

**Elucidating the physiological function of the RNA binding
protein EWS and its role in post-transcriptional gene
regulation.**

Inaugural-Dissertation

Zur Erlangung des Doktorgrades der Mathematisch-Naturwissenschaftlichen
Fakultät der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Sujitha Smruthy Duggimpudi

aus Velgonda/Indien

Düsseldorf, im Juli 2015

Aus der Klinik für Kinder-Onkologie, Hämatologie und Klinische Immunologie des
Universitätsklinikum Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Referent: Prof. Dr. Arndt Borkhardt
Klinik für Kinder-Onkologie,-Hämatologie und Klinische Immunologie
Universitätsklinikum der Heinrich-Heine-Universität Düsseldorf

Koreferent: Prof. Dr. Dieter Willbold
Institut für Physikalische Biologie
Heinrich-Heine-Universität Düsseldorf

Tag der mündlichen Prüfung: 14.07.2015

Dedicated to my parents Annamary and Sudhakar and to my aunt and uncle Sujatha and Prathap

Table of contents

SUMMARY OF THE THESIS:	7
ZUSAMMENFASSUNG:	8
1. INTRODUCTION	9
1.1 POST-TRANSCRIPTIONAL MODIFICATIONS	9
1.2 RNA-BINDING PROTEINS (RBPs).....	11
1.3 FUNCTIONS OF RBPs	12
1.3.1 Capping	13
1.3.2 Splicing	13
1.3.3 RNA editing	13
1.3.4 mRNA stability	14
1.3.5 Alternative splicing.....	14
1.3.6 Alternative cleavage and polyadenylation	14
1.3.7 mRNA export.....	15
1.3.8 mRNA localization	15
1.3.9 Translation	15
1.3.10 miRNA maturation.....	16
1.4 RNA-BINDING PROTEINS AND DISEASES	16
1.4.1 Translocations of RBPs.....	17
1.4.2 Loss of function effects by RBPs.....	17
1.4.3 Gain of function effects by RBPs	18
1.4.4 Altered expression of RBPs	18
1.5 EXPERIMENTAL METHODS TO IDENTIFY THE TARGETS OF RNA-BINDING PROTEINS	19
1.5.1 Microarray based immunoprecipitation	19
1.5.2 SELEX and RNAcompete	19
1.5.3 Yeast three-hybrid system.....	20
1.5.4 Ribonucleoprotein Immunoprecipitation (RIP)	21
1.5.5 CLIP (Crosslinking and Immunoprecipitation)	21
1.6 PAR-CLIP	22
1.6.1 Brief overview of the procedure	23
1.6.2 Bioinformatic methods to analyze the CLIP NGS datasets	24
1.6.3 Limitations of PAR-CLIP	25

1.7 FET FAMILY RNA-BINDING PROTEINS.....	26
1.7.1 Functions of FET family proteins	28
1.8 EWING SARCOMA BREAK POINT REGION 1 (EWSR1).....	29
1.8.1 Ewing Sarcoma	29
1.8.2 Functions of EWS	31
2. RESULTS AND DISCUSSION.....	34
2.1 PUBLICATION I: THE CELL CYCLE REGULATOR CCDC6 IS A KEY TARGET OF RNA- BINDING PROTEIN EWS.	35
2.1.1 Abstract	35
2.1.2 Results:.....	35
2.1.3 Discussion	38
2.2 PUBLICATION II: RNA TARGETS OF WILD TYPE AND MUTANT FET FAMILY PROTEINS.....	44
2.2.1 Abstract	44
2.2.2 Results.....	44
2.2.3 Discussion	47
3. REFERENCES	52
4. ABBREVIATIONS	60
5. PUBLICATION I.....	65
6. SUPPLEMENTARY INFORMATION FOR PUBLICATION I.....	77
6.1 SUPPLEMENTARY FIGURES	77
6.2 SUPPLEMENTARY TABLES	80
7. PUBLICATION II.....	84
8. SUPPLEMENTARY INFORMATION FOR PUBLICATION II	88
8.1 SUPPLEMENTARY FIGURES	88
8.2 SUPPLEMENTARY TABLES.....	92
8.3 SUPPLEMENTARY METHODS	94
9. ACKNOWLEDGEMENTS.....	102
CURRICULUM VITAE	103
AFFIRMATION.....	105

Table of Figures:

Figure 1-Functions of RNA binding proteins inside the cell.	12
Figure 2-Network of RBPs in human diseases.....	16
Figure 3-Schematic representation of PAR-CLIP	24
Figure 4-Protein domain organisation of FET proteins.	27
Figure 5-Protein domain organization of EWSR1 and FLI1.	30

Summary of the thesis:

In eukaryotes extensive post-transcriptional processing occurs resulting in the production of mature mRNAs from pre-mRNA. This processing includes splicing, polyadenylation, editing etc., which are mainly mediated by RNA binding proteins (RBPs). There are around ~1540 RBPs in humans, a number which represents 7.5% of all protein coding genes in humans. These RBPs control every aspect of RNA biology, from transcription to RNA modification, transport, localization, turnover and translation. Perturbations in the expression of RBPs affects their target RNAs thus leading to multiple diseases. Among various RBP families, FET family proteins have attracted wide attention from the scientific community since all members are involved in genomic rearrangements causing sarcomas and other cancers. The FET family consists of the three different proteins FUS, EWS and TAF15. They play a significant role in mRNA biogenesis and functions. We focussed our attention onto the second member of the FET family protein, EWS, which is –when translocated- known to cause Ewing Sarcoma. So far, research has focused on the chimeric transcription factors, while the putative physiological function of heterozygous EWSR1 loss in these tumors has not been thoroughly investigated. We have identified various mRNAs bound to all three FET members using PAR-CLIP. To identify the regulated targets among bound targets, we performed microarrays after knocking down EWS in HEK 293 T cells and analyzed the targets which were significantly down- or upregulated. We demonstrate that CCDC6, a known cell cycle regulator protein, is a novel target regulated by EWS. siRNA mediated down regulation of EWS caused an elevated rate of apoptosis in cells in a CCDC6-dependant manner. This effect was rescued upon re-expression of CCDC6. We observed a decrease in the number of cells in S and G2/M phase upon EWS knockdown and an increased cell number with a sub G0/G1 DNA content typical of apoptotic cells. This study provides evidence for a novel functional mechanism through which wild-type EWS operates in a target-dependant manner in Ewing Sarcoma.

