIN28B for the expression and maintenance of their identity. A possible outcome of this discovery is that EwSs expressing LIN28B may be sensitive to its inhibition, warranting further investigation of LIN28B as a therapeutic target in some cases of EwS.

The RBPs LIN28A and LIN28B abrogate the processing of primary and precursor let-7 family hairpins into mature let-7 miRNAs (Mayr and Heinemann, 2013; Thornton and Gregory, 2012; Viswanathan and Daley, 2010) and help maintain normal development and pluripotency by preventing let-7-induced differentiation in mouse embryonic stem cells (Zhang et al., 2016). Conversely, mature let-7 represses LIN28B translation as part of a negative feedback loop (Thornton and Gregory, 2012; Zhang et al., 2016). The maintenance of pluripotency and the protection of numerous oncogenes from let-7-mediated silencing are thought to be the principal mechanisms by which LIN28B promotes cancer growth and progression. However, as our observations suggest, LIN28B may also exert oncogenic properties by directly binding to mRNA (Mayr and Heinemann, 2013), which is consistent with the notion that the most abundant class of RNAs bound by LIN28B is protein coding transcripts (Hafner et al., 2013). A recently identified LIN28B target mRNA was *TLS/FUS*, a close relative of *EWS*, which, similar to *EWS*, partners with several genes to generate oncogenic fusion proteins (Wilbert et al., 2012).

The introduction of LIN28B into LIN28B⁻ primary EwS cells conferred increased proliferation and clonogenicity, as well as accelerated tumor-initiating capacity onto the cells, all of which are consistent with the effects that LIN28B exerts on diverse tumor cell types (Viswanathan et al., 2009). The mechanisms underlying these effects are most likely a combination of let-7 maturation suppression and modulation of the stability of the

Figure 6. LIN28B Controls EWS-FLI1 Transcript Stability in LIN28B⁺ EwS Cells

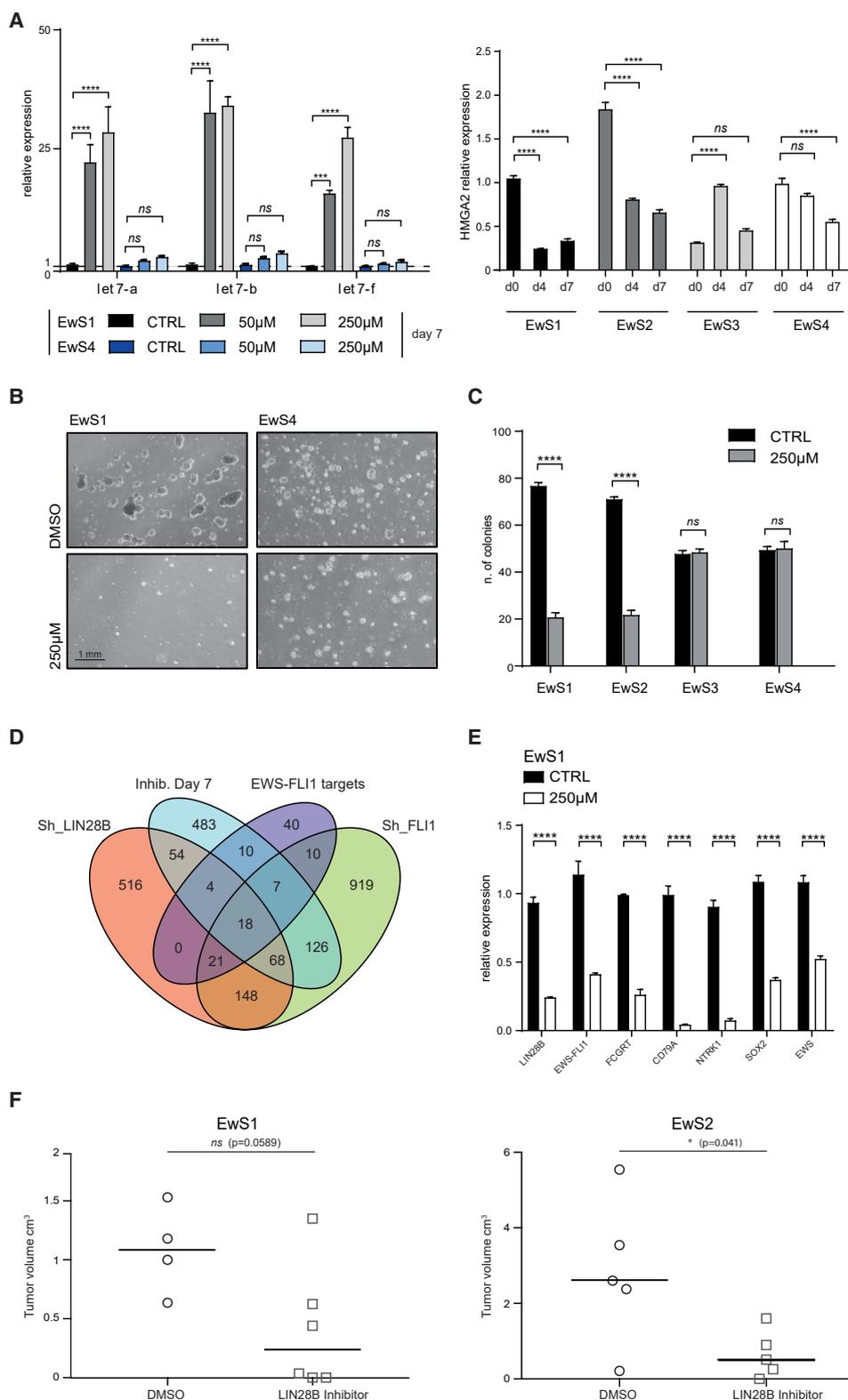
(A) RNA immunoprecipitation assay (RIP). qRT-PCR analysis of *EWS-FLI1* and WT *EWS* transcripts in EwS1 and EwS2 cells following LIN28B protein immunoprecipitation (upper panel). The *18S* gene was used as an internal control, and results were compared to non-specific isotype-matched immunoglobulin G (IgG) immunoprecipitates. Student's t test was used for statistical analysis.

(B) Nascent RNA assay for *EWS-FLI1* and EWS-FLI1 target transcripts in the presence and absence of LIN28B in EwS1 cells. LIN28B-targeting shRNA has no effect on nascent *EWS-FLI1* mRNA, whereas it strongly depletes EWS-FLI1 target transcripts (means \pm SDs of 3 technical replicates).

(C) Analysis of transcript levels upon ActD treatment, suggesting that LIN28B regulates *EWS-FLI1* RNA stability in EwS2 cells. EwS2 cells were subjected to LIN28B depletion for 36 h, after which 10 μ M ActD or solvent (equivalent volume of DMSO) was added for an additional 24 h, during which *LIN28B* and *EWS-FLI1* expression was measured (upper panel). To determine the half-life of *EWS-FLI1*, the decay rate was measured by qRT-PCR at 5 time points over 8 h in the presence of ActD (lower panel). *EWS-FLI1* transcript half-life, as estimated from the expression curves, is indicated.

(D) Left panel: expression of EWS-FLI1, LIN28B, and a panel of known direct EWS-FLI1 target genes in hpMSC1, infected with plnd EWS-FLI1 or co-infected with plnd EWS-FLI1 and pLIV LIN28B, respectively (mean values \pm SDs of 3 technical replicates). EWS-FLI1 was induced by doxycycline treatment for 8 days. Student's t test was used for statistical analysis. Right panel: western Blot analysis of EWS-FLI1 and LIN28B expression in hpMSC1 infected with the indicated lentiviral vectors and cultured with (+ doxy) or without doxycycline for 8 days.

* $p \leq 0.05$, ** $p \leq 0.01$; **** $p \leq 0.0001$. See also Figures S3 and S4.



(legend on next page)

available repertoire of direct LIN28B target transcripts. However, exogenous LIN28B affected neither EWS-FLI1 expression nor that of its target gene repertoire in EwS cells. Thus, the changes in the behavior of these cells in response to the introduction of LIN28B were not the result of altered EWS-FLI1 expression or function but rather the superposition of the respective oncogenic properties of LIN28B and EWS-FLI1.

In stark contrast was the effect of depleting LIN28B⁺ EwS cells of their endogenous LIN28B, which included a dramatic decrease in EWS-FLI1 expression along with that of its direct target genes, accompanied by the loss of self-renewal and tumor initiation. By far the most dominant trait of the transcriptome change associated with LIN28B depletion was the suppression of direct EWS-FLI1 target transcripts. Changes related to let-7 maturation and binding of LIN28B to other mRNA targets were not comparable to those of the EWS-FLI1 target gene repertoire. From a functional standpoint, depletion of EWS-FLI1 or LIN28B alone from these cells decreased their proliferation to a similar degree, despite the fact that LIN28B transcripts increased more than 2-fold in response to EWS-FLI1 knockdown. LIN28B therefore appears to provide indispensable support for EWS-FLI1 expression and function in these cells, and the absence of let-7 recognition sites in *EWS-FLI1* argues that LIN28B itself bears responsibility for the stability of *EWS-FLI1*.

The depletion of LIN28B in LIN28B⁺ EwS cells resulted in a decrease in *EWS-FLI1* transcript half-life to a level in the vicinity of that observed in LIN28B⁻ cells. As EWS-FLI1 expression in LIN28B⁻ cells is sufficient to allow it to fulfill its oncogenic functions, these cells must either possess an alternative mechanism to ensure *EWS-FLI1* stability or lack the putative machinery that degrades *EWS-FLI1* and is countered by LIN28B in LIN28B⁺ cells. Alternatively, the stability of EWS-FLI1 may be ensured at the protein level in LIN28B⁻ EwS cells. Regardless of the mechanism, and unlike hpMSCs, LIN28B⁻ EwS cells appear to have established, possibly over time, regulatory mechanisms of *EWS-FLI1* expression that are independent of and unresponsive to LIN28B.

A more fundamental question is why should LIN28B be indispensable for EWS-FLI1 expression in a subset of EwS? One possibility is that LIN28B provides the permissiveness required for EWS-FLI1-mediated transformation of the primary cells from which the LIN28B⁺ EwS subclass originates. This notion would

suggest that LIN28B⁺ and LIN28B⁻ EwS may have different cells of origin. In response to LIN28B expression, hpMSCs increase their proliferation, upregulate EWS-FLI1 target genes, and augment EWS-FLI1 expression. However, they do not require LIN28B expression for long-term EWS-FLI1 maintenance (Riggi et al., 2008). In most malignancies in which LIN28B expression becomes induced as a consequence of reprogramming linked to transformation, its expression is heterogeneous and is often confined to poorly differentiated cells that may display CSC features (Balzeau et al., 2017). In contrast, LIN28B⁺ EwS express LIN28B in virtually all cells, suggesting that these tumors may arise from a stem or an early progenitor cell that has not turned off LIN28B expression. Interrogation of a published expression array of 33 normal adult tissues, which includes human neural crest stem cells (NCSCs) and MSCs (GEO accession: GSE68776), revealed high LIN28B expression in NCSCs and NC-derived MSCs (NC-MSCs) but low to undetectable expression in bone marrow MSCs (BM-MSCs) (Figure S4E). All 3 cell types are permissive for EWS-FLI1 expression (Riggi et al., 2005, 2008, 2010; von Levetzow et al., 2011), but their permissiveness may invoke different mechanisms. Thus, the expression of *EWS-FLI1* in NCSCs and NC-MSCs may require LIN28B and thereby lead to the development of LIN28B⁺ EwS, whereas in BM-MSCs, it may result in LIN28B⁻ tumors, which may rely on alternative mechanisms to ensure *EWS-FLI1* stability.

Treatment of EwS cells with the 1632 compound that blocks LIN28B binding to let-7 precursors and disrupts the LIN28B-mediated inhibition of let-7 maturation did not affect LIN28B⁻ cell behavior, but it did abrogate LIN28B⁺ cell self-renewal. Even limited pre-treatment of LIN28B⁺ cells before injection into mice blunted their tumorigenicity, reflecting the extent of their dependence on LIN28B. Although these effects are consistent with increased let-7 expression, the 1632 inhibitor also caused a decrease in EWS-FLI1 and its target gene expression, which are let-7 independent. There are at least two non-mutually exclusive explanations for this seemingly counterintuitive observation. As LIN28B mRNA-binding sites have not been exhaustively elucidated (Mayr and Heinemann, 2013), it is possible that the inhibitor may interfere with LIN28B binding to *EWS-FLI1* transcripts, particularly in view of the discovery that some *LIN28B* mRNA consensus binding sites are similar to those of

Figure 7. The LIN28 Inhibitor 1632 Acts Selectively on LIN28B⁺ EwS Cells, Leading to a Marked Reduction in Their Self-Renewal and Tumorigenic Properties

(A) Left panel: qRT-PCR analysis of mature let7-a, b, and f expression in EwS1 and EwS4 cells, treated with increasing concentrations of the 1632 inhibitor (50 and 250 μM), compared to that in the same cells treated with solvent alone (equivalent volume of DMSO, CTRL). Mean values ± SEMs of 3 independent experiments are shown. Right panel: qRT-PCR analysis of *HMG2* expression in EwS1–EwS4 cultures treated with 250 μM of the 1632 inhibitor and assessed at days 0, 4, and 7 (means ± SEMs of 3 technical replicates). Two-way ANOVA was used for statistical analysis.

(B) Micrographs of EwS1 and EwS4 sphere cultures after 96 h of treatment with DMSO (CTRL) or 250 μM of the 1632 inhibitor.

(C) Clonogenic assay on EwS1–EwS4 cells treated with DMSO (CTRL) or 250 μM of the 1632 inhibitor (means ± SDs of 3 technical replicates).

(D) Venn diagram showing overlap among 4 gene datasets, including *LIN28B* knockdown, *EWS-FLI1* knockdown, and treatment with 250 μM of the 1632 inhibitor at 7 days (common downregulated genes); and the list of 99 direct EWS-FLI1 target genes described in Figure 5C.

(E) qRT-PCR assessment of *LIN28B*, *EWS-FLI1*, and WT *EWS* expression, as well as that of a panel of known direct EWS-FLI1 target genes in EwS1 cells following 7 days of treatment with the 1632 compound compared to that in the same cells treated with solvent alone for the same period of time. Mean values ± SEMs from 3 independent experiments are shown. Two-way ANOVA was used for statistical analysis.

(F) *In vivo* tumorigenicity assay after the injection of 1 × 10⁴ EwS1 (left panel) or EwS2 (right panel) cells following 96 h of treatment with 250 μM of the 1632 inhibitor or solvent (equivalent volume of DMSO, CTRL). The bar indicates the mean value.

*p ≤ 0.05; ****p ≤ 0.0001; ns, non-significant. See also Figure S5.

let-7 precursors (Hafner et al., 2013; Wilbert et al., 2012). Alternatively, the depletion of *LIN28B* by increased let-7 maturation may result in the progressive loss of EWS-FLI1 expression and deconstruction of its target gene network. Although it may be relevant to a minority of patients, the exquisite dependence of EWS-FLI1 expression on *LIN28B* function may provide an unprecedented therapeutic perspective for EwS.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Primary tumors and cell lines cultures
 - Animal Studies
- METHOD DETAILS
 - Plasmids and Lentiviral Infection
 - RNA *In Situ* Hybridization, Immunohistochemistry and Western Blot Analyses
 - RNA Isolation and Real-Time qPCR
 - Cell vitality, proliferation assays and drug tests
 - Nascent RNA capture and cDNA synthesis
 - *In vivo* tumorigenicity assays
 - RNA Immunoprecipitation
 - Achilles scores and survival analysis
 - *LIN28B* expression in an Ewing sarcoma dataset
 - Analysis of RNA-seq data
 - MicroRNA expression analysis
 - Overlap between differentially expressed genes and gene sets
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.celrep.2019.12.053>.

ACKNOWLEDGMENTS

This work was supported by Swiss National Science Foundation grants 310030_150024 and 310030_169563, Switzerland and Oncosuisse grant KLS-3365-02-2014, Switzerland (to I.S.); and a Swiss National Science Foundation grant PP00P3_157468, Switzerland, an Oncosuisse grant KFS-3973-08-2016, Switzerland, and a MEDIC Foundation grant, Switzerland (to N.R.). M.N.R. is supported by an award from the V Foundation, United States.