Zusammenfassung:

Auf dem Weg zur reifen mRNA findet in Eukaryonten eine umfangreiche post-transkriptionelle Prozessierung statt. Diese wird hauptsächlich durch RNA-bindende Proteine (RBPs) ausgeführt und beinhaltet u.a. Spleißen, Polyadenylierung und RNA-Editierung. Es gibt ca. 1540 RBPs im Menschen, welche zusammen 7,5% aller kodierenden Gene ausmachen. RBPs kontrollieren alle Aspekte der Biologie von RNAs, von der Transkription zu RNA Modifizierungen, Transport, subzelluläre Lokalisierung, Lebensdauer der RNAs und Translation. Eine veränderte Expression von RBPs resultiert in einer gestörten (Target) Regulation und in der Folge in der Entstehung von Krankheiten. Die FET RBP Familie, die aus den drei Mitgliedern FUS, EWS und TAF15 besteht und wichtige Aufgaben in der mRNA Biogenese übernimmt, ist Gegenstand intensiver Forschung, da genomische Rearrangements aller drei Proteine in Sarkomen und anderen Krebsarten beobachtet werden. Inhalt dieser Arbeit ist eine Analyse der Funktionen des RBPs EWS, dessen Translokation zu der Entstehung von Ewing-Sarkomen führt. Bisher wurden hauptsächlich die Auswirkungen des aus dieser Translokation resultierenden Transkriptionsfaktors untersucht, wohingegen nur wenig über die physiologischen Aufgaben des Proteins und die Auswirkungen des heterozygoten Verlustes des Proteins in Tumorzellen bekannt ist. Wir haben die durch die FET Proteine gebundenen mRNAs in HEK293 T Zellen identifiziert und anschließend durch Knockdownexperimente und Microarrays das Subset der durch EWS regulierten Gene. Wir können zeigen, dass die mRNA von CCDC6 -einem Protein mit Aufgaben in der Zellzyklusregulation- durch EWS gebunden und in seiner Expression reguliert wird. Eine durch siRNAs vermittelte verringerte EWS Expression führt zu einer erhöhten, CCDC6-abhängigen Apoptoserate, die durch eine CCDC6-Reexpression rückgängig gemacht werden kann. Hierzu passend führt eine verringerte EWS Expression zu einer erhöhten Anzahl von Zellen in sub G0/G1 Zellzyklusphasen, wie sie typisch für apoptotische Zellen sind. Zusammenfassend beschreiben wir einen neuen Mechanismus, durch den EWS seine Funktionen in der Zelle wahrnimmt und der in Ewing-Sarkomen vermutlich auf Grund der EWS Haploinsuffizienz gestört

1. Introduction

1.1 Post-transcriptional modifications

For many years, mRNAs were merely described by the scientific world as ‘just’ a carrier of information from DNA to protein. This simplistic term undermined the complexity, diversity of the roles that are played by mRNAs. Lately, RNA emerged from this naive state and etched a niche for itself, which we now call “RNA biology”. RNA comes in many kinds where some RNAs like messenger RNA, transfer RNA and ribosomal RNA play roles in protein synthesis, while others like small nucleolar RNA (snoRNA), small nuclear RNA (snRNA), guide RNA have roles in post-transcriptional modification or DNA replication. With advances in sequencing technologies many new kinds of RNAs, with roles in transcriptional and translational regulation are being discovered, which are together called regulatory/catalytic RNAs. These include microRNAs (miRNAs), small interfering RNAs (siRNAs), long non-coding RNAs (lncRNAs), PIWI interacting RNAs (piRNAs) etc. Extensive research is now been carried out to study these different forms of RNA only to reveal the complexities and dynamic roles that are played by them.

The central dogma of life states that the information of life flows from DNA to RNA and further to protein. DNA initially only gives birth to naked RNA which is escorted through its life’s journey from nucleus until its death in cytoplasm by multitudes of proteins, which are called RNA-binding proteins (RBPs). Along with RBPs several small RNAs like microRNAs and non-coding RNAs also bind to mRNAs. The mRNAs together with RBPs and small RNAs constitute messenger ribonucleo protein particles (mRNPs). There are many different kinds of mRNP complexes like eIF4e, eIF4G, EJC, PABPs and SR proteins, which bind to

specific sites on the mRNAs. These complexes target mostly two regions in the mRNAs, namely the 5' end and 3' ends. The 5' end of all RNAII polymerase transcripts is characterized by the presence of a 7-methylguanosine cap (by cap binding proteins like CBC20/80, eIF4G, eIF4e) [1] and the 3' end is characterized by the poly A tail (by poly A binding proteins) [2]. The other type of mRNP complexes which bind independently of the sequence of RNA include Y box containing proteins, which are required for the packing of mRNAs [3] and SR complexes which shuttle between nucleus and cytoplasm and coordinate multiple functions like splicing, export, translation and degradation of mRNAs [4]. The other regions where these mRNP complexes bind are exon junctions which are called exon junction complexes (EJCs). These EJCs bind sequence specifically and aid in splicing [5]. Apart from the above mentioned regions of mRNA binding, the majority of the RNA-binding factors recognize so-called "RNA recognition elements", namely specific sequences or specific structure of the mRNAs. These elements are mostly present in the 3' and 5' untranslated regions of mRNA.