AUTHOR CONTRIBUTIONS

Conception & Design, N.R. and I.S.; Investigation & Methodology, M.T., G.B., S.C.-B., P.W., N.C., A.B., A.V., A.K., T.K., T.G., C.F., and P.M.; Data Analysis, M.N.R., P.P., N.R., S.I., and M.L.S.; Resources, I.L., S.C., G.M.C., E.C., A.D., M.M., I.C., P.G.N., A.M.C., J.M., R.R., and I.S.; Writing, Review, & Editing, G.B., N.R., M.N.R., and I.S.; Study Supervision, N.R. and I.S.; Funding Acquisition, I.S.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 2, 2018

Revised: June 19, 2019

Accepted: December 13, 2019

Published: March 31, 2020; corrected online April 6, 2020

REFERENCES

- Agarwal, V., Bell, G.W., Nam, J.W., and Bartel, D.P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. *eLife* 4, e05005.
- Aguirre, A.J., Meyers, R.M., Weir, B.A., Vazquez, F., Zhang, C.Z., Ben-David, U., Cook, A., Ha, G., Harrington, W.F., Doshi, M.B., et al. (2016). Genomic Copy Number Dictates a Gene-Independent Cell Response to CRISPR/Cas9 Targeting. *Cancer Discov.* 6, 914–929.
- Aken, B.L., Achuthan, P., Akanni, W., Amode, M.R., Bersndorff, F., Bhai, J., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., et al. (2017). Ensembl 2017. *Nucleic Acids Res.* 45 (D1), D635–D642.
- Baizeau, J., Menezes, M.R., Cao, S., and Hagan, J.P. (2017). The *LIN28/let-7* Pathway in Cancer. *Front. Genet.* 8, 31.
- Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehár, J., Kryukov, G.V., Sonkin, D., et al. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483, 603–607.
- Boulay, G., Sandoval, G.J., Riggi, N., Iyer, S., Buisson, R., Naigles, B., Awad, M.E., Rengarajan, S., Volorio, A., McBride, M.J., et al. (2017). Cancer-Specific Retargeting of BAF Complexes by a Prion-like Domain. *Cell* 171, 163–178.e19.
- Boulay, G., Volorio, A., Iyer, S., Broye, L.C., Stamenkovic, I., Riggi, N., and Rivera, M.N. (2018). Epigenome editing of microsatellite repeats defines tumor-specific enhancer functions and dependencies. *Genes Dev.* 32, 1008–1019.
- Brohl, A.S., Solomon, D.A., Chang, W., Wang, J., Song, Y., Sindiri, S., Patidar, R., Hurd, L., Chen, L., Shern, J.F., et al. (2014). The genomic landscape of the Ewing Sarcoma family of tumors reveals recurrent *STAG2* mutation. *PLoS Genet.* 10, e1004475.
- Carmel-Gross, I., Bollag, N., Armon, L., and Urbach, A. (2016). *LIN28*: A Stem Cell Factor with a Key Role in Pediatric Tumor Formation. *Stem Cells Dev.* 25, 367–377.
- Carvalho, B.S., and Irizarry, R.A. (2010). A framework for oligonucleotide microarray preprocessing. *Bioinformatics* 26, 2363–2367.
- Chatterji, P., and Rustgi, A.K. (2018). RNA Binding Proteins in Intestinal Epithelial Biology and Colorectal Cancer. *Trends Mol. Med.* 24, 490–506.
- Chatterji, P., Hamilton, K.E., Liang, S., Andres, S.F., Wijeratne, H.R.S., Mizuno, R., Simon, L.A., Hicks, P.D., Foley, S.W., Pitarresi, J.R., et al. (2018). The *LIN28B-IMP1* post-transcriptional regulon has opposing effects on oncogenic signaling in the intestine. *Genes Dev.* 32, 1020–1034.
- Choi, S.H., Kim, S.H., Shim, K.W., Han, J.W., Choi, J., Kim, D.S., Lyu, C.J., Kim, J.W., Suh, C.O., and Cho, J. (2016). Treatment Outcome and Prognostic Molecular Markers of Supratentorial Primitive Neuroectodermal Tumors. *PLoS One* 11, e0153443.
- Chuang, T.H., and Ulevitch, R.J. (2004). *Triad3A*, an E3 ubiquitin-protein ligase regulating Toll-like receptors. *Nat. Immunol.* 5, 495–502.
- Cornaz-Buros, S., Riggi, N., DeVito, C., Sarre, A., Letovanec, I., Provero, P., and Stamenkovic, I. (2014). Targeting cancer stem-like cells as an approach to defeating cellular heterogeneity in Ewing sarcoma. *Cancer Res.* 74, 6610–6621.
- Crompton, B.D., Stewart, C., Taylor-Weiner, A., Alexe, G., Kurek, K.C., Calicchio, M.L., Kiezun, A., Carter, S.L., Shukla, S.A., Mehta, S.S., et al. (2014). The genomic landscape of pediatric Ewing sarcoma. *Cancer Discov.* 4, 1326–1341.

- De Vito, C., Riggi, N., Suvà, M.L., Janiszewska, M., Horlbeck, J., Baumer, K., Provero, P., and Stamenkovic, I. (2011). Let-7a is a direct EWS-FLI-1 target implicated in Ewing's sarcoma development. *PLoS One* 6, e23592.
- De Vito, C., Riggi, N., Cornaz, S., Suvà, M.L., Baumer, K., Provero, P., and 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, 807–821.
- Dunker, A.K., and Uversky, V.N. (2010). Drugs for 'protein clouds': targeting intrinsically disordered transcription factors. *Curr. Opin. Pharmacol.* 10, 782–788.
- Engert, F., Schneider, C., Weiß, L.M., Probst, M., and Fulda, S. (2015). PARP Inhibitors Sensitize Ewing Sarcoma Cells to Temozolomide-Induced Apoptosis via the Mitochondrial Pathway. *Mol. Cancer Ther.* 14, 2818–2830.
- Erkizan, H.V., Kong, Y., Merchant, M., Schlottmann, S., Barber-Rotenberg, J.S., Yuan, L., Abaan, O.D., Chou, T.H., Dakshanamurthy, S., Brown, M.L., et al. (2009). A small molecule blocking oncogenic protein EWS-FLI1 interaction with RNA helicase A inhibits growth of Ewing's sarcoma. *Nat. Med.* 15, 750–756.
- Filbin, M.G., Tirosh, I., Hovestadt, V., Shaw, M.L., Escalante, L.E., Mathewson, N.D., Neftel, C., Frank, N., Pelton, K., Hebert, C.M., et al. (2018). Developmental and oncogenic programs in H3K27M gliomas dissected by single-cell RNA-seq. *Science* 360, 331–335.
- Hafner, M., Max, K.E., Bandaru, P., Morozov, P., Gerstberger, S., Brown, M., Molina, H., and Tuschl, T. (2013). Identification of mRNAs bound and regulated by human LIN28 proteins and molecular requirements for RNA recognition. *RNA* 19, 613–626.
- Henchen, M., Stubbusch, J., Abarchan-EI Makhfi, I., Kramer, M., Deller, T., Pierre-Eugene, C., Janoueix-Lerosey, I., Delattre, O., Ernsberger, U., Schulte, J.B., and Rohrer, H. (2015). Lin28B and Let-7 in the Control of Sympathetic Neurogenesis and Neuroblastoma Development. *J. Neurosci.* 35, 16531–16544.
- Javelaud, D., Wietzerbin, J., Delattre, O., and Besançon, F. (2000). Induction of p21Waf1/Cip1 by TNF α requires NF- κ B activity and antagonizes apoptosis in Ewing tumor cells. *Oncogene* 19, 61–68.
- Jiang, J., Lee, E.J., Gusev, Y., and Schmittgen, T.D. (2005). Real-time expression profiling of microRNA precursors in human cancer cell lines. *Nucleic Acids Res.* 33, 5394–5403.
- Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12, 323.
- Madison, B.B., Liu, Q., Zhong, X., Hahn, C.M., Lin, N., Emmett, M.J., Stanger, B.Z., Lee, J.S., and Rustgi, A.K. (2013). LIN28B promotes growth and tumorigenesis of the intestinal epithelium via Let-7. *Genes Dev.* 27, 2233–2245.
- Madison, B.B., Jegathanan, A.N., Mizuno, R., Winslow, M.M., Castells, A., Cuatrecasas, M., and Rustgi, A.K. (2015). Let-7 Represses Carcinogenesis and a Stem Cell Phenotype in the Intestine via Regulation of Hmga2. *PLoS Genet.* 11, e1005408.
- Mayr, F., and Heinemann, U. (2013). Mechanisms of Lin28-mediated miRNA and mRNA regulation—a structural and functional perspective. *Int. J. Mol. Sci.* 14, 16532–16553.
- Nakhaei, P., Mesplede, T., Solis, M., Sun, Q., Zhao, T., Yang, L., Chuang, T.H., Ware, C.F., Lin, R., and Hiscott, J. (2009). The E3 ubiquitin ligase Triad3A negatively regulates the RIG-I/MAVS signaling pathway by targeting TRAF3 for degradation. *PLoS Pathog.* 5, e1000650.
- Picard, D., Miller, S., Hawkins, C.E., Bouffet, E., Rogers, H.A., Chan, T.S., Kim, S.K., Ra, Y.S., Fangusaro, J., Korshunov, A., et al. (2012). Markers of survival and metastatic potential in childhood CNS primitive neuro-ectodermal brain tumours: an integrative genomic analysis. *Lancet Oncol.* 13, 838–848.
- Pishas, K.I., and Lessnick, S.L. (2016). Recent advances in targeted therapy for Ewing sarcoma. *F1000Res.* 5, F1000 Faculty Rev-2077.
- Powers, J.T., Tsanov, K.M., Pearson, D.S., Roels, F., Spina, C.S., Ebright, R., Seligson, M., de Soysa, Y., Cahan, P., Theißen, J., et al. (2016). Multiple mechanisms disrupt the let-7 microRNA family in neuroblastoma. *Nature* 535, 246–251.
- R Development Core Team (2019). R: A language and environment for statistical computing (R Foundation for Statistical Computing).
- Riggi, N., and Stamenkovic, I. (2007). The biology of Ewing sarcoma. *Cancer Lett.* 254, 1–10.
- Riggi, N., Cironi, L., Provero, P., Suvà, M.L., Kaloulis, K., Garcia-Echeverria, C., Hoffmann, F., Trumpp, A., and Stamenkovic, I. (2005). Development of Ewing's sarcoma from primary bone marrow-derived mesenchymal progenitor cells. *Cancer Res.* 65, 11459–11468.
- Riggi, N., Suvà, M.L., Suvà, D., Cironi, L., Provero, P., Tercier, S., Joseph, J.M., Stehle, J.C., Baumer, K., Kindler, V., and Stamenkovic, I. (2008). EWS-FLI-1 expression triggers a Ewing's sarcoma initiation program in primary human mesenchymal stem cells. *Cancer Res.* 68, 2176–2185.
- Riggi, N., Suvà, M.L., De Vito, C., Provero, P., Stehle, J.C., Baumer, K., Cironi, L., Janiszewska, M., Petricevic, T., Suvà, D., et al. (2010). EWS-FLI-1 modulates miRNA145 and SOX2 expression to initiate mesenchymal stem cell reprogramming toward Ewing sarcoma cancer stem cells. *Genes Dev.* 24, 916–932.
- Riggi, N., Knoechel, B., Gillespie, S.M., Rheinbay, E., Boulay, G., Suvà, M.L., Rossetti, N.E., Boonseng, W.E., Oksuz, O., Cook, E.B., et al. (2014). EWS-FLI1 utilizes divergent chromatin remodeling mechanisms to directly activate or repress enhancer elements in Ewing sarcoma. *Cancer Cell* 26, 668–681.
- Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43, e47.
- Roos, M., Pradère, U., Ngondo, R.P., Behera, A., Allegrini, S., Civenni, G., Zagalak, J.A., Marchand, J.R., Menzi, M., Towbin, H., et al. (2016). A Small-Molecule Inhibitor of Lin28. *ACS Chem. Biol.* 11, 2773–2781.
- Savola, S., Klami, A., Myllykangas, S., Manara, C., Scotlandi, K., Picci, P., Knuutila, S., and Vakkila, J. (2011). High Expression of Complement Component 5 (C5) at Tumor Site Associates with Superior Survival in Ewing's Sarcoma Family of Tumour Patients. *ISRN Oncol.* 2011, 168712.
- Shinoda, G., Shyh-Chang, N., Soysa, T.Y., Zhu, H., Seligson, M.T., Shah, S.P., Abo-Sido, N., Yabuuchi, A., Hagan, J.P., Gregory, R.I., et al. (2013). Fetal deficiency of lin28 programs life-long aberrations in growth and glucose metabolism. *Stem Cells* 31, 1563–1573.
- Suvà, M.L., Riggi, N., Stehle, J.C., Baumer, K., Tercier, S., Joseph, J.M., Suvà, D., Clément, V., Provero, P., Cironi, L., et al. (2009). Identification of cancer stem cells in Ewing's sarcoma. *Cancer Res.* 69, 1776–1781.
- Thornton, J.E., and Gregory, R.I. (2012). How does Lin28 let-7 control development and disease? *Trends Cell Biol.* 22, 474–482.
- Tirole, F., Laud-Duval, K., Prieur, A., Delorme, B., Charbord, P., and Delattre, O. (2007). Mesenchymal stem cell features of Ewing tumors. *Cancer Cell* 11, 421–429.
- Tirole, F., Surdez, D., Ma, X., Parker, M., Le Deley, M.C., Bahrami, A., Zhang, Z., Lapouble, E., Grossetête-Lalami, S., Rusch, M., et al.; St. Jude Children's Research Hospital–Washington University Pediatric Cancer Genome Project and the International Cancer Genome Consortium (2014). Genomic landscape of Ewing sarcoma defines an aggressive subtype with co-association of STAG2 and TP53 mutations. *Cancer Discov.* 4, 1342–1353.
- Tomazou, E.M., Sheffield, N.C., Schmidl, C., Schuster, M., Schönegger, A., Datlinger, P., Kubicek, S., Bock, C., and Kovar, H. (2015). Epigenome mapping reveals distinct modes of gene regulation and widespread enhancer reprogramming by the oncogenic fusion protein EWS-FLI1. *Cell Rep.* 10, 1082–1095.
- Urbach, A., Yermalovich, A., Zhang, J., Spina, C.S., Zhu, H., Perez-Atayde, A.R., Shukrun, R., Charlton, J., Sebire, N., Mifsud, W., et al. (2014). Lin28 sustains early renal progenitors and induces Wilms tumor. *Genes Dev.* 28, 971–982.

Viswanathan, S.R., and Daley, G.Q. (2010). Lin28: a microRNA regulator with a macro role. *Cell* *140*, 445–449.

Viswanathan, S.R., Powers, J.T., Einhorn, W., Hoshida, Y., Ng, T.L., Toffanin, S., O'Sullivan, M., Lu, J., Phillips, L.A., Lockhart, V.L., et al. (2009). Lin28 promotes transformation and is associated with advanced human malignancies. *Nat. Genet.* *41*, 843–848.

von Levetzow, C., Jiang, X., Gweye, Y., von Levetzow, G., Hung, L., Cooper, A., Hsu, J.H., and Lawlor, E.R. (2011). Modeling initiation of Ewing sarcoma in human neural crest cells. *PLoS One* *6*, e19305.

Vormoor, B., and Curtin, N.J. (2014). Poly(ADP-ribose) polymerase inhibitors in Ewing sarcoma. *Curr. Opin. Oncol.* *26*, 428–433.

Wilbert, M.L., Huelga, S.C., Kapeli, K., Stark, T.J., Liang, T.Y., Chen, S.X., Yan, B.Y., Nathanson, J.L., Hutt, K.R., Lovci, M.T., et al. (2012). LIN28 binds messenger RNAs at GGAGA motifs and regulates splicing factor abundance. *Mol. Cell* *48*, 195–206.

Zhang, J., Ratanasirinrawoot, S., Chandrasekaran, S., Wu, Z., Ficarro, S.B., Yu, C., Ross, C.A., Cacchiarelli, D., Xia, Q., Seligson, M., et al. (2016). LIN28 Regulates Stem Cell Metabolism and Conversion to Primed Pluripotency. *Cell Stem Cell* *19*, 66–80.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal anti-EWS	Bethyl	Cat#A300-418A; RRID: AB_420958
Mouse monoclonal anti-Tubulin	Millipore	Cat#CP06; RRID: AB_2617116
Rabbit polyclonal anti-FLI1	Abcam	Cat#ab15289; RRID: AB_301825
Rabbit Anti-Human LIN28B Polyclonal Antibody	Proteintech	Cat# 16178-1-AP; RRID:AB_2135051
Rabbit Anti-Human LIN28B Polyclonal Antibody	Cell Signaling Technology	Cat# 4196; RRID:AB_2135047
Monoclonal Anti-GAPDH-Peroxidase antibody produced in mouse	Sigma-Aldrich	Cat# G9295; RRID:AB_1078992
Anti-Rabbit IgG (whole molecule)-Peroxidase antibody produced in goat	Sigma-Aldrich	Cat#A0545; RRID:AB_257896
Biological Samples		
EwS1-4	This paper	See Table S3
hpMSC1-2	This paper	See Method Details
Chemicals, Peptides, and Recombinant Proteins		
IMDM, GlutaMAX Supplement	Thermofisher Scientific	Cat#31980022
DMEM, high glucose, GlutaMAX Supplement, pyruvate	Thermofisher Scientific	Cat#31966021
KnockOut Serum Replacement - Multi-Species	Thermofisher Scientific	Cat#10828028
FBS Good Forte	PAN BIOTECH	Cat#P40-49500
Trypsin-EDTA (0.05%), phenol red	Thermofisher Scientific	Cat#25300062
MEM Non-Essential Amino Acids Solution	Thermofisher Scientific	Cat#11140035
Penicillin-Streptomycin (10,000 U/mL)	Thermofisher Scientific	Cat#15140122
EGF Human	PROSPEC Protein specialists	Cat#CYT-217
FGF 2 Human	PROSPEC Protein specialists	Cat#CYT-218
Fugene 6	Promega	Cat#E2692
Polybrene	Sigma-Aldrich	Cat#005557
Puromycin	InvivoGen	Cat#ant-pr-1
PowerUp SYBR Green Master Mix	Applied Biosystems	Cat#A25742
Calcein AM cell-permeable-dye	Life Technologies	Cat#C1430
Lin28 1632	Tocris Bioscience	Cat#6068
Dimethyl sulfoxide	Sigma-Aldrich	Cat#41640
Actinomycin D	Sigma-Aldrich	Cat#A1410
VectaMount® AQ Aqueous Mounting Medium	Vector Laboratories	Cat#H-5501
M-MLV Reverse Transcriptase	Promega	Cat#M170B
RNasin Ribonuclease inhibitor	Promega	Cat#N211B
RPMI 1640 Medium, GlutaMAX Supplement	Thermofisher Scientific	Cat#61870010
Lin28B RNAscope® Probe	ACD	Cat#596361
dNTP set	MP Biomedicals	Cat#11NTACG100-CF
Doxycycline hyclate	Sigma-Aldrich	Cat#D9891
Geneticin	Thermofisher Scientific	Cat#10131-027
FBS (tetracyclin-free)	PAN BIOTECH	Cat#P30-3602
Ketasol-100	Dr. E. Graeub AG	Cat#668.51
Rompun 2%	Provet AG	Cat#1315

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
PDGF-BB	PeproTech	Cat#100-14B
Critical Commercial Assays		
miRCURY RNA Isolation Kit - Cell & Plant	Exiqon	Cat#300110
Click-iT Nascent RNA Capture Kit	Life Technologies	Cat#10365
NucleoSpin miRNA kit	Macherey-Nagel	Cat#740971
EZ-Magna RIP™ RNA-Binding Protein Immunoprecipitation kit	MerckMillipore	Cat#PP64B
CellTiter 96® AQueous One Solution Cell Proliferation Assay	Promega	Cat#G3582
Deposited Data		
Gene expression data for Ewing sarcoma spheres generated for this project	This Paper	GEO: GSE122632
Gene expression data for primary Ewing sarcomas	Savola et al., 2011	GEO: GSE17618
Gene expression data for primary Ewing sarcomas	Brohl et al., 2014	https://pub.abcc.ncifcrf.gov/cgi-bin/JK
Gene sets for functional enrichment analysis	Broad Institute	Molecular Signatures Database, RRID:SCR_016863
ATARIS gene scores	Aguirre et al., 2016	https://depmap.org/portal/download/
Experimental Models: Cell Lines		
Lenti-X 293T	Clontech	Cat#632180
A-673	ATCC	Cat# CRL-1598, RRID:CVCL_0080
Experimental Models: Organisms/Strains		
NNOD.Cg-Prkdc ^{scid} Il2rg ^{tm1Wjl} /SzJ	The Jackson Laboratory	RRID:IMSR_JAX:005557
Oligonucleotides		
Random Primers	Promega	Cat#C118A
Primer sequences for real-time PCR	This paper	see Table S5
sgRNA targeting Lin28B (exon2): CACCGCATCGACTGGAATATCCAAG	Powers et al., 2016	N/A
Recombinant DNA		
shRNA Lin28B n.1	Broad Institute, RNAi Consortium	Cat#TRCN0000219859
shRNA Lin28B n.2	Broad Institute, RNAi Consortium	Cat#TRCN0000122191
shRNA targeting EWS-FLI-1	Tirode et al., 2007	N/A
pInd EWS-FLI1	Boulay et al., 2017	N/A
Software and Algorithms		
FlowJo version 9.9.4	FLOWJI, LLC	https://www.flowjo.com/solutions/flowjo/ , RRID:SCR_008520
Adobe Illustrator CC 2015	Adobe	https://www.adobe.com/products/illustrator.html , RRID:SCR_010279
GraphPad Prism version 8	GraphPad Software	https://www.graphpad.com/ , RRID:SCR_002798
Oligo	Carvalho and Irizarry, 2010	Bioconductor, RRID:SCR_006442
Limma	Ritchie et al., 2015	Bioconductor, RRID:SCR_006442
R Project for Statistical Computing	R Development Core Team, 2019	R Project for Statistical Computing, RRID:SCR_001905
RSEM	Li and Dewey, 2011	https://github.com/deweylab/RSEM
Other		
Ensembl	Aken et al., 2017	Ensembl Genome Browser, RRID:SCR_013367
TargetScan	Agarwal et al., 2015	TargetScan, RRID:SCR_010845

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Ivan Stamenkovic (Ivan.Stamenkovic@chuv.ch). All unique/stable reagents generated in this study are available from the Lead Contact with a completed Materials Transfer Agreement.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Primary tumors and cell lines cultures

Primary EwS and MSC samples from the University Hospital of Lausanne (CHUV) were obtained at surgery with the approval of the Ethics Committee of the Canton de Vaud (project authorization No. 131/12). For samples received from the Hospital Sant Joan de Déu (HSJD, Barcelona), written informed consent was obtained from all patients prior to inclusion of their samples in the tumor bio-bank, following procedures approved by the Ethical Committee for Clinical Research at HSJD (M. 1608-C). All human samples were anonymized before analysis and were exempted from informed consent in accordance with the law of the Canton de Vaud. [Table S3](#) lists LIN28B status, age, sex, tumor site, tumor staging, treatment status, EwS-FLI1 translocation details of EwS1-4 samples. Primary EwS spheres were cultured in IMDM (GIBCO), supplemented with 20% KO serum (GIBCO), 10ng/mL human recombinant EGF and bFGF (PROSPEC), and 1% Pen/Strep (GIBCO) in ultra-low attachment flasks (Corning), as previously described ([Suvà et al., 2009](#)).

HpMSCs were obtained from bone marrow of two healthy pediatric patients undergoing corrective surgery as described previously ([Riggi et al., 2008](#)). Both HpMSC1 and HhMSC2 were obtained from male patients. HpMSCs were cultured at low confluence in IMDM (GIBCO), 10% FCS (PAN BIOTECH), 10ng/mL PDGF-BB (PeproTech), and 1% Pen/Strep (GIBCO), and were tested for multi-lineage differentiation into adipocytes, chondrocytes, and osteoblasts ([Riggi et al., 2008](#)).

The Lenti-X 293T cell line was purchased from Clontech (cat no. 632180) and grown in DMEM medium (GIBCO) supplemented with 10% FCS (PAN BIOTECH) and 1% Pen/Strep (GIBCO). The A673 cell line was purchased from ATCC (cat no. CRL-1598) and grown in RPMI (GIBCO) supplemented with 10% FCS (PAN BIOTECH) and 1% Pen/Strep (GIBCO).

Cell cultures were maintained at 37°C and 5% CO₂ in humidified culture incubators.

Animal Studies

Experimental protocols involving mice were approved by the Veterinary Service of the Canton of Vaud, Switzerland (Etat de Vaud, Service Vétérinaire) under the authorization number VD2488. NOD-*scid* IL2Rgamma^{null} mice (NSG) purchased from The Jackson Laboratory, USA (stock number 005557) were used in this study. All mice were 6-8 weeks old males. Mice were maintained in a pathogen-free environment in individual ventilated cages and fed with autoclaved food and water at the animal Facility of the University of Lausanne.

METHOD DETAILS

Plasmids and Lentiviral Infection

Stable LIN28B depletion was obtained using either pLKO.1 lentiviral shRNA vectors purchased from the RNAi Consortium (shRNA n.1 ref. TRCN0000219859; shRNA n.2 ref. TRCN0000122191), or a previously described sgRNA targeting LIN28B exon 2 ([Powers et al., 2016](#)) cloned into the lentiCRISPRv2 vector. Control cells were infected with shRNA and sgRNA sequences targeting the *GFP* transcripts (GCAAGCTGACCCTGAAGTTCAT). The shRNA sequence used for EWS-FLI1 depletion was previously described and validated ([Tirode et al., 2007](#)). For LIN28B overexpression, the *LIN28B* cDNA was amplified from A673 cells and cloned into the pLIV lentiviral vector. Lentiviruses were produced using Lenti-X 293T packaging cells, transfected with the plasmid of interest, GAG/POL and VSV using FuGene 6 (Promega). In all cases, EwS cell suspensions from dissociated spheres were subject to lentiviral infection for 8hrs in the presence of 6μg/mL Polybrene (Sigma-Aldrich), and cells were selected with 1μg/mL puromycin (InvivoGen) for 48hrs prior to further analysis, as previously described ([De Vito et al., 2012](#)). HpMSCs infected with the plnd EWS-FLI1 V5 lentiviral vector ([Boulay et al., 2017](#)) were selected with 1500μg/mL geneticin (ThermoFisher Scientific) for 8 days and then treated with 1 μg/mL doxycycline (Sigma-Aldrich) for 8 supplementary days with medium renewal every other day.

RNA *In Situ* Hybridization, Immunohistochemistry and Western Blot Analyses

RNAscope technology (Advanced Cell Diagnostics, ACD) was used for RNA *in situ* hybridization following the manufacturer's instructions as described previously ([Filbin et al., 2018](#)). Briefly, slides were baked for 1hr at 60°C, deparaffinized and dehydrated. The tissue was pretreated with Hydrogen Peroxide for 10 minutes at room temperature and with Target Retrieval Reagent for 15 minutes at 98°C. Protease Plus was then applied for 30 minutes at 40°C. LIN28B probe (ACD) was hybridized for 2hrs at 40°C, followed by signal amplification. Tissue was counterstained with Gill's hematoxylin followed by mounting with VectaMount mounting media (Vector Laboratories).

For immunohistochemistry (IHC) EwS spheres were re-suspended in 1.5% low-melting agarose and included in paraffin blocks. The blocks were sectioned and stained using LIN28B-specific polyclonal antibody (Proteintech). The signal was revealed using goat anti-rabbit Ig conjugated to HRP (Sigma-Aldrich).

Western blots were analyzed according to standard procedures. The antibodies used were: rabbit polyclonal anti-human LIN28B antibody (Cell Signaling Technology), mouse monoclonal anti-GAPDH peroxidase conjugated antibody (Sigma-Aldrich), mouse monoclonal anti- α -tubulin antibody (Millipore), rabbit polyclonal anti-FLI1 antibody (Abcam), rabbit polyclonal anti-EWS antibody (Bethyl Laboratories).

RNA Isolation and Real-Time qPCR

Total RNA was isolated using miRCURY RNA Isolation Kit – Cell & Plant (Exiqon) according to the manufacturer's recommendations. Real-time qPCR was performed as previously described (Riggi et al., 2010) using PowerUp SYBR Green (Applied Biosystems). Primer sequences used for gene expression are listed in Table S5. Relative gene expression levels were calculated with the DDCT method after normalization of the Ct values to the geometric mean of the Ct values of three housekeeping genes (*GAPDH*, *36B4* and *TBP*). SYBR Green primers for miRNA immature forms were previously published (Jiang et al., 2005). For miRNA mature form expression quantification, RT-qPCR was performed as previously described (Cornaz-Buros et al., 2014).

Cell vitality, proliferation assays and drug tests

For cell viability assays, freshly dissociated spheres were stained with 1 μ g/mL Calcein AM cell-permeable-dye (ThermoFisher Scientific) and analyzed by FACS. For clonogenic assays Calcein-positive single cells were FACS-sorted onto ultra-low attachment 96-well plates (Corning) and cultured in complete KO medium for 15 days, when sphere formation was assessed. Cell proliferation was determined by MTS assays (CellTiter 96, Promega), according to the manufacturer's recommendations.

For pharmacological targeting of LIN28B, the 1632 chemical inhibitor (Tocris) was dissolved in DMSO (Sigma-Aldrich) according to the manufacturer's recommendations. EwS spheres were dissociated and re-suspended in 6-well plates (Corning) in complete medium containing either solvent (DMSO) or the 1632 inhibitor at day 0. The drug was added at a concentration of 50 or 250 μ M, and an equal volume of DMSO without the drug was used as a control. Media were aspirated on day 4 and replaced with fresh medium containing DMSO or the inhibitor.

For clonogenic assays, spheres were pre-treated with DMSO or the 1632 inhibitor for 4 days, then resuspended, stained with 1 μ g/mL Calcein AM cell-permeable-dye (ThermoFisher Scientific) and single Calcein-positive cells were FACS-sorted onto ultra-low attachment 96-well plates (Corning) containing medium with DMSO or the 1632 inhibitor. Media were removed on day 7 and replaced with fresh medium containing DMSO or the inhibitor. For *in vivo* tumorigenicity assays, EwS spheres were pre-treated in culture for 4 days with DMSO or the inhibitor prior to injection.

For the analysis of *EWS-FLI1* RNA stability, EwS2 cells were treated with 10 μ g/mL of ActD (Sigma-Aldrich), or an equal volume of solvent (DMSO) as control. Cells were infected with shCTRL or sh2 LIN28B lentiviral vectors. ActD or DMSO were added 36hrs later, and RNA was harvested at 0, 4, 8, 12 and 24hrs post-ActD treatment. Relative gene expression levels were calculated by real-time qPCR using DDCT method normalized to *GAPDH* expression levels.

Nascent RNA capture and cDNA synthesis

Nascent RNA capture was performed with a Click-iT Nascent RNA Capture Kit (Life Technologies) according to the manufacturer's instructions. Briefly, cells were incubated with 5-EU at a concentration of 0.5mM for 30 minutes. Total RNA was isolated using NucleoSpin miRNA kit (Macherey-Nagel) and biotinylated. Following RNA precipitation, Dynabeads provided in the kit were used to isolate the nascent RNA with incorporated 5-EU. For cDNA synthesis from the bound nascent RNA, the bead suspension was warmed at 70°C for 5 minutes. Random primers (Promega) and dNTP (MP Biomedicals) were immediately added to the mixture. The suspension was left to cool down to room temperature for 30 minutes under constant rotation. M-MLV Reverse Transcriptase (Promega) and RNasin Ribonuclease inhibitor (Promega) were subsequently added to the suspension, which was then warmed to 42°C for 1hr with gentle vortexing. Finally, the cDNA was collected after heating the solution for 5 minutes at 85°C.

In vivo tumorigenicity assays

For kidney subcapsular injections mice were anesthetized using 100mg/kg Ketamine (Ketasol-100, Graueb AG) and 16mg/kg Xylazine (Rompun 2%, Provet AG) dissolved in PBS and 1x10⁴ sphere-derived cells were injected beneath the renal capsule. Tumor growth was monitored weekly using ultrasound imaging (Vevo 2100 Ultrasound Device, 40-MHz probe, VisualSonics, Canada). Tumor volume was calculated by the following formula: $V = 4/3 \pi \times (Dd \times Ds \times Dt)/8$, where *Dd* corresponds to tumor height, and *Ds* and *Dt* to tumor lengths measured in long- and short-axis views, respectively. For the experiment in Figure 2, mice were sacrificed when tumors reached 1cm³. For the experiments in Figures 3 and 6, mice were sacrificed when control tumors reached 1cm³. For the experiment in Figure 5, mice were sacrificed 8 weeks after injection. Harvested tumors were further processed for RNA extraction, hematoxylin/eosin staining, and IHC.

RNA Immunoprecipitation

RNA immunoprecipitation (RIP) on total EwS1 and EwS2 cell lysates was conducted using the EZ-Magna RIPTM RNA-Binding Protein Immunoprecipitation kit (MerckMillipore), according to the manufacturer's instructions. Rabbit polyclonal anti-human LIN28B antibody (Cell Signaling Technology) was diluted 1:50, as recommended by the manufacturer. Briefly, following gentle lysis of the cells at –80°C, total lysates were incubated overnight with anti-LIN28B antibody pre-bound to magnetic beads provided with the kit. The

beads were then rinsed and the antibody-protein-RNA complexes were dissociated by proteinase K digestion for 30 minutes. RNA was purified using the miRCURY RNA Isolation Kit - Cell & Plant (Exiqon) and enrichment of the selected LIN28B-bound transcripts was assessed by RT-qPCR. Ct values obtained for the genes of interest were normalized to Ct values obtained from 10% of the input (corresponding to the non-immunoprecipitated RNA obtained from the same sample).

Achilles scores and survival analysis

ATARIS scores were obtained from the project Achilles CRISPR screening data (Aguirre et al., 2016). The median ATARIS score for each gene was used to select the top 100 genes associated with the strongest reduction in EwS cells growth. To determine the correlation between the expression of these genes and survival in Ewing sarcoma we used the results of (Savola et al., 2011), available as GSE17618. Cox univariate analysis in the R computing environment (R Development Core Team, 2019, <https://www.R-project.org>) was used to determine the correlation between the expression of each gene and event-free survival.