These mRNPs collectively determine the future of the mRNAs they are bound to, by subjecting them to various modifications called post-transcriptional gene regulation (PTGR). This regulation orchestrates processes like maturation, transport, stability and translation of mRNAs. Post-transcriptional gene regulatory mechanisms thus drive the steady-state protein translation, thereby maintaining the dynamic functions of the cells. Our focus in this thesis will be on RBPs, which form the core elements of PTGR.

1.2 RNA-binding proteins (RBPs)

The story of RBPs dates back to 1950's, when scientists observed that elongating transcripts of lampbrush chromosomes under electron microscope were packed with proteins, suggesting that RNAs are associated with different kinds of proteins [6]. For almost 20 years this issue was not taken up for study, since research on DNA had been in the focus. Only in 1970 several labs started working on the composition of RNA and discovered key mRNA components. Since then several hundreds of new proteins that are bound to RNA have been discovered up until today. This is partly due to the major advances in the molecular approaches to study the subcellular components of the cell. Modern techniques like microarrays, mass spectrometry and –more recently- next-generation sequencing are acting as virtual binoculars to look into the depths and hidden crevices of a cell. To date, in budding yeast alone there are at least ~500 proteins identified to be bound to RNA [7] and this number is certain to rise in the more complex higher forms like humans [8]. The RNA-binding domain, RRM, alone is present in nearly 500 human proteins [9]. Other frequently found RNA binding domains include the K homology domain (KH domain), zinc finger domains, RGG boxes, double stranded RNA binding domains (dsRBD) etc.[10]. A recent publication by Gerstberger et al has pushed the number of RBPs to 1542, a number which represents 7.5% of all protein coding genes in humans. Among these 1542 genes, the authors identified ~600 structurally distinct RNA-binding domains, where only 20 classes of RBDs have more than 10 gene members each and others had only 1 or 2 members, thus revealing the complexities posed in identifying RBPs based on their domains. Among the ~700 mRBPs, 405 contained RNA recognition motif, a K homology domain, a DEAD motif, a double-stranded RNA-binding motif or a zinc finger domain whereas the ~170 ribosomal proteins contained 119 distinct domains exclusively confined to them [11].

1.3 Functions of RBPs

Although many RBPs have been identified so far, the biological functions of many of these proteins have not been studied yet in detail [11]. Functions of those already studied have revealed the diverse roles played by RBPs ranging from influencing transcription, alternative splicing, polyadenylation to RNA modifications, localization, transport, translation and turnover [12]

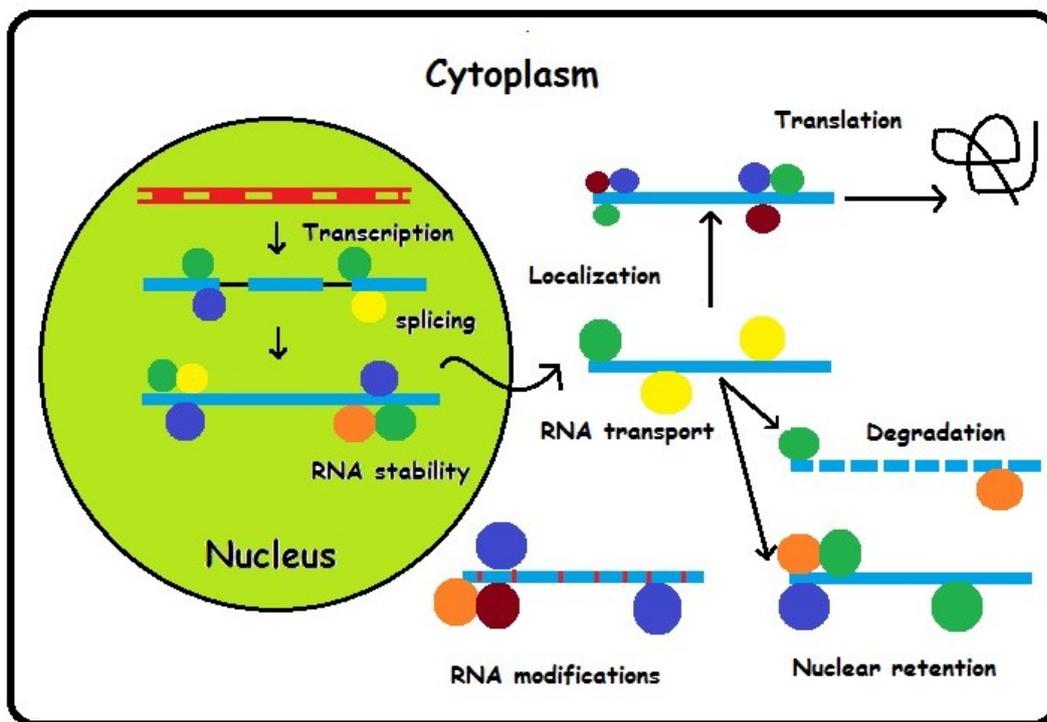


Figure 1-Functions of RNA binding proteins inside the cell.

RBPs are expressed either ubiquitously, and interact with almost any type of RNA, regulating basic events in gene expression (e.g. the RBP PABP1), while others are expressed only in particular cell types and are involved in regulating only small subsets of RNA targets like Musashi1 [13]. Almost all categories of RBPs are directly or indirectly taking part in protein synthesis. Nearly 700 RBPs are involved in mRNA binding, around 170 are ribosomal proteins and 130 proteins are involved in tRNA biogenesis, while others are taking part in

rRNA and snoRNA biogenesis [11]. Perturbations in the complex post-transcriptional events regulated by RBPs naturally leads to disease. Therefore it is important to understand the functions of any given RBP.