LIN28B expression in an Ewing sarcoma dataset

We retrieved the gene-level normalized data of 65 primary Ewing tumors (Brohl et al., 2014) from Oncogenomicsdb and represented expression as $\log_2(\text{RPKM}+1)$. Microarray data were retrieved as normalized expression levels from the respective GEO records (GSE17618; GSE34620; GSE12102). As the detection threshold, we used the median expression over all genes and samples. Correlation between gene expression and survival was evaluated by Cox univariate analysis on logarithmic expression values using event-free survival times. To assess possible associations between LIN28B status and somatic mutations in known oncogenes, we used Fisher exact test to determine whether tumors expressing LIN28B were significantly overrepresented among those carrying somatic mutations in *STAG2*, *CKDN2A*, *TP53*, or *BRCA2*, as reported in Table S5 of Brohl et al. (2014).

Analysis of RNA-seq data

Gene-level counts were obtained by aligning reads to the human transcriptome (Ensembl version 79 (Aken et al., 2017)) using RSEM (Li and Dewey, 2011). When comparing LIN28B-positive versus negative spheres we considered as differentially expressed the genes with absolute \log_2 fold-change greater than one in all four possible comparisons. For the shRNA and 1632 compound experiments, we considered as differentially expressed between two conditions the genes with $\log_2(\text{TPM}+1)$ greater than 3 in at least one of the conditions and absolute \log_2 fold-change between the two conditions greater than one. The condition $\log_2(\text{TPM}+1)$ in at least one of our samples was also used to select the genes used in analyzing the overlap between our differentially expressed genes and gene lists derived from external databases.

MicroRNA expression analysis

We used the Affymetrix miRNA 4.0 platform to compare miRNA expression between two LIN28B⁺ and two LIN28B⁻ tumors. RMA normalization and background subtraction were performed using the “oligo” Bioconductor package (Carvalho and Irizarry, 2010). Differential expression was evaluated using the “limma” Bioconductor package (Ritchie et al., 2015) and miRNAs with absolute \log_2 of fold change greater than 1 and nominal $p < 0.01$ were considered differentially expressed. Their expression z-score are shown in the heatmap of Figure 5A, together with the p value of the enrichment of their targets among genes that were found differentially expressed in LIN28B-positive versus negative spheres. Specifically, we evaluated, by exact Fisher test, the overlap between predicted targets of up- (down-) regulated miRNAs (obtained from TargetScan [Agarwal et al., 2015] and genes that are down- (up-) regulated in LIN28B⁺ spheres. The size of the overlap and the corresponding p values are shown beside the heatmap.

Overlap between differentially expressed genes and gene sets

The overlap between lists of differentially expressed genes and gene sets shown in the Venn diagrams was evaluated for statistical significance using exact Fisher test. The “chemical and genetic perturbations” gene sets of the Broad Institute MSIGDB database were obtained from their website. The heatmaps represent the expression z-score of the common genes in the conditions indicated.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis of wet lab experiments (Student's t test, 2-way ANOVA, Kaplan-Meier test, Fisher's exact test) were performed by Prism GraphPad Software 8.00. All statistical tests and sample numbers are disclosed in respective Figure Legends/Supplementary Tables.

Statistical analysis of bioinformatics data are described in details in the “Method Details” section.

DATA AND CODE AVAILABILITY

Gene expression data generated for this project are available for the Gene Expression Omnibus repository under accession GSE122632 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE122632>). All the software used in the analysis and the data generated for other publications are publicly available as detailed in the Key Resources Table.

Cell Reports, Volume 30

Supplemental Information

LIN28B Underlies the Pathogenesis

of a Subclass of Ewing Sarcoma

Tugba Keskin, Arnaud Bakaric, Patricia Waszyk, Gaylor Boulay, Matteo Torsello, Sandrine Cornaz-Buros, Nadja Chevalier, Thibaud Geiser, Patricia Martin, Angela Volorio, Sowmya Iyer, Anupriya Kulkarni, Igor Letovanec, Stéphane Cherix, Gregory M. Cote, Edwin Choy, Antonia Digkila, Michael Montemurro, Ivan Chebib, Petur G. Nielsen, Angel M. Carcaboso, Jaume Mora, Raffaele Renella, Mario L. Suvà, Carlo Fusco, Paolo Provero, Miguel N. Rivera, Nicolò Riggi, and Ivan Stamenkovic

Supplemental Information

Content:

Five Supplemental Figures

Five Supplemental Tables

Supplemental Figure 1

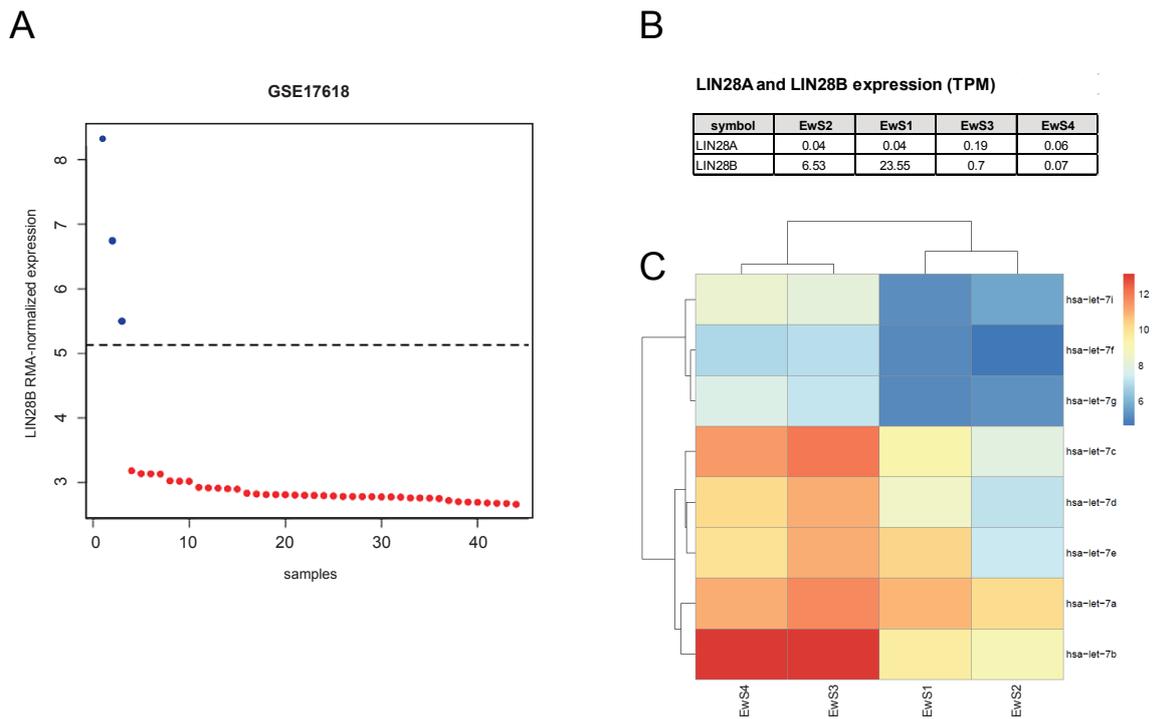


Figure S1 (refers to Figures 1 and 2)

(A) Box plot depicting *LIN28B* transcript expression in a cohort of 44 EwS samples (Savola et al., 2011; GSE17618). The horizontal dashed line represents the overall median gene expression value. (B and C) Expression of LIN28A and LIN28B (B) and heatmap of let-7 isoform expression (C) in the four primary EwS cell cultures.

Supplemental Figure 2

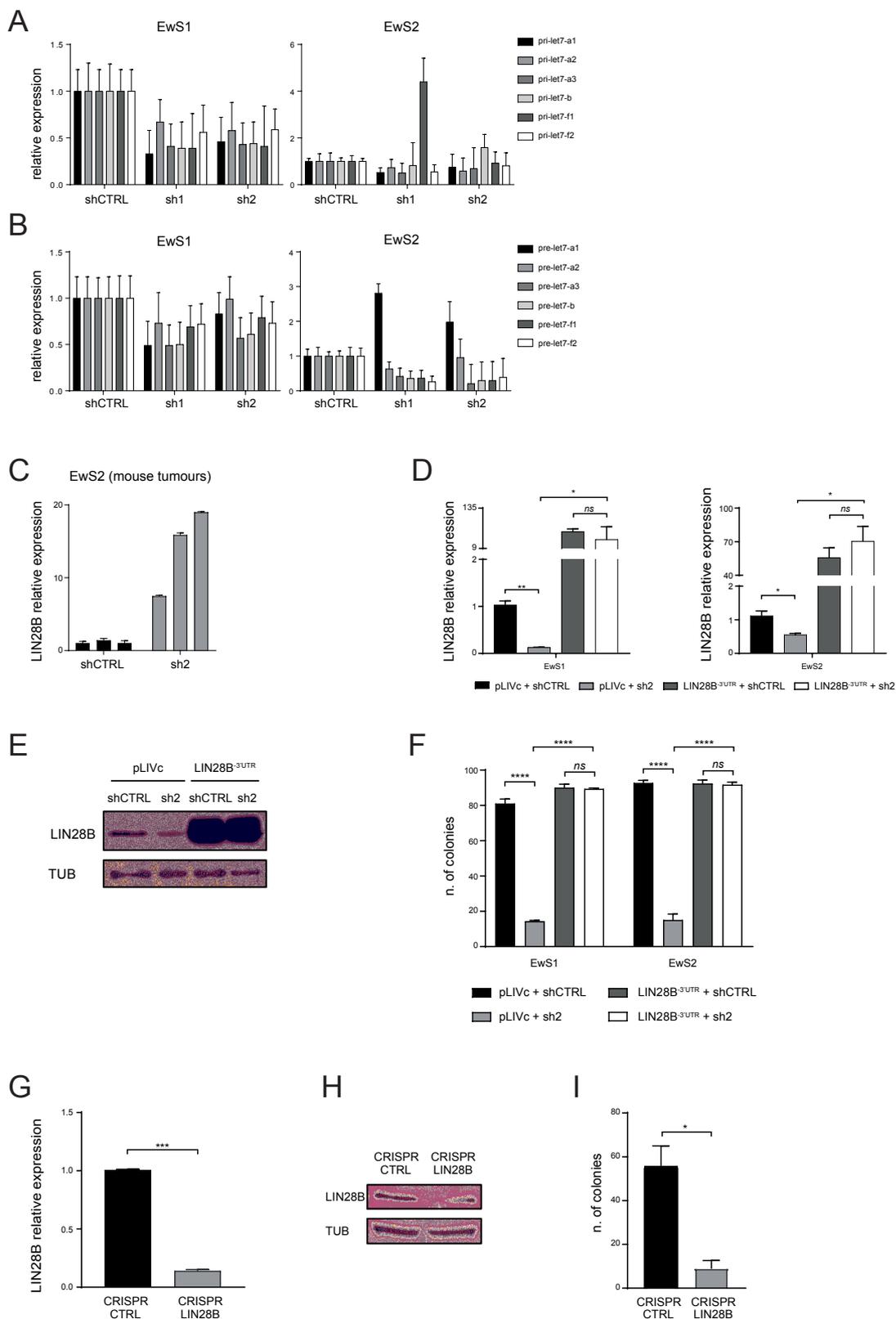


Figure S2 (refers to Figure 3)

(A and B) RT-qPCR analysis of let-7 microRNA precursor forms (pri-let7 and pre-let7) in control or LIN28B-depleted EwS1 and EwS2 cells (mean \pm SEM). Let-7 depletion by targeting shRNAs is relative to let-7 expression in the same cells infected with vectors bearing non-targeting shRNAs. **(C)** RT-qPCR analysis of *LIN28B* expression in three tumor xenografts derived from either control or *LIN28B*-depleted

EwS2 cells (mean \pm SEM). **(D)** RT-qPCR and **(E)** Western Blot analyses of LIN28B expression in EwS1 (*left*) and EwS2 (*right*) cells transduced with the indicated combinations of lentiviral constructs. pLIVc: empty vector; *LIN28B-3'UTR*: lentiviral vector containing *LIN28B* cDNA lacking the 3'UTR (mean \pm SEM). Expression of LIN28B is compared to cells infected with an empty vector (pLIVc) and a *GFP*-targeting shRNA. The data represent the mean values of three independent experiments. **(F)** Clonogenic assay of *LIN28B*-depleted and *LIN28B-3'UTR*-rescued EwS1 (*left*) and EwS2 (*right*) cells, as in **(D)**, showing complete restoration of spherogenic ability upon introduction of *LIN28B-3'UTR* (mean \pm SD). **(G)** RT-qPCR and **(H)** Western Blot assessment of LIN28B expression in EwS1 spheres transduced with a CRISPR/Cas-9 lentiviral vector targeting either *GFP* or *LIN28B* sequences (CRISPR CTRL and *LIN28B*, respectively; mean \pm SEM). **(I)** Clonogenic assay of CRISPR CTRL and *LIN28B* EwS1 cells showing that CRISPR-mediated depletion of LIN28B leads to a decrease in spherogenic ability comparable to that caused by shRNA (**Figure 3E**, mean \pm SD). Statistical analysis was performed with two-way ANOVA. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$; *ns*, non-significant.

Supplemental Figure 3

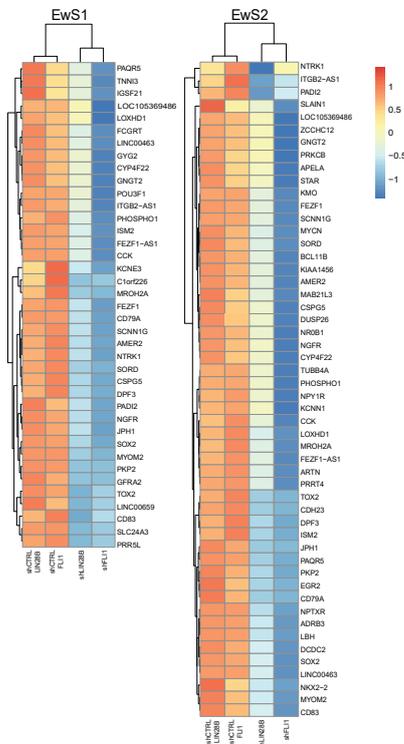
A

set	P value	common
PHONG_TNF_RESPONSE_NOT_VIA_P38	2.208 ⁻¹⁹	65
ELVIDGE_HYPOXIA_UP	2.551 ⁻¹⁷	41
MARTENS_TRETINOIN_RESPONSE_UP	1.694 ⁻¹⁶	66
ELVIDGE_HYPOXIA_BY_DMOG_UP	1.819 ⁻¹⁶	35
ONDER_CDHL_TARGETS_2_DN	2.417 ⁻¹⁶	59
RIGGI_EWING_SARCOMA_PROGENTOR_UP	4.819 ⁻¹⁶	66
...
MIYAGAWA_TARGETS_OF_EWSR1_ETS_FUSION_UP	5.88 ⁻¹⁰	42
ZHANG_TARGETS_OF_EWSR1_FL11_FUSION	1.65 ⁻⁰⁹	23
KINSEY_TARGETS_OF_EWSR1_FL11_FUSION_UP	0.098701	85

B

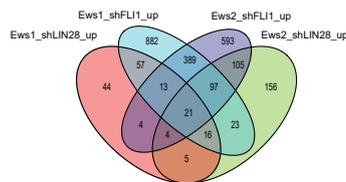
EwS1							
Set 1	Set 2	N_1	N_2	N_Background	Common	Expected	P_value
Sh-LIN28B	Sh-FLI1	829	1317	13256	255	82.36	8.61 ⁻⁶⁸
Sh-LIN28B	EWS-FLI1 target	829	99	13256	43	6.191	4.582 ⁻²⁶
Sh-FLI1	EWS-FLI1 target	1317	99	13256	56	9.836	7.447 ⁻³¹

EwS2							
Set 1	Set 2	N_1	N_2	N_Background	Common	Expected	P_value
Sh-LIN28B	Sh-FLI1	568	1281	13256	320	54.89	1.355 ⁻¹⁸³
Sh-LIN28B	EWS-FLI1 target	568	99	13256	56	4.242	6.785 ⁻²⁶
Sh-FLI1	EWS-FLI1 target	1281	99	13256	82	9.567	6.104 ⁻⁶⁷

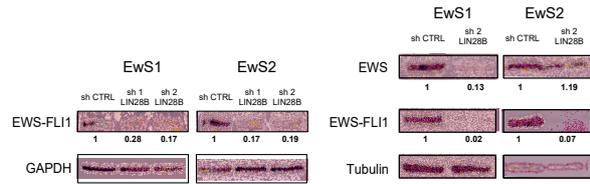


C

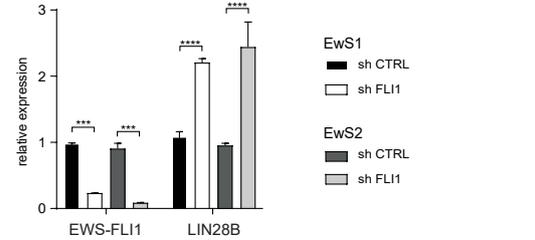
set_1	set_2	n_1	n_2	n_background	common	expected	p_value
Ews1_shLIN28B_up	Ews1_shFLI1_up	164	1498	13256	107	18.53	1.810E-61
Ews1_shLIN28B_up	Ews2_shFLI1_up	164	1226	13256	42	15.17	6.512E-10
Ews2_shLIN28B_up	Ews1_shFLI1_up	427	1498	13256	157	48.23	8.351E-45
Ews2_shLIN28B_up	Ews2_shFLI1_up	427	1226	13256	227	39.49	1.759E-124



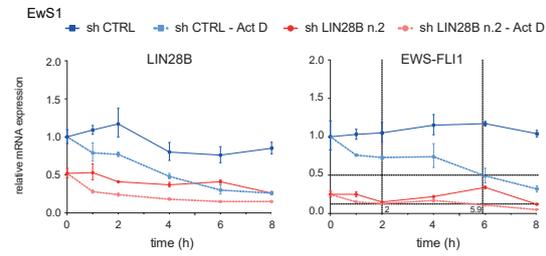
D



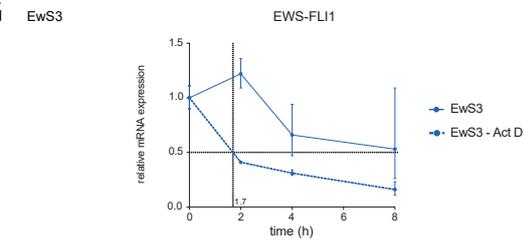
E



F



G



H

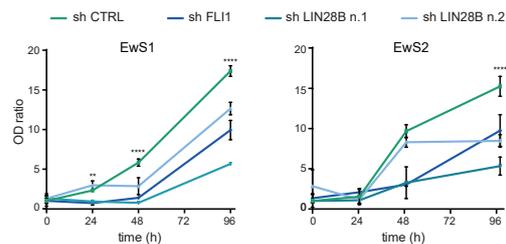


Figure S3 (refers to Figures 5 and 6)

(A) Functional GSEA pathway analysis of LIN28B-depleted EwS1 cells. (B) Upper panel: Statistical comparison of RNA-seq data between EwS1/2 cells depleted of either LIN28B (shLIN28B) or EWS-FLI1 (shFLI1), and a manually curated list of 99 direct EWS-FLI1 target genes. Lower panel: Heatmaps depicting the common downregulated genes in LIN28B and EWS-FLI1-depleted EwS1 or EwS2 cells. (C) Venn

diagram illustrating the overlap between genes down regulated by LIN28B and EWS-FLI1. **(D)** Western Blot analysis of EWS-FLI1 (*left panel*) and both EWS-FLI1 and EWS (*right panel*) expression levels in LIN28B-depleted EwS1 and EwS2 cells. The numerical values beneath the panels indicate the decrease in expression level normalized to GAPDH and tubulin loading controls. **(E)** RT-qPCR analysis of *LIN28B* expression in EWS-FLI1-depleted EwS1 or EwS2 cells. Two-way ANOVA was used for statistical analysis. Mean values \pm SEM of three independent experiments are shown. **(F)** Analysis of transcript levels upon ActD treatment in EwS1 cells. The *LIN28B* (*left panel*) and *EWS-FLI1* (*right panel*) decay rate was measured by RT-qPCR at five time points. The *EWS-FLI1* transcript half-life is shown to shift from 5.9hrs in the presence of LIN28B to 2.0hrs upon its depletion. **(G)** Analysis of *EWS-FLI1* transcript half-life in wild-type EwS3 cells (measured as in **F**) revealed a half-life of approximately 1.7 hours. **(H)** Proliferation of EwS1 and 2 spheres transduced with either control, *LIN28B*- or *EWS-FLI1*-targeting shRNAs measured by MTS. Values were normalized to timepoint 0 (mean \pm SD). Two-way ANOVA was used for statistical analysis. ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$.

Supplemental Figure 4

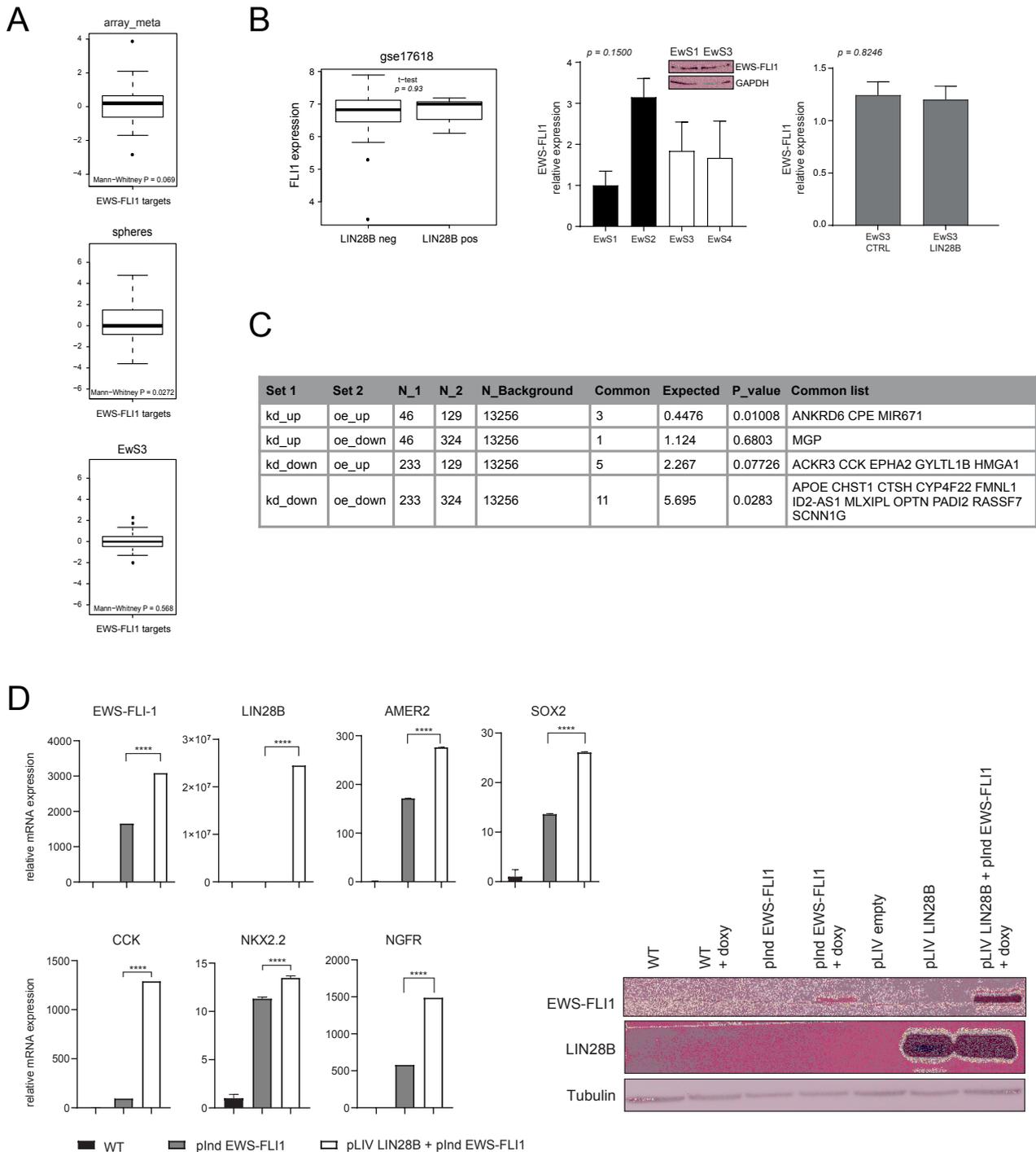


Figure S4 (refers to Figures 5 and 6)

(A) Upper panel: Boxplot comparing the log₂ fold change of 99 direct EWS-FLI1 target genes in LIN28B+ and LIN28B- tumor samples. The fold changes are averaged over 3 microarray datasets (GSE12102; GSE17618; GSE34620). When multiple microarray probes were associated to a gene, we selected, for each dataset, the one with the highest mean expression. **Middle panel:** Boxplot comparing the log₂ fold change of 99 direct EWS-FLI1 target genes in LIN28B+ and LIN28B- spheres. The fold changes refer to the mean expression in the two LIN28B+ samples vs the mean expression in the two LIN28B- samples. **Lower panel:** Boxplot comparing the log₂ fold change of 99 direct EWS-FLI1 targets in EwS3 cells expressing LIN28B-3'UTR and control EwS3 cells. **(B) Left panel:** The boxplots compare the logarithmic normalized expression of FLI1 in the respective dataset in LIN28B+ vs LIN28B- samples. Samples were defined as

LIN28B+ if *LIN28B* expression was higher than the median expression of all genes over all samples. When multiple microarray probes were associated to a gene, we selected the one with the highest mean expression. *Middle panel*: RT-qPCR and western Blot analysis of EWS-FLI1 expression in EwS 1-4 cells. *Right panel*: *EWS-FLI1* expression in control and *LIN28B-3'UTR* expressing EwS3 cells. Two-way ANOVA (*middle panel*) and Student's t-test (*lower panel*) were used for statistical analysis. **(C)** Comparison of differentially expressed genes shared by EwS1 and EwS2 cells depleted of LIN28B (kd) and EwS3 cells overexpressing LIN28B (oe). There is no significant overlap between the genes that move in opposite directions in the two data sets. Significant overlap is observed only for a small number of genes that move in the same direction upon LIN28B overexpression in EwS3 and depletion in EwS1/EwS2 cells. **(D)** *Left panel*: Expression of *EWS-FLI1*, *LIN28B* and a panel of known direct EWS-FLI1 target genes (*AMER2*, *CCK*, *NKX2.2*, *SOX2* and *NGFR*) in hpMSC2, infected with pInd EWS-FLI1 or co-infected with pInd EWS-FLI1 and pLIV LIN28B respectively (mean values \pm SD are shown). EWS-FLI1 was induced by doxycycline treatment for 8 days. Student's t-test was used for statistical analysis. *Right panel*: Western Blot analysis of EWS-FLI1 and LIN28B expression in hpMSC2 infected with the indicated lentiviral vectors and cultured with (+ doxy) or without doxycycline for 8 days. * $p < 0.05$; **** $p < 0.0001$.

Supplemental Figure 5

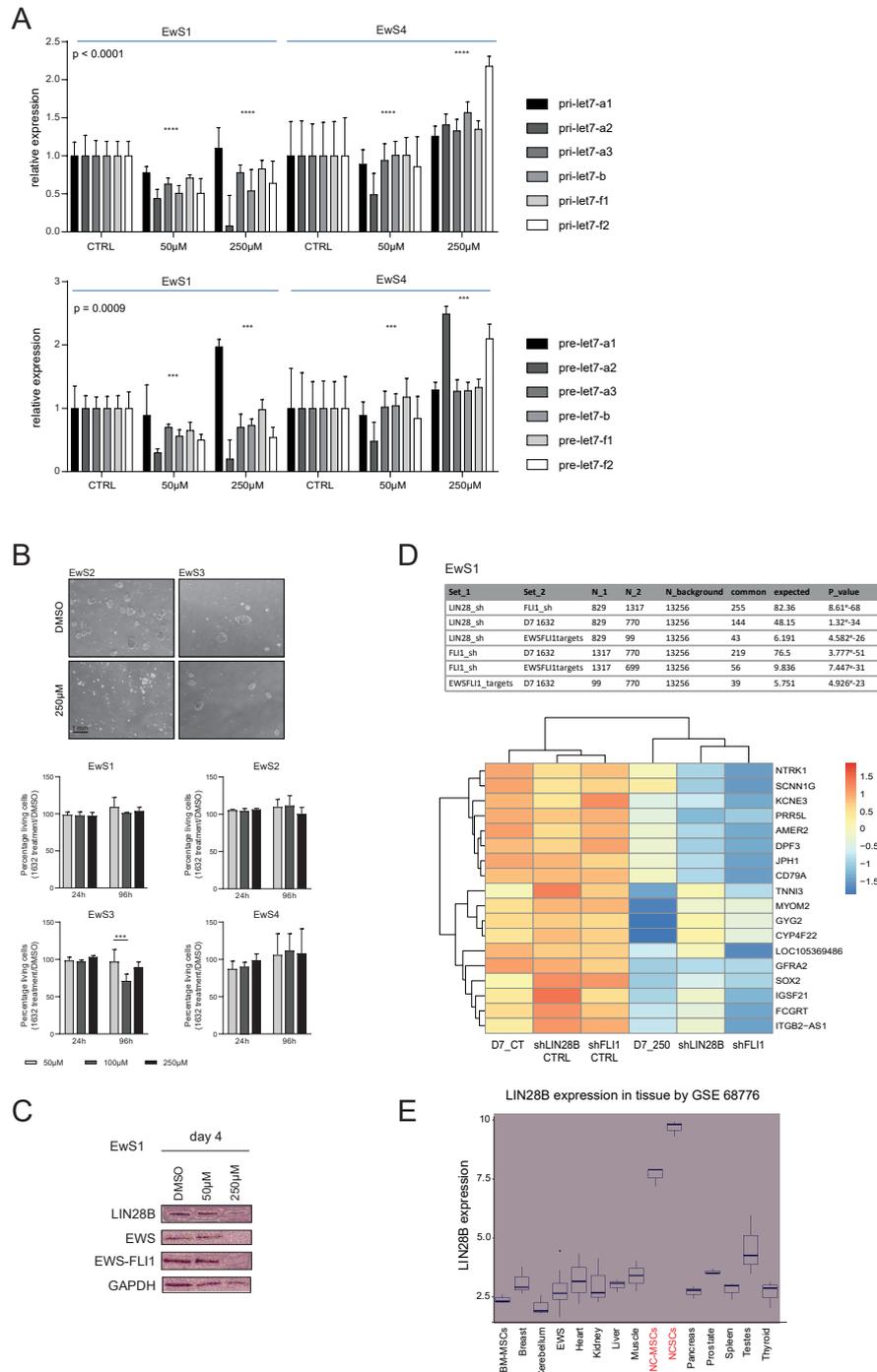


Figure S5 (refers to Figure 7)

(A) RT-qPCR analysis of let-7 miRNA precursor forms (pri-let7 and pre-let7) in EwS1 and EwS4 cells treated either with DMSO (CTRL), or increasing concentrations of the 1632 inhibitor (mean \pm SEM). (B) *Upper panel*: micrographs depicting EwS2 and EwS3 sphere cultures, after 96hrs of treatment with either DMSO or 250 μ M of the LIN28B inhibitor 1632. *Lower panel*: viability of the 4 primary EwS cultures following 96hrs of treatment with the inhibitor at the indicated concentrations. (C) Western Blot analysis of LIN28B, EWS and EWS-FLI1 expression in EwS1 cells after 4 days of treatment with the 1632 inhibitor. (D) Statistical comparison of RNA-seq data between EwS1 cells depleted of LIN28B (shLIN28B), EWS-FLI1 (shFLI1) or treated with the 1632 inhibitor (D7 1632), and a manually curated list of 99 direct EWS-FLI1 targets genes. *Lower panel*: Heatmap depicting the common downregulated genes in LIN28B or EWS-FLI1-depleted and inhibitor-treated EwS1 cells. (E) LIN28B expression in normal tissues and cells. BM-MSCs: bone marrow mesenchymal stem cells; NC-MSCs: neural crest derived mesenchymal stem cells; NCSCs: neural crest stem cells. *** $p < 0.001$; **** $p < 0.0001$.