In the following, PTGR regulated by RBPs are explained in more detail.

1.3.1 Capping

Capping occurs on the nascent mRNA as soon as it emerges after transcription. The cap structure protects the mRNA from degradation by exonucleases. It also helps with the optimal translation of the mRNA as well as ribosomal binding. Two major classes of cap-binding proteins are eukaryotic translation initiation factor 4E (eIF4E) in the cytoplasm and nuclear cap binding complex (nCBC) (Reviewed in [14]).

1.3.2 Splicing

Splicing is a process which removes the introns and non-coding regions in order to generate the mature mRNA which further gets translated to protein. Spliceosomes bind on either side of an intron, looping the intron into a circle and then cleaving it off. The two ends of the exons are then joined together. The splicing machinery is highly dynamic and consists of small nuclear RNAs and proteins like Brr2, Snu114 and Prp24 etc. (Reviewed in [15]).

1.3.3 RNA editing

RNA editing is a complex process, which results in sequence variation in the RNA molecule, and is catalyzed by enzymes called Adenosine Deaminase Acting on RNA (ADAR) enzymes, which convert specific adenosine residues to inosine in an mRNA molecule by hydrolytic deamination. These enzymes can alter splicing, translational machinery, the double stranded RNA structures and the binding affinity between RNA and RBPs. In humans, three ADAR proteins are known: ADAR1, ADAR2 and ADAR3. The first two proteins are ubiquitously expressed while the third one only being expressed in brain. Dysregulation of RNA editing

through ADAR enzymes was reported to cause several diseases ranging from cancer to ALS (Reviewed in [16]).

1.3.4 mRNA stability

mRNA stability is defined as the half-life of any mRNA which will ultimately determine the steady-state levels of that particular mRNA. The control of mRNA stability is carried out by specific cis-acting elements (sequence specific control elements) and trans-acting elements (RBPs and some miRNAs). The cis-regulating elements are usually the adenylate- and uridylate (AU)-rich elements (ARE) which are specifically bound by trans-acting RBPs. These elements ultimately determine whether mRNA decay is delayed or facilitated. Dysregulation of mRNA stability is associated with human diseases such as cancer and Alzheimers disease (Reviewed in [17]).

1.3.5 Alternative splicing

Nearly 74% of proteins are products of alternative splicing mechanism [18]. Few examples of RBPs regulating alternative splicing include the neuronal specific Nova proteins and the TAR DNA binding protein (TDP43). Nova proteins bind to the pre-messenger RNAs of genes like neogenin, Flamingo 1 and JNK2 and control their alternative splicing. These mRNAs have roles in maintaining neuronal plasticity and the loss of Nova proteins causes several motor neurological disorders [19]. TDP43 on other hand controls the alternative splicing of Cystic fibrosis CFTR mRNA which encodes a chloride channel [20].

1.3.6 Alternative cleavage and polyadenylation

Almost all eukaryotic pre-mRNAs are processed at the 3' end by cleavage and polyadenylation (APA), which terminates transcription and adds the poly(A) tail. The addition of the poly(A) tail is also called polyadenylation. As the name suggests, it consists of a stretch of RNA that is made solely of adenine bases which are added to the 3' end of mRNA after the

3' UTR regions. It protects the mRNA from 3' exonucleases to increase the half-life of the mRNA. It governs the stability, nuclear transport and translation of mRNAs (Reviewed in [21]). APA expands the repertoire of transcripts expressed from the genome and is highly regulated by various RBPs. Some examples are CPSF-73 and PABPN1 which together activate the poly (A) polymerase which is inactive on its own. Mutated PABPN1 causes oculopharyngeal muscular dystrophy (OPMD) (Reviewed in [22]).

1.3.7 mRNA export

Only the fully processed mRNAs should be transported to the cytoplasm from the nucleus for protein production and therefore it is essential that a mechanism exists which distinguishes fully processed mRNAs from those which are not. This mechanism is controlled by RBPs and is a multi step process. The TAP/NXF1:p15 heterodimer is a key player in mRNA transport [23].

1.3.8 mRNA localization

Spatial localization of mRNAs in the cell is critical for protein production. This has been well studied in lower life forms like *S.cerevisiae* and *D.melanogaster*. For example, She2p and She3p RBPs are important for actively localizing Ash mRNA to the bud region of the daughter cell where it associates with myosin and actin [24]. Another RBP, ZBP1 is essential for localization of β -actin mRNA to the cytoplasm [25].

1.3.9 Translation

Silencing of some mRNAs to prevent their translation is common in many species. This mostly happens during development where some mRNAs remain silent to be only translated in future stages. This silencing is usually aided by binding of several RBPs, which stop the initiation factors to bind and initiate translation. An example in *C.elegans* is Gld-1 which represses pal-1 mRNA in the blastomere stages but gets reactivated in later oocyte and

embryonic stages [26]. Another example is ZBP1, an RBP which not only localizes β actin mRNA but also aids in its translation [25].

1.3.10 miRNA maturation

Micro RNAs are produced by Dicer as pre-miRNAs and are further processed to become mature miRNAs. Recently, the RBP hnRNP A1 was for the first time shown to regulate miRNA maturation binding to human pre-mir18a, the precursor of miR-18a, and to facilitate its Drosha-mediated processing [27].

1.4 RNA-binding proteins and diseases

RBPs, as seen above, control the entire gamut of post-transcriptional modifications and therefore dysregulation of these RNA and protein interactions through mutations, translocations or deletions will pose a serious threat to the normal functions of cells. Several diseases have been identified where RBPs expression is altered (See Figure 2)

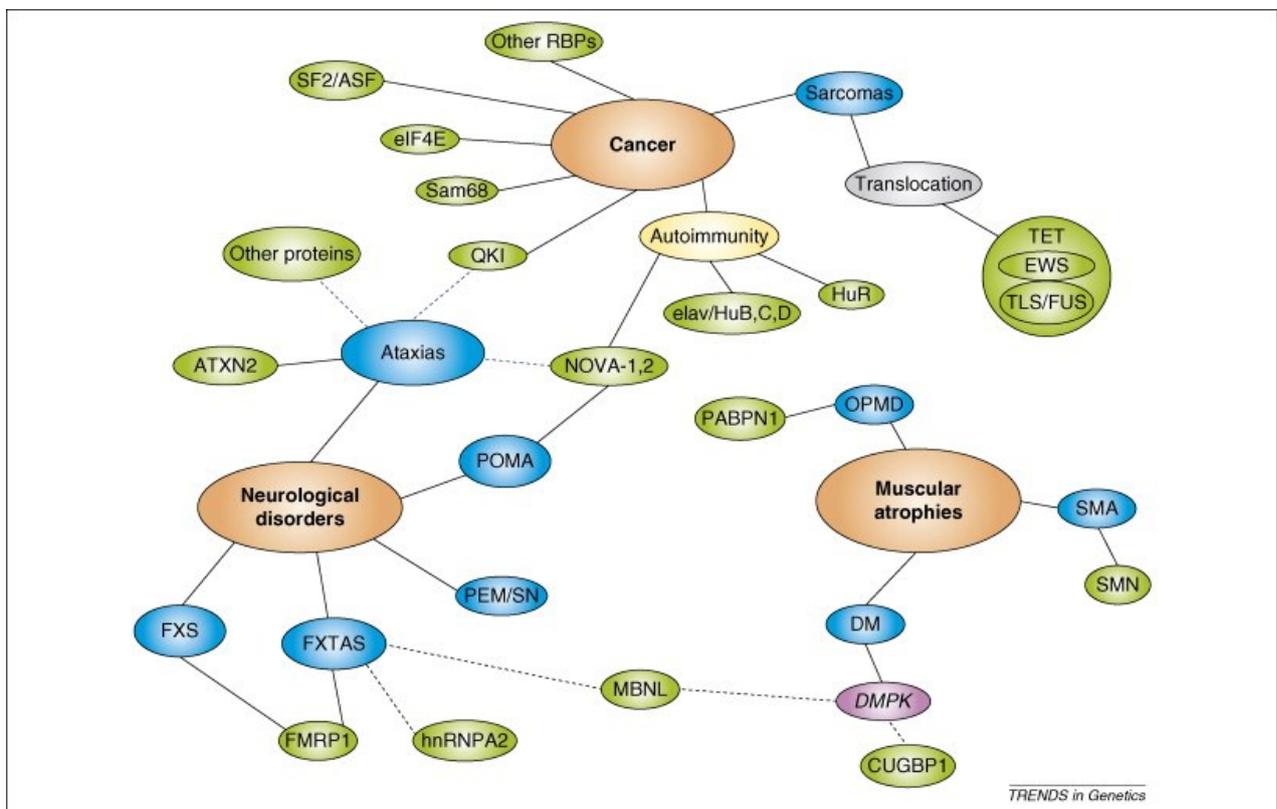


Figure 2-Network of RBPs in human diseases. [30]

1.4.1 Translocations of RBPs

The classic examples of RBPs where translocations are commonly occurring and are known to cause sarcomas are FET family of RBPs, which consists of the three proteins FUS, EWSR1 and TAF15. FUS fuses to transcription factors such as CHOP, ATF1, ERG and causes myxoid sarcomas while EWSR1 fuses to FLI1, ERG, ETV1 to cause Ewing Sarcoma family tumors (ESFTs). Recently, the third family member TAF15 was reported to fuse with ZNF384 to cause acute lymphoblastic leukemia [28]. MSI2, a musashi family RBP was also reported to fuse to HOXA9 to form in frame fusion transcript, playing a role in the disease progression of chronic myeloid leukemia [29]. These fusion proteins possess transforming properties that are sufficient to confer oncogenic transformation [30].

1.4.2 Loss of function effects by RBPs

Loss of function effect is rendered by inactivation of RBPs which is caused due to genetic mutations or antibody mediated autoimmune diseases [30]. Some examples of such diseases are fragile X syndrome which is the most common hereditary form of mental retardation. It is caused when the trinucleotide repeat expands from ~50 to more than 200 repeats in the 5' UTR regions of the FMR1 gene rendering the gene inactive [31, 32]. Spinal muscular atrophy is a common autosomal recessive neuromuscular disorder, which results in loss of motor neurons in spinal cord. This is caused due to deletions in the survival motor neuron (SMN1) gene and retention of SMN2 gene. SMN2 gene differs from SMN1 by lacking exon7 which oligomerizes less efficiently and is rapidly degraded [33]. Paraneoplastic syndromes are autoimmune diseases caused due to autoantibodies, which are generated as the reaction of the body to certain cancers. These autoantibodies bind to RBPs like Hu family of RBPs and Nova RBPs and cause paraneoplastic encephalomyelitis/sensory neuropathy (PEM/SN) and paraneoplastic opsoclonus-myoclonus ataxia (POMA) [34]. Loss of function mutations in

tRNA splicing components and aminoacyl tRNA synthetases typically cause encephalopathy, neuropathy and Charcot-Marie-Tooth disease [35].