Supplemental Table 1

gene	A673_BONE	CADOES1_BONE	EWS8_BONE	EWS502_BONE	MHHES1_BONE	RDES_BONE	SKES1_BONE	TC32_BONE	TC71_BONE
FLI1	-1.7155747	-0.005276916	-2.19884966	-3.546416131	-2.075481018	-2.98938317	-2.251644225	-1.399897726	-1.6272414
ETV6	-2.67701418	-0.874143404	-1.9861892	-1.488600508	-1.770666689	-2.27100231	-1.195739071	-0.57521201	-1.7164513
CDK4	-1.36415395	-1.444025616	-1.00994362	-2.424460582	-1.659558954	-1.95001239	-1.461991612	-0.78796882	-0.4577825
CITED2	-1.43081782	-2.445382734	-1.07566435	-1.757784496	-0.80499902	-1.07156063	-0.887528474	-1.74205236	-1.2808429
STAG1	-0.00944325	-0.232443173	-1.67041687	-2.088117241	-1.162394246	-1.32056758	-1.381279711	-0.73354232	-0.4486620
C15orf41	-1.37182045	-0.570835489	-1.23742018	-2.11962023	0.065848758	0.179156087	-1.175648982	-0.9003163	-1.1613294
GPX4	-0.9373859	-1.483892875	-0.89503612	-2.198552657	-0.472156712	-1.93474208	-1.451348598	-0.05884521	-1.1427977
TTC7B	-1.27731268	-0.280817702	-1.06977738	-1.04935969	-0.60273645	-0.79391264	-1.623967329	0.163792402	-1.0416359
GS2G	-0.09663661	-0.047591905	-1.1016923	-2.102151665	-1.129995156	-2.13687387	-0.983034428	-0.24949218	0.35815253
RNF216	-0.66180625	-0.408781487	-1.01283579	-0.44037517	-0.943013486	-0.84105355	-1.112733776	-1.0552749	-1.1790997
EP300	-0.15035004	-0.814022004	-0.93533882	-1.190407467	-0.916935733	-1.43336124	-1.109543315	-1.05898419	-0.6082613
BCL11B	-0.75409822	-1.335171142	-1.3280633	-0.406406063	-1.225330955	-0.48098249	-1.541453461	-0.90595819	-0.4774327
ATP2C1	-0.22535265	0.118447426	-1.68892752	-1.542413538	-1.030175336	0.719391265	-0.028368974	-0.94078428	-0.8997896
TRIM8	-0.74690158	-0.712969341	-1.33653805	-0.83789118	-0.874959154	-1.49320845	-0.955022269	-0.26866084	-1.2092549
MEITL11B	-0.89317429	-0.867721014	-0.52431885	-1.074406859	0.487171121	-0.69162906	-0.886855777	-0.87175566	-0.218194
SOX1	-0.94675988	-0.86099667	-1.4023891	0.103057027	-0.426469612	-0.97683437	-1.207057675	-1.43727426	-0.1174114
IGF2BP1	-0.78751838	-0.944919399	-0.85689458	-0.714365211	-0.64607726	-1.15736585	-0.563041168	-0.86742969	-0.8925977
BCL2	-0.54811673	-2.698268976	0.14999185	0.295988904	0.154819946	-1.45127578	-2.589690848	-0.85822602	-1.7103273
G6PD	-0.03115683	-1.065652655	0.01709692	-3.00918875	-1.518983881	-1.49397807	-0.288712929	-0.8055086	-0.7372615
PAG1	0.381954409	-1.790576577	0.77311069	-0.1978261	-0.874701735	-1.17388469	-0.835409385	-0.8225488	-0.2121864
CDAN1	-0.82137145	-1.447851792	-0.78496606	-1.885251745	0.25135121	0.28795309	-1.889023681	-0.66044015	-0.6642811
PI4KB	-0.91312798	0.08409012	-0.80471277	-0.04197277	-0.433021938	-1.29891637	-0.768913755	-0.8218728	-0.2521514
DROSHA	-0.43164771	-0.410113404	0.41857646	-0.786060907	-1.022481557	-0.80124889	-0.754753068	-0.88787275	0.05359999
WDR59	-0.80421219	-0.280742899	0.07117579	-1.712098852	-0.855543758	-1.1585276	-0.608987805	-0.75133971	0.02856956
CHML	-0.17297085	-0.987205374	-1.01805773	0.524371022	0.133516488	-0.73992828	-0.993181968	-1.69355716	-0.0842567
USP14	-1.43769896	-0.026455181	-2.69950404	-0.73432158	0.059190553	-1.580100627	0.18794755	-0.133979221	-2.2949959
EEFSEC	-1.44519948	-0.729682787	-0.78377124	-1.599342563	-0.686763461	-1.94302841	-0.40191436	-0.0558256	-0.1377755
ZFHX4	0.015236565	-0.339376752	0.40039742	-0.878512199	-0.761381779	-0.85938781	-0.727330677	-1.0491658	-0.3482241
HS3ST2	-0.82087113	-0.700602763	-0.18469986	-0.870674038	-0.537219252	-0.53440858	-0.872639907	-0.84781823	-1.0391282
CA6	0.668062555	-0.791193824	0.43321276	-0.029128803	-1.133300183	-0.70028019	-1.032493962	-0.73863657	0.12267635
DNAJC19	-0.70017274	0.000636276	-0.74948567	-1.506097463	-1.255909962	0.17758216	-1.104531327	-0.259948742	0.7054356
FAM167A	-0.85668201	-0.179739271	0.36623967	-0.696562359	0.097461567	-0.70629681	-0.552484343	-0.76871214	-1.0956378
CENPU	-0.7552295	0.023825539	0.61925949	-1.211257526	-1.028199917	-0.6954606	0.548904768	-0.69226304	0.27570615
WDR43	-1.20322937	-0.788651444	-0.45419897	0.311231174	-0.352768115	-0.82493601	-0.687695553	-0.73694006	0.63530771
DIAPH3	-1.38673481	-0.190694598	-0.91653523	-0.028560751	-1.375761222	-0.29368567	-0.822647458	-0.39488409	-0.6618471
CDH11	-0.67092605	0.342145189	-1.94378538	-0.65972319	-0.592760267	0.82114928	-0.127055283	-0.8005459	-0.6789128
SLC25A33	0.360111659	-0.768619669	-1.12093875	-0.140418254	-0.664132916	-0.65602414	-0.7031274	0.023705184	-0.62574019
CLDN15	0.170943701	-0.923663303	0.44314773	-0.390273762	-0.768041016	-0.64826308	-0.73424046	-0.78578258	0.63960768
KIF17	-0.9184185	-0.26441843	-0.64493191	-0.748362534	0.05649786	-0.66422912	-0.617986512	-0.64646968	-1.2091092
SMEK1	-1.23349036	0.083066765	-1.06169649	-0.772111546	-0.669285528	0.027651659	-0.130114651	0.380169755	-0.6445153
KBTBD2	-0.64175548	-0.283480113	-0.25410972	-0.125536954	-0.043304738	-1.96514006	-0.95341823	-0.86182872	-1.5973673
MAGOH	-1.44490773	-0.104705699	-1.04936393	-0.707844107	0.25580894	0.051954315	-0.522498398	-0.6314724	-0.6409029
LRG2	-0.09358316	-0.375230456	-0.77264693	-0.628095056	0.448841682	-0.84165483	-0.675317932	0.020286775	-0.6846163
BCL2	-0.18935689	-1.626396452	0.49541672	-0.057931277	-0.62660832	-0.83578115	-0.623858533	-0.92541929	-0.7215172
MIR6821	0.158873483	-0.786810849	-0.62092658	-0.17292043	-1.321849789	-1.02283955	0.457126967	-0.65147458	0.27082362
SLC38A2	-0.42486122	-0.652552763	-0.79422206	-0.759915186	0.147614397	-0.80813413	-0.455021088	-0.61874579	-0.2311168
HOXD11	-0.71250874	-0.610616574	-0.10210502	-0.80509877	0.217334617	-1.04168605	-0.701281583	-0.527468	-0.5846563
SETDB1	-0.62095594	-0.80839779	-0.60036195	-1.789260163	0.201395654	-0.73348014	-0.124866304	-1.03899885	0.3447292
CLMP	-0.60648972	-0.427981428	-0.6571859	-0.201965553	-0.123278847	-1.19146193	-0.663388346	-1.2226093	-0.5015005
MPV17L2	-0.60614378	0.105170122	-0.81837086	-0.612047911	-0.654388305	-0.01423151	-0.60051282	0.347041319	-1.3174403
VPS26B	-0.62136022	-0.608438277	-0.02157988	0.412123672	-0.789479853	-0.6035864	-0.491466386	-0.82365498	-0.2247578
TCTE3	-0.60768922	-0.609646716	0.47004169	-0.123675408	-0.106458432	-0.60278347	0.030886688	-0.65404112	-0.7746337
FZD8	-0.43196967	-0.73343319	-0.60124881	-0.227182988	-0.257708416	-0.0726287	-0.616357221	-1.061581	-0.846281
SIPA1L2	-0.60213238	-0.020117508	0.31726472	-0.109175205	-0.835170445	-0.62163588	-0.692826048	-0.56308464	-0.6010934
SYBU	-0.0636752	-0.600610495	0.0155393	-0.529330196	0.00711446	-0.6666384	-0.813158586	-0.76308048	-0.733032
ZEB2	-0.86105403	-0.760649831	0.09798716	-0.599510474	-0.045988914	-0.80836661	-0.866242575	-0.51719804	-0.4371078
IGF1R	0.044922443	0.08961898	0.06310936	-0.598494629	0.115393506	-2.86107978	-0.97638994	-0.83216803	-1.0737047
MMD	-0.61150691	-0.597195927	0.12388469	0.124163533	-0.619490208	-0.32607078	-0.635802225	-0.282334936	-0.5133662
RGN	-0.52772235	-0.985850384	-1.52587827	-0.595635511	0.009095543	-0.23414507	-0.618082297	-0.44733818	-0.845902
NR1H3	-0.7537608	-0.083495836	-0.59369844	-0.835147377	-0.232903679	-0.80216357	-0.474585817	0.063267617	-0.6290334
SLC2A5	-0.30299139	-0.590216435	-0.16447504	0.04295468	-1.095406632	-0.94347672	-0.783510692	-0.63051734	-0.11873384
LIN28B	-1.10492807	0.054174751	-0.07132055	-0.69783342	0.196395306	-0.8535356	-0.977065072	-0.08843985	-0.5860031
HDAC8	0.321833391	-0.292201153	-0.88046476	-0.585185669	-0.740072656	-0.35870268	-0.690451934	-0.83889098	-0.4015946
IFI44L	-0.5848162	-0.147916095	-0.89477384	0.146640487	-0.715395165	-0.72820788	0.257524108	-0.17759704	-0.6233263
WSB2	0.449019727	-0.682494165	-0.78519598	0.084552537	0.12850296	-1.6152846	-0.796497555	0	-0.5823799
CDSN	0.029540461	-0.685273719	-0.61542097	-0.625470849	-0.574591176	-0.05121159	0.513165244	0.474518564	-1.1528232
HBO1	-0.51141706	-0.574333167	-0.7635567	-0.64222798	-0.399771822	-0.05689955	0.113372833	-0.78556985	-1.1925935
FEN1	-0.8016389	0.16542149	-0.74540782	0.056435973	-1.631914334	0.523202974	-0.547232046	-0.57362668	-0.6076718
LPAR4	-0.84122915	-0.224324122	-0.57214362	-0.585557267	-0.826114885	-0.17678148	-0.962645026	-0.31952554	-0.560628
SEC14L1	-0.92461916	-0.680414129	-0.20152906	-0.550102242	-0.147593286	-0.25022803	-0.790385675	-0.57160444	-0.6479306
IRS2	-0.57012447	-0.395970461	-0.50272359	0.459231375	0.483979768	-2.11552319	-1.131670011	-0.67372003	-1.6112626
HS3ST4	-0.38284909	-0.598916	-1.02087937	-0.558253509	-0.124351966	-0.30687189	-0.689646959	-0.56951558	-0.6890174
MCL1	-1.06200808	-0.566024334	0.17810104	0.106310779	-0.829790799	-1.16379852	-0.389025773	-0.7887672	0.01167893
PI3M3	-0.32617897	-0.998876734	-0.23574436	-0.561670108	-0.699977332	-0.9876032	-0.540799006	-1.16961215	-0.4139014
COQ10A	0.089196736	-0.360043907	-0.60298449	-0.747211587	0.309614373	-0.60260481	-0.148454121	-0.86181726	-0.5614994
HOXC9	-0.05318342	-0.328283158	-0.26937379	-0.595905058	-0.561455204	-0.7479862	-0.420682253	-1.22005371	-0.564927
SP9	-0.90037518	0.398616298	-0.71679138	0.056219993	-0.561287165	0.353676105	-0.872634642	0.091574936	-0.9099277
ADAMTS4	-0.44978619	-0.721147859	-0.00482881	0.539457654	-0.983204574	-0.56087027	0.273767777	-1.29476005	-0.7207352
UBE2N	-0.55851695	0.06137263	-0.90667017	-0.847433003	-0.40813083	-0.89002415	-0.619443287	-0.11040496	0.00159909
EWSR1	-0.45623899	-1.298549522	-0.55730905	-1.433555741	-0.508364269	-1.25393538	0.020346192	-0.73833003	-0.193067
DDI2	-1.12476159	0.074342179	-1.02741443	-1.287874488	-0.556328367	-0.267324	-0.027439695	-0.15807215	-

Supplemental Table 2

Col_1	Col_2	N_tot	N_1	N_2	N_common	Odds_ratio	P_value
LIN28_pos	STAG2	64	6	17	2	1.425	0.6522
LIN28_pos	CKDN2A	64	6	9	1	1.245	1
LIN28_pos	TP53	64	6	7	3	12.43	0.01442
LIN28_pos	BRCA2	64	6	4	0	0	1
LIN28_pos	ANY	64	6	29	5	6.889	0.08342

Table S2 (refers to Figure 1)

Correlation between mutational status of selected genes from Supplementary Table S5 of [Brohl 2014] and *LIN28B* expression as determined from RNA-seq data.

Supplemental Table 3

EwS samples in the study

sample name	EwS1	EwS2	EwS3	EwS4
LIN28B status	pos	pos	neg	neg
age	pediatric	pediatric	adult	pediatric
tumor site	scapula	lung	vertebra	cranium
tumor staging	primary	metastatic	metastatic	primary
treated before sampling	no	yes	yes	no
EWS-FLI1 translocation	type 1	type 2	type 1	type 1

Table S3 (refers to Figure 2)

Clinical and genetic data relevant to the four primary EwS cultures.

Supplemental Table 4

EWS-FLI1 Most Responsive Direct Target Genes (111)

ACACB	HIST1H2AJ	OLFML3
ADGRG7	HIST1H3C	PADI2
ADRB3	HIST1H3G	PAQR5
AKAP7	HIST1H3J	PCDH20
ALK	HS3ST4	PHOSPHO1
AMER2	IGSF21	PKP2
ANKRD24	ISM2	POU3F1
APELA	ITGB2-AS1	PPP1R1A
ARTN	JPH1	PREX1
BCL11B	KCNA2	PRKCB
C1orf226	KCNE3	PRR5L
CCK	KCNG3	PRRT4
CD79A	KCNN1	RAMP1
CD83	KIAA0226L	RAP1GAP
CDH23	KIAA1456	RASGEF1B
CLSTN2	KMO	RBM11
CSPG5	LBH	RHOH
CYP4F22	LINC00463	RNF219
DCDC2	LINC00659	RRM2
DHRS13	LIPI	SAA2
DHRS2	LOC101927503	SCNN1G
DISC1FP1	LOC105369486	SH3GL3
DLG2	LOXHD1	SLAIN1
DPF3	MAB21L3	SLC15A2
DUSP26	MAP2K6	SLC24A3
EGR2	MKI67	SORD
EXO1	MND1	SOX2
FCGRT	MROH2A	SPIN2A
FEZF1	MYCN	STAR
FEZF1-AS1	MYOM2	STEAP2
FGF14	NCKAP1L	TNNI3
GFRA2	NGFR	TOX2
GNGT2	NKX2-2	TUBB4A
GSTM4	NPTXR	UGT3A2
GYG2	NPY1R	UTS2
HIST1H1B	NR0B1	ZCCHC12
HIST1H1D	NTRK1	ZNF620

Table S4 (refers to Figure 5)

List of the 111 most responsive genes to EWS-FLI1 (Boulay et al., 2018).

Supplemental Table 5

<i>Gene</i>	<i>Forward primer</i>	<i>Reverse primer</i>
<i>GAPDH</i>	5'-GGTCTCCTCTGACTTCAACA-3'	5'-GTGAGGGTCTCTCTCTTCCT-3'
<i>36B4</i>	5'-GCAATGTTGCCAGTGTCTGT-3'	5'-GCCTTGACCTTTTCAGCAAG-3'
<i>TBP</i>	5'-CGGCTGTTTAACTTCGCTTC-3'	5'-CACACGCCAAGAAACAGTGA-3'
<i>LIN28B</i>	5'-CTGTTTAGGAAGTGAAAGAAGAC-3'	5'-CACTTCTTTGGCTGAGGAGGTAG-3'
<i>HMG2</i>	5'-GCGCCTCAGAAGAGAGGAC-3'	5'-GTCTTCCCCTGGGTCTCTTAG-3'
<i>EWS-FLI1</i>	5'-AGCAGCCTCCACTAGTTAC-3'	5'-CCAAGCTCCTCTTCTGACTG-3'
<i>EWS</i>	5'-GTTCTCTCCTGGTCCGAAA-3'	5'-CAGCCTCCCACTAGTTACCC-3'
<i>18S</i>	5'-GGCCCGAAGCGTTTACTTTG-3'	5'-TTTCGCTCTGGTCCGTCTTG-3'
<i>FCGRT</i>	5'-TGGCGATGAGCACCCTAC-3'	5'-GATCCCACCACGAGCAC-3'
<i>CD79A</i>	5'-CAAGAACCGAATCATCACAGC-3'	5'-CGTTCTGCCATCGTTTCC-3'
<i>NTRK1</i>	5'-GGGCCTCTCCTTACAGGAAC-3'	5'-AGCTTCTGTTTCAGGCACTCC-3'
<i>SOX2</i>	5'-TTGCTGCCTCTTTAAGACTAGGA-3'	5'-TAAGCCTGGGGCTCAAAC-3'
<i>CCK</i>	5'-GCCCTGCTGGCAAGATAC-3'	5'-GCAGGTTCTTAACGATGGACA-3'
<i>RRM2</i>	5'-CACGGAGCCGAAAATAAGC-3'	5'-TCTGCCTTCTTATACATCTGCCA-3'
<i>AMER2</i>	5'-GCATCAAGCACCTGACCAACCT-3'	5'-TGAATCTGCGGACCAGCACTTT-3'
<i>NKX2.2</i>	5'-CAGCGACAACCCGTACAC-3'	5'-GACTTGGAGCTTGAGTCTGA-3'
<i>NGFR</i>	5'-TCATCCCTGTCTATTGCTCCA-3'	5'-TGTTCTGCTTGACAGCTGTTTC-3'

Table S5 (refers to STAR Methods)

Primer sequences for real-time qPCR.