1.4.3 Gain of function effects by RBPs

Gain-of-function is usually observed when microsatellite-expansion repeats are transcribed into mRNAs resulting in the entrapment of RBPs that associate with the repeats and interferes with the normal function of RBPs. Examples include myotonic dystrophy type 1 (DM1), in which a CUG trinucleotide expansion in the 3'-UTR of the myotonic dystrophy protein kinase (DMPK) mRNA resulting in the entrapment and gain-of-function of the RBP muscleblind-like protein 1 (MBNL1). Similarly, myotonic dystrophy type II (DM2) is caused due to CCTG repeat in intron 1 of ZNF9 protein which results in the entrapment of the RBP CUG-binding protein 1 (CUGBP1) [36]. Another example of gain of function is Fragile-X-associated tremor/ataxia syndrome (FXTAS) which is an adult-onset neurodegenerative disorder caused by a gain-of-function in which *FMRI* mRNAs indirectly mediate FXTAS by the induction of intranuclear inclusions in neurons and astrocytes [37] [38].

1.4.4 Altered expression of RBPs

Expression patterns of various RBPs are reported to cause cancers. EWS, FUS and TAF15 were shown to be upregulated in liposarcoma [39]. Similarly, MSI2 was reported to be upregulated in AML, CML and gastric tumors [40, 41]. EIF4E and SF2/ASF are also among a growing list of RBPs with altered expression in human cancer [42]. The STAR family of RBPs also shows an altered expression in cancer. For example Sam68 is overexpressed in breast and prostate cancer cells [43, 44], whereas another member QKI (Quaking) is downregulated in gliomas [45].

In general, it is however not yet clear in these cases whether the altered expression of these RBPs is a cause or a consequence of the respective cancers [30]. Further investigation is

required to establish the true nature of the relationship between altered expression of RBPs and cancer.

1.5 Experimental methods to identify the targets of RNA-binding proteins

Identifying the mRNA targets of an RBP provides the framework to study the functional and regulatory activity of that particular RBP [12]. Comprehensive analysis should therefore identify the mRNAs an RBP binds, and the proteins with which it interacts to regulate the transcripts. Many *in vitro* and *in vivo* techniques are available to study the sequence binding preferences of RBPs and their *in vivo* binding sites.

1.5.1 Microarray based immunoprecipitation

Following the initial discovery of RBPs, many techniques were developed to study their bound targets. Initially, microarray technology was applied. This approach identified enriched transcripts following RBP immunoprecipitation. This method was first applied to *S.cerevisiae* where 36 of all 600 RBPs were studied, revealing extensive mRNA and protein interactions [46]. Although microarrays can give the interaction of RNAs bound by certain proteins, they do not provide the information on the exact binding region, therefore an ideal experimental set up to get deeper insights into the function of RBP should be able to identify both the mRNA targets and also specifically give the information about the exact recognition site.

1.5.2 SELEX and RNAcompete

SELEX (systematic evolution of ligands by exponential enrichment) is a low-throughput method for *in vitro* detection of RBP sequence-binding preferences. The high affinity binding sites are selected after multiple rounds of binding to a purified protein followed by PCR amplification. The sequences are then cloned and Sanger sequenced thus identifying the short sequences, which are preferably bound by the RBPs. This method also gives insights into the

structural preferences of the binding sites of the RBP [47]. The only major disadvantage of SELEX is that only high affinity bound sequences can be identified because of the multiple rounds of purification and amplification. Recently an adapted version of SELEX, HT-SELEX was developed which encompasses a high-throughput sequencing technology, which involves a smaller number of binding reactions but millions of different RNA oligos [48].

Another *in vitro* method called RNAcompete was recently developed which is related to HT-SELEX but replaces the large, complex random initial RNA oligo pool used by HT-SELEX with a smaller, pre-designed pool that is synthesized with the help of a custom microarray. This is comparatively cheaper than HT-SELEX [49]. The disadvantage of this technique is that it fails to detect RNAs with a stable secondary structure, therefore RBPs with strict structural requirements on their binding sites are less successfully studied in this assay [50].

1.5.3 Yeast three-hybrid system

The yeast two-hybrid system was adapted to identify RNA ligands of an RBP[51]. In this assay, a protein-RNA interaction is detected by the reconstitution of a transcriptional activator using two hybrid proteins and a hybrid RNA. The RNA molecule is tethered to the promoter of a reporter gene, which by binding to a hybrid protein leads to activation of the reporter gene, allowing for selection of desired sequences and a measurement of the interaction strength [51]. The disadvantage of this technique is that it is not possible to study the interactions with a large repertoire of targets and cannot provide insights into binding specificity and binding regions. RNA library of RNAs that bound to yeast Snp1 protein was studied using this method.

1.5.4 Ribonucleoprotein Immunoprecipitation (RIP)

This technique can be efficiently applied to large scale interactions studies of RBPs. This technique uses the principle of immunoprecipitation where RNAs bound by particular RBP are pulled down using an antibody directed against the RBP of interest and the bound RNA sequences are then studied using either RNAseq in recent times or micro arrays (RIP-CHIP) [52] [53]. The only drawback of this technique is that it is hard to pinpoint the exact binding sites of a particular RBP.

1.5.5 CLIP (Crosslinking and Immunoprecipitation)

CLIP is another large scale method which uses ultra violet rays (254nm) to form permanent bonds between RNA and its bound RBPs. Although treating with UV might not be practically applicable in some cell lines and tissues, the strong covalent crosslinking allows for stringent washes required to reduce unspecific binding. This allows identifying the true binding targets thus reducing false positives. CLIP also pinpoints biologically relevant interactions through sequencing of the target transcripts. Covalent crosslinking allows one to partially digest the RNA while retaining the core element involved in protein binding such that short nucleotides of 60-100 bases can be purified, allowing both the identification of the bound species and location of the binding site [54].

Adapted versions of CLIP techniques like HITS-CLIP, which uses high throughput sequencing for target analysis [55] and i-CLIP which uses the property of reverse transcriptase to terminate at crosslink sites [56] were developed. These methods identify RBP binding sites at nucleotide resolution. The other adapted version of CLIP is called PAR-CLIP which utilises photoactivatable ribonucleoside analogues to facilitate crosslinking [57]. We used PAR-CLIP in the analysis of our RBP of interest and therefore this will be discussed in more detail.