Acknowledgments

I would like to thank all the members of my thesis committee, Prof. Marques-Vidal, Prof. Mazzolai, Prof. Petrova, Prof. Radtke and Prof. Schäfer, for following me during my thesis and for the positive discussions we had about this project.

I also want to thank the people who provided us the material that we used to generate all our results. Prof. C. Antonescu for sending us CDS frozen samples and Dr. Petur Nielsen for sending us CDS tumor slides. Prof. T.Nakamura, for allowing us to work with his CIC-DUX4 overexpression construct as well as a CDS primary cell line. Dr. M. Kawamura-Saito, Dr. R. Oyama and Dr. M. Yoshimatsu for establishing and sharing CDS cell lines.

I would like to thank all the facilities for their collaboration on this work : The EPFL histology core facility, the UNIL mouse pathology facility, the UNIL mouse facility, the UNIL flow cytometry facility and the UNIL cellular imaging facility.

Many thanks to the faculty of biology and medicine of the university of Lausanne as well as to the swiss institute for cancer research (ISREC) for providing me the funding and encourage the MD-PhD research program.

Je souhaite remercier le Professeur Stamenkovic pour m'avoir ouvert la porte qui mène au monde passionnant de la recherche et pour m'avoir dirigé dans le laboratoire du Professeur Riggi pour mon MD-PhD.

Merci Nik pour ta supervision, ta patience et ton soutien tout au long de ce travail. Malgré les nombreux rebondissements, les joies et les déceptions que m'ont procuré ce projet, tu as su me guider et m'aider à garder le cap. Pour résumer je souhaite citer Garry Kasparov : "To make 13 your special number makes you different from other people because other people don't like 13".

Je souhaite également remercier mes collègues des laboratoires Stamenkovic et Riggi : Thank you Sanal for your precious help and advices, you teached me everything I know about ChIP-seq and I'll always have a good time remembering our discussions. Merci Liliane pour maintenir un certain niveau d'organisation au labo malgré la tendance à dévier vers un état d'haute entropie. Thanks Santosh for your help in trying to catch the white whale (CIC-DUX4). Un grand merci à Patricia Martin qui génère une énorme partie de tous les résultats qui sortent des deux labos et avec qui j'apprécie discuter entre deux expériences. Many thanks to Tugba and Patricia Waszyk for your great collaboration on the LI***B project. Thanks to Carlo and Luisa for sharing your knowledge and experience with me on this project. Thanks to Emely and Sampath for the discussion we had in the lab during those long years. Merci aussi aux étudiants et étudiantes de nos deux labos qui fournissent un travail précieux pour l'avancée des recherches.

Merci à mes amis pour les rires et les innombrables souvenirs que je garde en votre compagnie. En particulier, merci à toi Marcone pour m'aider à résoudre la grande énigme du puzzle cosmique.

Merci à ma famille qui forme depuis toujours un environnement réconfortant qui participe beaucoup à la réalisation de mes projets. Bien sûr, si je suis arrivé jusqu'ici c'est bien grâce à mes parents qui m'ont toujours soutenu dans mes choix et mes décisions. Merci Catherine pour ton soutien inébranlable et ta capacité à me comprendre dans n'importe quelle situation. Merci Ozren pour m'avoir transmis l'intérêt pour la recherche et la curiosité du monde qui nous entoure.

La plus belle histoire qui ait démarré dans ce laboratoire est celle que je vis avec toi Giulia. Tu me comprends tous les jours et tu me supportes quand rien ne va avec ce projet. Cette thèse t'es entièrement dédiée car c'est grâce à toi que je suis allé jusqu'au bout de ces 4 années de doctorat. Avec Morgan, vous êtes ce qu'il y a de plus précieux à mes yeux et ce pourquoi je continuerai à aller

de l'avant quoi qu'il arrive. Je vous aime et je vous remercie pour tout l'amour que vous me donnez en retour.

References

1. Alaggio R, Coffin CM. The Evolution of Pediatric Soft Tissue Sarcoma Classification in the Last 50 Years. *Pediatr Dev Pathol* [Internet]. 2015 Nov 1 [cited 2020 Mar 24];18(6):481–94. Available from: <http://journals.sagepub.com/doi/10.2350/15-07-1666-MISC.1>
2. Doyle LA. Sarcoma classification: An update based on the 2013 World Health Organization Classification of Tumors of Soft Tissue and Bone. *Cancer* [Internet]. 2014 Jun 15 [cited 2018 Aug 14];120(12):1763–74. Available from: <http://doi.wiley.com/10.1002/cncr.28657>
3. Antonescu C. Round cell sarcomas beyond Ewing: emerging entities. *Histopathology* [Internet]. 2014 Jan 1 [cited 2018 Aug 14];64(1):26–37. Available from: <http://doi.wiley.com/10.1111/his.12281>
4. Specht K, Sung Y-S, Zhang L, Richter GH, Fletcher CD, Antonescu CR. Distinct Transcriptional Signature and Immunoprofile of CIC- DUX4–Fusion Positive Round Cell Tumors Compared to EWSR1- Rearranged Ewing Sarcomas – Further Evidence Toward Distinct Pathologic Entities. *Genes Chromosom Cancer* [Internet]. 2014 [cited 2017 Aug 21];53(7):622–33. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4108073/pdf/nihms612380.pdf>
5. Watson S, Perrin V, Guillemot D, Reynaud S, Coindre J-M, Karanian M, et al. Transcriptomic definition of molecular subgroups of small round cell sarcomas. *J Pathol* [Internet]. 2018 May 1 [cited 2018 Aug 20];245(1):29–40. Available from: <http://doi.wiley.com/10.1002/path.5053>
6. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell*. 2000.
7. Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. *Cell*. 2011.
8. Gröbner SN, Worst BC, Weischenfeldt J, Buchhalter I, Kleinheinz K, Rudneva VA, et al. The landscape of genomic alterations across childhood cancers. *Nature*. 2018;555(7696):321–7.
9. Filbin M, Monje M. Developmental origins and emerging therapeutic opportunities for childhood cancer. *Nat Med* [Internet]. 2019;25(3):367–76. Available from: <http://dx.doi.org/10.1038/s41591-019-0383-9>

10. Bhakta N, Force LM, Allemani C, Atun R, Bray F, Coleman MP, et al. Childhood cancer burden: a review of global estimates. *The Lancet Oncology*. 2019.
11. Ward ZJ, Yeh JM, Bhakta N, Frazier AL, Girardi F, Atun R. Global childhood cancer survival estimates and priority-setting: a simulation-based analysis. *Lancet Oncol* [Internet]. 2019 Jul 1 [cited 2020 Feb 12];20(7):972–83. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/31129029>
12. Alejandro Sweet-Cordero E, Biegel JA. The genomic landscape of pediatric cancers: Implications for diagnosis and treatment. *Science (80-)*. 2019;363(6432):1170–5.
13. Dupont C, Armant DR, Brenner CA. Epigenetics: Definition, mechanisms and clinical perspective. Vol. 27, *Seminars in Reproductive Medicine*. NIH Public Access; 2009. p. 351–7.
14. Allis CD, Jenuwein T. The molecular hallmarks of epigenetic control. *Nat Rev Genet* [Internet]. 2016;17(8):487–500. Available from: <http://dx.doi.org/10.1038/nrg.2016.59>
15. Dawson MA, Kouzarides T. Cancer epigenetics: from mechanism to therapy. *Cell* [Internet]. 2012 Jul 6 [cited 2014 Jul 11];150(1):12–27. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22770212>
16. Jones DTW, Banito A, Grünewald TGP, Haber M, Jäger N, Kool M, et al. Molecular characteristics and therapeutic vulnerabilities across paediatric solid tumours. *Nat Rev Cancer* [Internet]. 2019;19(8):420–38. Available from: <http://dx.doi.org/10.1038/s41568-019-0169-x>
17. Mertens F, Johansson B, Fioretos T, Mitelman F. The emerging complexity of gene fusions in cancer [Internet]. Nature Publishing Group. 2015 [cited 2020 Mar 23]. Available from: www.nature.com/reviews/cancer
18. Jo VY, Doyle LA. Refinements in Sarcoma Classification in the Current 2013 World Health Organization Classification of Tumours of Soft Tissue and Bone. Vol. 25, *Surgical Oncology Clinics of North America*. W.B. Saunders; 2016. p. 621–43.
19. Oda Y, Yamamoto H, Kohashi K, Yamada Y, Iura K, Ishii T, et al. Soft tissue sarcomas: From a morphological to a molecular biological approach. *Pathol Int* [Internet]. 2017 Sep 1 [cited

- 2020 Mar 23];67(9):435–46. Available from: <http://doi.wiley.com/10.1111/pin.12565>
20. Steele CD, Pillay N. The genomics of undifferentiated sarcoma of soft tissue: Progress, challenges and opportunities. *Semin Cancer Biol.* 2020;61(October 2019):42–55.
 21. Laetsch TW, Roy A, Xu L, Black JO, Coffin CM, Chi Y-Y, et al. Undifferentiated sarcomas in children harbor clinically-relevant oncogenic fusions and gene copy-number alterations: A report from the Children’s Oncology Group HHS Public Access. *Clin Cancer Res.* 2018;24(16):3888–97.
 22. Antonescu CR, Owosho AA, Zhang L, Chen S, Deniz K, Hury JM, et al. Sarcomas With CIC-rearrangements Are a Distinct Pathologic Entity With Aggressive Outcome A Clinicopathologic and Molecular Study of 115 Cases [Internet]. 2017 [cited 2018 Aug 22]. Available from: www.ajsp.com.
 23. Peters TL, Kumar V, Polikepahad S, Lin FY, Sarabia SF, Liang Y, et al. BCOR–CCNB3 fusions are frequent in undifferentiated sarcomas of male children. *Mod Pathol* [Internet]. 2015 Apr [cited 2018 Aug 22];28(4):575–86. Available from: <http://www.nature.com/articles/modpathol2014139>
 24. Haidar A, Arekapudi S, DeMattia F, Abu-Isa E, Kraut M. High-grade undifferentiated small round cell sarcoma with t(4;19)(q35;q13.1) CIC-DUX4 fusion: Emerging entities of soft tissue tumors with unique histopathologic features – A case report and literature review. *Am J Case Rep.* 2015 Feb 16;16:87–94.
 25. Italiano A, Sung YS, Zhang L, Singer S, Maki RG, Coindre JM, et al. High prevalence of CIC fusion with double-homeobox (DUX4) transcription factors in EWSR1-negative undifferentiated small blue round cell sarcomas. *Genes Chromosom Cancer.* 2012 Mar;51(3):207–18.
 26. Tamada H, Kobayashi M, Sano K, Uehara T, Matsumoto Y, Tateishi A, et al. Ultrastructure of CIC-DUX4 sarcoma: the first pathological report. *Ultrastruct Pathol* [Internet]. 2020;00(00):1–8. Available from: <https://doi.org/10.1080/01913123.2020.1737610>
 27. Le Guellec S, Velasco V, Pérot G, Watson S, Tirode F, Coindre JM. ETV4 is a useful marker

- for the diagnosis of CIC-rearranged undifferentiated round-cell sarcomas: A study of 127 cases including mimicking lesions. *Mod Pathol*. 2016;29(12):1523–31.
28. Siegele B, Roberts J, Black JO, Rudzinski E, Vargas SO, Galambos C. DUX4 Immunohistochemistry Is a Highly Sensitive and Specific Marker for CIC-DUX4 Fusion-positive Round Cell Tumor [Internet]. 2016 [cited 2018 Dec 4]. Available from: www.ajsp.com.
29. Richkind KE, Romansky SG, Finklestein JZ. t(4;19)(q35;q13.1): A recurrent change in primitive mesenchymal tumors? *Cancer Genet Cytogenet* [Internet]. 1996 Mar 1 [cited 2018 Aug 22];87(1):71–4. Available from: <https://www.sciencedirect.com/science/article/pii/0165460895002405?via%3Dihub>
30. Kawamura-Saito M, Yamazaki Y, Kaneko K, Kawaguchi N, Kanda H, Mukai H, et al. Fusion between CIC and DUX4 up-regulates PEA3 family genes in Ewing-like sarcomas with t(4;19)(q35;q13) translocation. *Hum Mol Genet* [Internet]. 2006 [cited 2017 Aug 21];15(13):2125–37. Available from: https://oup.silverchair-cdn.com/oup/backfile/Content_public/Journal/hmg/15/13/10.1093/hmg/ddl136/2/ddl136.pdf?Expires=1503421251&Signature=JZpB9Bzy7K35DhFngRHqP3KXyA8IrBZTJIPDo6uBA CHM9v9q4JL09jNhr6x9E7SEnlH-KiSqilx~k7KP6lhCP63EG-XEUz32gE2CBoS2nNjI3bULldg
31. Tsukamoto Y, Futani H, Yoshiya S, Watanabe T, Kihara T, Matsuo S, et al. Primary undifferentiated small round cell sarcoma of the deep abdominal wall with a novel variant of t(10;19) CIC-DUX4 gene fusion. *Pathol Res Pract* [Internet]. 2017 Oct [cited 2018 Dec 13];213(10):1315–21. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0344033817302558>
32. Jiménez G, Guichet A, Ephrussi A, Casanova J. Relief of gene repression by Torso RTK signaling: Role of capicua in Drosophila terminal and dorsoventral patterning. *Genes Dev*. 2000;14(2):224–31.
33. Jiménez G, Shvartsman SY, Paroush Z. The Capicua repressor - A general sensor of RTK

- signaling in development and disease. *J Cell Sci.* 2012;125(6):1383–91.
34. Keenan SE, Blythe SA, Marmion RA, Djabrayan NJ-V, Wieschus EF, Shvartsman SY. Rapid Dynamics of Signal-Dependent Transcriptional Repression by *Capicua*. *Dev Cell* [Internet]. 2020;1–8. Available from: <https://doi.org/10.1016/j.devcel.2020.02.004>
 35. Suisse A, He DQ, Legent K, Treisman JE. COP9 signalosome subunits protect *capicua* from MAPK-dependent and -independent mechanisms of degradation. *Dev.* 2017;144(14):2673–82.
 36. Roch F, Jiménez G, Casanova J. EGFR signalling inhibits *Capicua*-dependent repression during specification of *Drosophila* wing veins. *Development.* 2002;129(4):993–1002.
 37. Simón-Carrasco L, Jiménez G, Barbacid M, Drosten M. The *Capicua* tumor suppressor: a gatekeeper of Ras signaling in development and cancer. *Cell Cycle* [Internet]. 2018 Mar 19 [cited 2018 Nov 14];17(6):702–11. Available from: <https://www.tandfonline.com/doi/full/10.1080/15384101.2018.1450029>
 38. Pascual J, Jacobs J, Sansores-Garcia L, Natarajan M, Zeitlinger J, Aerts S, et al. Hippo Reprograms the Transcriptional Response to Ras Signaling. *Dev Cell* [Internet]. 2017;42(6):667-680.e4. Available from: <http://dx.doi.org/10.1016/j.devcel.2017.08.013>
 39. Papagianni A, Forés M, Shao W, He S, Koenecke N, Andreu MJ, et al. *Capicua* controls Toll/IL-1 signaling targets independently of RTK regulation. *Proc Natl Acad Sci U S A.* 2018;115(8):1807–12.
 40. Tanaka M, Yoshimoto T, Nakamura T. A double-edged sword: The world according to *Capicua* in cancer. *Cancer Sci* [Internet]. 2017 Dec 1 [cited 2018 Nov 14];108(12):2319–25. Available from: <http://doi.wiley.com/10.1111/cas.13413>
 41. Wong D, Lounsbury K, Lum A, Song J, Chan S, LeBlanc V, et al. Transcriptomic analysis of CIC and ATXN1L reveal a functional relationship exploited by cancer. *Oncogene* [Internet]. 2018 Aug 9 [cited 2018 Dec 13]; Available from: <http://www.nature.com/articles/s41388-018-0427-5>
 42. Dissanayake K, Toth R, Blakey J, Olsson O, Campbell DG, Prescott AR, et al. ERK/p90(RSK)/14-3-3 signalling has an impact on expression of PEA3 Ets transcription