1.6 PAR-CLIP

Although the CLIP method improved the identification of the global target spectrum of any given RBP, it still had its limitations, the main one being the low efficiency of the UV-254 nm RNA-protein crosslinking. Moreover, the exact location of the crosslink was not readily identifiable within the crosslinked fragments. Additionally, separation from signal (true RNA targets) to noise (background binding) was not easily achieved. To overcome these limitations, Hafner et al. improved the CLIP method for isolation of segments of RNA bound by RBPs which they termed PAR-CLIP.

PAR-CLIP stands for PhotoActivatable Ribonucleoside enhanced CrossLinking and ImmunoPrecipitation [57]. This method makes use of photoreactive thionucleoside analogues such as 4-thiouridine (4SU) or 6-thioguanosine (6SG), which readily incorporate into newly synthesized RNA. The recommended concentration of these analogues in the cell culture does not render any toxic effects during endogenous labelling therefore they can be directly added to the culture medium. The cells are then exposed to low energy (365 nm) UV light, which minimizes damage to the rest of the RNA and at the same time, increases the efficiency of the crosslinking. The other advantage of crosslinking inside the cells will eliminate any possibility of post-lysis re-associations of interacting partners [58]. Additionally, thionucleosides, when crosslinked, produce characteristic nucleotide conversions at the step of reverse transcription which pinpoint the exact place of protein-RNA contact. Crosslinked 4SU additionally leads to T-to-C substitutions during reverse transcription, 6SG causes G-to-A conversions. The T-to-C and G-to-A transition frequency of clusters of sequence reads is used to separate clusters derived from crosslinked RNA segments from those originating from non crosslinked RNA background. The background usually refers to error free reads matching abundant cellular RNAs like rRNA, miRNA, tRNA or other bacterial RNAs. These

background RNAs are devoid of T-to-C transitions since only the mRNAs bound by RBP contain them [59]. Due to the ability to separate signal from noise efficiently, PAR-CLIP requires less sequence reads to identify the crosslink site compared to UV-254 CLIP approaches. This advantage is well suited for characterization of RBPs binding with lower sequence specificity and many binding sites.

1.6.1 Brief overview of the procedure

Once the cells are crosslinked after UV exposure they are subjected to lysis and the RNA is partially digested to yield short RNA tags. The complex of protein of interest crosslinked to RNA is subjected to immunoprecipitation. If an antibody against endogenous protein is not available, the proteins can be cloned using a 5' or 3' tag and this tag can then be used to pull down the protein. The crosslinked protein-RNA complex is then radiolabelled, denatured and then resolved on a denaturing gel. The band corresponding to the molecular weight of the protein is cut out. This step reduces the background of unbound RNAs and unspecific interactions. Protein from the complex is digested using proteinase K and RNA is isolated using phenol chloroform isolation. Adapters which are specific to the sequencing platform are ligated to the RNA. The RNA is then converted to cDNA with primers specific to the adapters followed by PCR amplification. The cDNA library is finally analyzed by next generation sequencing (Figure 3).

Detailed protocol on how to perform the experiment step by step is given in the publication from Spitzer et al [60] and a visual protocol is also available [61].