- factors via the transcriptional repressor capicúa. *Biochem J* [Internet]. 2011 Feb 1 [cited 2018 Dec 13];433(3):515–25. Available from: <http://biochemj.org/lookup/doi/10.1042/BJ20101562>
43. Choi N, Park J, Lee JS, Yoe J, Park GY, Kim E, et al. miR-93/miR-106b/miR-375-CIC-CRABP1: A novel regulatory axis in prostate cancer progression. *Oncotarget*. 2015;6(27):23533–47.
 44. Zhou Y, Wang M, Shuang T, Liu Y, Zhang Y, Shi C. MiR-1307 influences the chemotherapeutic sensitivity in ovarian cancer cells through the regulation of the CIC transcriptional repressor. *Pathol Res Pract* [Internet]. 2019;215(10):152606. Available from: <https://doi.org/10.1016/j.prp.2019.152606>
 45. Miao LJ, Yan S, Zhuang QF, Mao QY, Xue D, He XZ, et al. miR-106b promotes proliferation and invasion by targeting capicua through MAPK signaling in renal carcinoma cancer. *Onco Targets Ther*. 2019;12:3595–607.
 46. Weissmann S, Cloos PA, Sidoli S, Jensen ON, Pollard S, Helin K. The Tumor Suppressor CIC Directly Regulates MAPK Pathway Genes via Histone Deacetylation. *Cancer Res* [Internet]. 2018 Aug 1 [cited 2018 Nov 16];78(15):4114–25. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29844126>
 47. Hwang I, Zheng H, Paik J. CIC is a critical regulator of neuronal differentiation Find the latest version : Caryl and Israel Englander Institute for Precision Medicine , NewYork-Presbyterian Correspondence : 2020;
 48. Wong D, Yip S. Making heads or tails – the emergence of capicua (CIC) as an important multifunctional tumour suppressor. *J Pathol*. 2020;
 49. Chittaranjan S, Chan S, Yang C, Yang KC, Chen V, Moradian A, et al. Mutations in CIC and IDH1 cooperatively regulate 2-hydroxyglutarate levels and cell clonogenicity. *Oncotarget* [Internet]. 2014 Sep 15 [cited 2018 Dec 14];5(17):7960–79. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25277207>
 50. Yang R, Chen LH, Hansen LJ, Carpenter AB, Moure CJ, Liu H, et al. Cic Loss Promotes Gliomagenesis via Aberrant Neural Stem Cell Proliferation and Differentiation. *Cancer Res*

- [Internet]. 2017 Nov 15 [cited 2018 Nov 16];77(22):6097–108. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28939681>
51. Ahmad ST, Rogers AD, Chen MJ, Dixit R, Adnani L, Frankiw LS, et al. Capicua regulates neural stem cell proliferation and lineage specification through control of Ets factors. *Nat Commun* [Internet]. 2019;10(1):1–17. Available from: <http://dx.doi.org/10.1038/s41467-019-09949-6>
 52. Park S, Lee S, Lee C-G, Park GY, Hong H, Lee J-S, et al. Capicua deficiency induces autoimmunity and promotes follicular helper T cell differentiation via derepression of ETV5. *Nat Commun* [Internet]. 2017 Jul 12 [cited 2018 Nov 15];8:16037. Available from: <http://www.nature.com/doi/10.1038/ncomms16037>
 53. Sizemore GM, Pitarresi JR, Balakrishnan S, Ostrowski MC. The ETS family of oncogenic transcription factors in solid tumours. *Nat Rev Cancer* [Internet]. 2017 Apr 28 [cited 2018 Dec 17];17(6):337–51. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28450705>
 54. Han F, Zhang J, Ma S, Chen X, Liu W, He X, et al. Altered capicua transcriptional repressor gene expression exhibits distinct prognostic value for isocitrate dehydrogenase-mutant oligodendroglial tumors. *Oncol Lett*. 2018;15(2):1459–68.
 55. Bunda S, Heir P, Metcalf J, Li ASC, Agnihotri S, Pusch S, et al. CIC protein instability contributes to tumorigenesis in glioblastoma. *Nat Commun* [Internet]. 2019;10(1). Available from: <http://dx.doi.org/10.1038/s41467-018-08087-9>
 56. Tan Q, Brunetti L, Rousseaux MWC, Lu H-C, Wan Y-W, Revelli J-P, et al. Loss of Capicua alters early T cell development and predisposes mice to T cell lymphoblastic leukemia/lymphoma. *Proc Natl Acad Sci U S A* [Internet]. 2018 Feb 13 [cited 2018 Dec 14];115(7):E1511–9. Available from: <http://www.pnas.org/lookup/doi/10.1073/pnas.1716452115>
 57. Yoe J, Kim D, Kim S, Lee Y. Capicua restricts cancer stem cell-like properties in breast cancer cells. *Oncogene* [Internet]. 2020; Available from: <http://dx.doi.org/10.1038/s41388-020-1230->

58. Kim E, Kim D, Lee J-S, Yoe J, Park J, Kim C-J, et al. Capicua suppresses hepatocellular carcinoma progression by controlling the ETV4-MMP1 axis. *Hepatology* [Internet]. 2018 Jun [cited 2018 Dec 14];67(6):2287–301. Available from: <http://doi.wiley.com/10.1002/hep.29738>
59. Okimoto RA, Breitenbuecher F, Olivas VR, Wu W, Gini B, Hofree M, et al. Inactivation of Capicua drives cancer metastasis. *Nat Genet* [Internet]. 2017 Jan 21 [cited 2018 Dec 14];49(1):87–96. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27869830>
60. Hashiba T, Yamashita T, Okada H, Nio K, Hayashi T, Asahina Y, et al. Inactivation of transcriptional repressor Capicua confers sorafenib-resistance in human hepatocellular carcinoma. *Cell Mol Gastroenterol Hepatol* [Internet]. 2020; Available from: <https://doi.org/10.1016/j.jcmgh.2020.02.009>
61. Gabriëls J, Beckers M-C, Ding H, De Vriese A, Plaisance S, van der Maarel S., et al. Nucleotide sequence of the partially deleted D4Z4 locus in a patient with FSHD identifies a putative gene within each 3.3 kb element. *Gene* [Internet]. 1999 Aug 5 [cited 2018 Nov 19];236(1):25–32. Available from: <https://www.sciencedirect.com/science/article/pii/S037811199900267X?via%3Dihub>
62. Gatica LV, Rosa AL. A complex interplay of genetic and epigenetic events leads to abnormal expression of the DUX4 gene in facioscapulohumeral muscular dystrophy. *Neuromuscul Disord* [Internet]. 2016 Dec 1 [cited 2018 Nov 19];26(12):844–52. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0960896616300876>
63. van der Maarel SM, Tawil R, Tapscott SJ. Facioscapulohumeral muscular dystrophy and DUX4: breaking the silence. *Trends Mol Med* [Internet]. 2011 May 1 [cited 2018 Nov 19];17(5):252–8. Available from: <https://www.sciencedirect.com/science/article/pii/S1471491411000025?via%3Dihub>
64. Choi SH, Gearhart MD, Cui Z, Bosnakovski D, Kim M, Schennum N, et al. DUX4 recruits p300/CBP through its C-terminus and induces global H3K27 acetylation changes. *Nucleic Acids Res* [Internet]. 2016 Jun 20 [cited 2018 Nov 20];44(11):5161–73. Available from:

<https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkw141>

65. Mitsuhashi H, Ishimaru S, Homma S, Yu B, Honma Y, Beermann M Lou, et al. Functional domains of the FSHD-associated DUX4 protein. *Biol Open*. 2018;7(4).
66. Bosnakovski D, da Silva MT, Sunny ST, Ener ET, Toso EA, Yuan C, et al. A novel P300 inhibitor reverses DUX4-mediated global histone H3 hyperacetylation, target gene expression, and cell death. *Sci Adv*. 2019;5(9).
67. Resnick R, Wong CJ, Hamm DC, Bennett SR, Skene PJ, Hake SB, et al. DUX4-Induced Histone Variants H3.X and H3.Y Mark DUX4 Target Genes for Expression. *Cell Rep*. 2019;29(7):1812-1820.e5.
68. De Iaco A, Planet E, Coluccio A, Verp S, Duc J, Trono D. DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nat Genet* [Internet]. 2017 May 1 [cited 2018 Nov 20];49(6):941–5. Available from: <http://www.nature.com/doi/10.1038/ng.3858>
69. Young JM, Whiddon JL, Yao Z, Kasinathan B, Snider L, Geng LN, et al. DUX4 Binding to Retroelements Creates Promoters That Are Active in FSHD Muscle and Testis. *PLoS Genet*. 2013;9(11).
70. Lim KRQ, Nguyen Q, Yokota T. Dux4 signalling in the pathogenesis of facioscapulohumeral muscular dystrophy. *Int J Mol Sci*. 2020;21(3):1–17.
71. Shadle SC, Zhong JW, Campbell AE, Conerly ML, Jagannathan S, Wong CJ, et al. DUX4-induced dsRNA and MYC mRNA stabilization activate apoptotic pathways in human cell models of facioscapulohumeral dystrophy. *PLoS Genet*. 2017;13(3):1–25.
72. Dmitriev P, Bou Saada Y, Dib C, Ansseau E, Barat A, Hamade A, et al. DUX4-induced constitutive DNA damage and oxidative stress contribute to aberrant differentiation of myoblasts from FSHD patients. *Free Radic Biol Med* [Internet]. 2016;99:244–58. Available from: <http://dx.doi.org/10.1016/j.freeradbiomed.2016.08.007>
73. Serra C, Wagner KR, Serra C, Wagner KR. It ' s not all about muscle : fibroadipogenic progenitors contribute to facioscapulohumeral muscular dystrophy It ' s not all about muscle :

- fibroadipogenic progenitors contribute to facioscapulohumeral muscular dystrophy. 2020;
74. Bury M, Le Calvé B, Lessard F, Dal Maso T, Saliba J, Michiels C, et al. NFE2L3 Controls Colon Cancer Cell Growth through Regulation of DUX4, a CDK1 Inhibitor. *Cell Rep.* 2019;29(6):1469-1481.e9.
 75. Zhang J, McCastlain K, Yoshihara H, Xu B, Chang Y, Churchman ML, et al. Deregulation of DUX4 and ERG in acute lymphoblastic leukemia. *Nat Genet* [Internet]. 2016 Dec 24 [cited 2018 Nov 21];48(12):1481–9. Available from: <http://www.nature.com/articles/ng.3691>
 76. Chew GL, Campbell AE, De Neef E, Sutliff NA, Shadle SC, Tapscott SJ, et al. DUX4 Suppresses MHC Class I to Promote Cancer Immune Evasion and Resistance to Checkpoint Blockade. *Dev Cell* [Internet]. 2019;50(5):658-671.e7. Available from: <https://doi.org/10.1016/j.devcel.2019.06.011>
 77. Dmitriev P, Kiseleva E, Kharchenko O, Ivashkin E, Pichugin A, Dessen P, et al. Dux4 controls migration of mesenchymal stem cells through the Cxcr4-Sdf1 axis. *Oncotarget.* 2016;7(40):65090–108.
 78. Preussner J, Zhong J, Sreenivasan K, Günther S, Engleitner T, Künne C, et al. Oncogenic Amplification of Zygotic Dux Factors in Regenerating p53-Deficient Muscle Stem Cells Defines a Molecular Cancer Subtype. *Cell Stem Cell.* 2018;23(6):794-805.e4.
 79. Forés M, Simó n-Carrasco L, Ajuria L, ria Samper N, Gonzá lez-Crespo S, Drosten M, et al. A new mode of DNA binding distinguishes Capicua from other HMG-box factors and explains its mutation patterns in cancer. 2017 [cited 2018 Nov 23]; Available from: <http://cancer.sanger.ac.uk/cosmic>
 80. Yoshimoto T, Tanaka M, Homme M, Yamazaki Y, Takazawa Y, Antonescu CR, et al. CIC-DUX4 induces small round cell sarcomas distinct from ewing sarcoma. *Cancer Res.* 2017;77(11):2927–37.
 81. Okimoto RA, Wu W, Nanjo S, Olivas V, Lin YK, Ponce RK, et al. CIC-DUX4 oncoprotein drives sarcoma metastasis and tumorigenesis via distinct regulatory programs. *J Clin Invest.* 2019;129(8):3401–6.

82. Lazo de la Vega L, Hovelson DH, Cani AK, Liu C-J, McHugh JB, Lucas DR, et al. Targeted next-generation sequencing of CIC-DUX4 soft tissue sarcomas demonstrates low mutational burden and recurrent chromosome 1p loss. *Hum Pathol* [Internet]. 2016 Dec 1 [cited 2018 Dec 17];58:161–70. Available from: <https://www.sciencedirect.com/science/article/pii/S004681771630212X?via%3Dihub>
83. Smith SC, Buehler D, Choi EYK, McHugh JB, Rubin BP, Billings SD, et al. CIC-DUX sarcomas demonstrate frequent MYC amplification and ETS-family transcription factor expression. *Mod Pathol* [Internet]. 2015;28(1):57–68. Available from: <http://dx.doi.org/10.1038/modpathol.2014.83>
84. Oyama R, Takahashi M, Yoshida A, Sakumoto M, Takai Y, Kito F, et al. Generation of novel patient-derived CIC-DUX4 sarcoma xenografts and cell lines. *Sci Rep*. 2017;7(1):1–13.
85. Yoshimatsu Y, Noguchi R, Tsuchiya R, Kito F, Sei A, Sugaya J, et al. Establishment and characterization of NCC-CDS2-C1: a novel patient-derived cell line of CIC-DUX4 sarcoma. *Hum Cell* [Internet]. 2020;(0123456789):1–10. Available from: <https://doi.org/10.1007/s13577-019-00312-x>
86. Nakai S, Yamada S, Outani H, Nakai T, Yasuda N, Mae H, et al. Establishment of a novel human CIC-DUX4 sarcoma cell line, Kitra-SRS, with autocrine IGF-1R activation and metastatic potential to the lungs. *Sci Rep*. 2019 Dec 1;9(1):1–15.
87. Oyama R, Takahashi M, Yoshida A, Sakumoto M, Takai Y, Kito F, et al. Generation of novel patient-derived CIC- DUX4 sarcoma xenografts and cell lines. *Sci Rep* [Internet]. 2017 Jul 5 [cited 2018 Dec 6];7(1):4712. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28680140>
88. Riggi N, Cironi L, Provero P, Suva M-L, Kaloulis K, Garcia-Echeverria C, et al. Development of Ewing’s Sarcoma from Primary Bone Marrow-Derived Mesenchymal Progenitor Cells. *Cancer Res* [Internet]. 2005 Dec 15 [cited 2016 Nov 15];65(24):11459–68. Available from: <http://cancerres.aacrjournals.org/cgi/doi/10.1158/0008-5472.CAN-05-1696>
89. Lasko LM, Jakob CG, Edalji RP, Qiu W, Montgomery D, Digiammarino EL, et al. Discovery

- of a selective catalytic p300/CBP inhibitor that targets lineage-specific tumours. *Nature* [Internet]. 2017 Sep 27;550:128. Available from: <https://doi.org/10.1038/nature24028>
90. Riggi N, Knoechel #1 B, Gillespie #1 SM, Rheinbay E, Boulay G, Suvà ML, et al. EWS-FLI1 utilizes divergent chromatin remodeling mechanisms to directly activate or repress enhancer elements in Ewing sarcoma The aberrant transcription factor EWS-FLI1 drives Ewing sarcoma yet its molecular function is incompletely understood. We find that EWS-FLI1 reprograms gene regulatory circuits in Ewing HHS Public Access. *Cancer Cell* [Internet]. 2014 [cited 2017 Aug 23];26(5):668–81. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4492343/pdf/nihms-639914.pdf>
91. Riggi N, Suvà ML, Suvà D, Cironi L, Provero P, Tercier S, et al. EWS-FLI-1 expression triggers a ewing’s sarcoma initiation program in primary human mesenchymal stem cells. *Cancer Res* [Internet]. 2008 Apr 1 [cited 2016 Nov 15];68(7):2176–85. Available from: <http://cancerres.aacrjournals.org/cgi/doi/10.1158/0008-5472.CAN-07-1761>
92. Riggi N, Knoechel B, Gillespie SM, Rheinbay E, Boulay G, Suvà ML, et al. EWS-FLI1 utilizes divergent chromatin remodeling mechanisms to directly activate or repress enhancer elements in Ewing sarcoma. *Cancer Cell* [Internet]. 2014 Nov 10 [cited 2018 Dec 18];26(5):668–81. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25453903>
93. Lasko LM, Jakob CG, Edalji RP, Qiu W, Montgomery D, Digiammarino EL, et al. Discovery of a selective catalytic p300/CBP inhibitor that targets lineage-specific tumours. *Nature* [Internet]. 2017 Sep 27 [cited 2018 Dec 4];550(7674):128. Available from: <http://www.nature.com/doi/10.1038/nature24028>