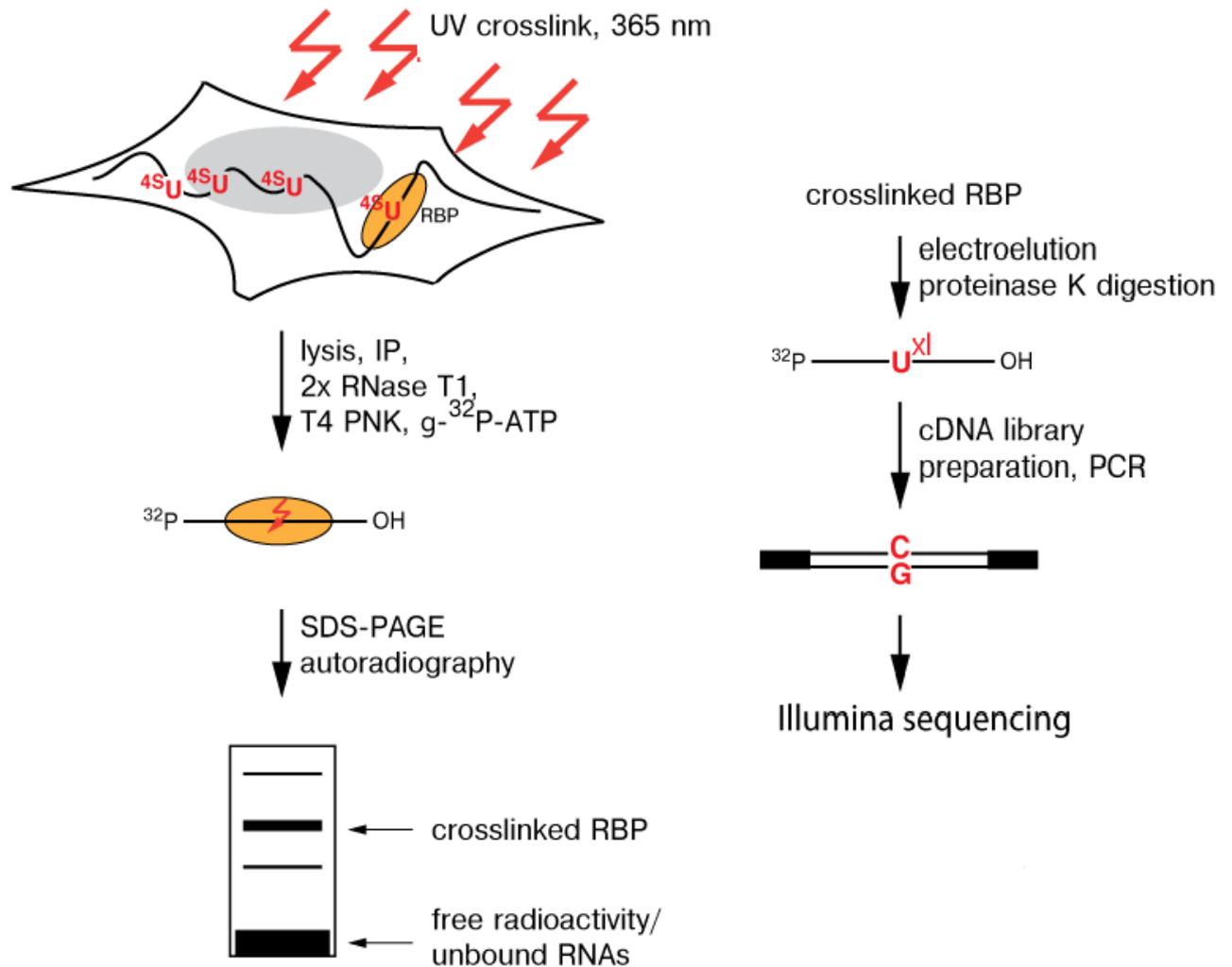


Figure 3-Schematic representation of PAR-CLIP[57]

1.6.2 Bioinformatic methods to analyze the CLIP NGS datasets

The output of the Illumina sequencing pipeline is a text based FASTQ format storing both a nucleotide sequence and its corresponding quality scores. These raw NGS reads contain adapter sequences ligated to them at both 3' and 5' ends which have to be removed from all the reads. This is done by searching for matches between partial or the full adapter sequences and parts of a read [62]. Currently, read clippers like cutadapt or trimmomatic [63] are used. A potential problem with this approach is that it removes low quality regions which generate short reads which cannot be aligned to a reference genome. To avoid this problem, tools like SHREC [64] or Quake [65] are used, which correct errors caused by low base quality. Next,

the clipped reads are mapped to a reference genome such as the human genome sequence [66]. A common computational approach for mapping reads is to use Burrows-Wheeler Aligner (BWA) [67], Bowtie [68] and Bowtie2 [69]. These methods are altered in a way that they allow read alignment with mismatches, since PAR-CLIP reads contain T-to-C transitions. The aligned read sequences are then annotated using a software called HOMER [70], which directly annotates gene names and gene functions. This software can additionally identify the region of every read to inform the researcher whether it is part of an intron or exon or 3' or 5' UTR of a gene.

The advantage of PAR-CLIP is that it also allows one to identify the RNA Recognition Element (RRE) that is recognized by the RBP. Clustering of reads and assessing the clusters allows one to identify the RRE. This clustering is performed by hierarchical bottom-up clustering (Kaufmann et al Wiley and Sons, 2009-book). The clusters are piled up in a way such that every read of a cluster overlaps with at least one other read of the same cluster by at least a pre-specified minimum length. Next, the motif from the RRE sites is identified, for which several motif-finding methods are available. MEME suite is one method which identifies multiple RREs within a set of sequences. The other method employing the Meta-MEME model is also widely used and is built from known motifs from a motif database. Some other motif searching algorithms include cERMIT [71], mCarts [72], Mcast[73], Phylogibbs [74] and RNAcontext [75].

Please refer to Kloetgen et al [62] for a detailed review on bioinformatic methods for elucidating the RNA-protein interactions from PAR-CLIP.

1.6.3 Limitations of PAR-CLIP

The use of 4SU in PAR-CLIP makes it a uridine biased method in comparison to UV-254 nm crosslinking alone. The solution to resolve this problem would be to also use 6SG, but the G-

to-A transition events in cDNA are less frequent and may require deeper sequencing. Therefore it would be interesting to test other photoactivatable nucleotides and their respective signatures following crosslinking and reverse transcription [59]. The other limitation is the toxicity rendered by ribonucleosides on some cells and tissues which therefore limits the applicability of PAR-CLIP in living organisms. Apart from cell lines, this method has successfully been performed in the nematode *C.elegans* [76] but also has the potential to be introduced in a tissue specific way into more complex animals like *drosophila* using the so-called “TU tagging system” [77].

1.7 FET family RNA-binding proteins

Heterogeneous nuclear ribonucleoprotein particle (hnRNP) proteins are a class of RNPs which are ubiquitously expressed and are highly abundant in cells [78]. This class of proteins are predominantly localized to the nucleus but some shuttle between cytoplasm and nucleus [79]. They play a significant role in mRNA biogenesis and functions. FET family proteins are among the most important proteins in hnRNP family proteins. FET family proteins contain three different proteins FUS, EWS and TAF15; these are sometimes also referred as TET family proteins as FUS is also referred to as TLS [80]. Pulldown assays showed that FET family proteins interact with each other suggesting that they might be part of the same protein complexes [81]. FET members are RNA- and DNA-binding proteins, which interact with a multitude of transcripts and affect multiple steps in mRNA biogenesis. These proteins have attracted wide attention from the scientific community since all three are involved in genomic rearrangements causing sarcomas [82] and other cancers; additionally, mutations in FUS and TAF15 were shown to cause several neurological diseases like amyotrophic lateral sclerosis (ALS) and frontotemporal lobar dementia (FTLD) [83]

The FET protein family is highly conserved and found in multicellular organisms including plants, nematodes, insects and vertebrates, although in invertebrates only one protein is present but all vertebrates ranging from fish to mammals have three members [84]. All FET members share a common domain organization, including an N-terminal low complexity (LC) domain, three RGG domains (RGG1, RGG2 and RGG3), a zinc finger domain (ZnF), an RNA recognition motif (RRM) and a nuclear localization signal (NLS) (See Figure 4).

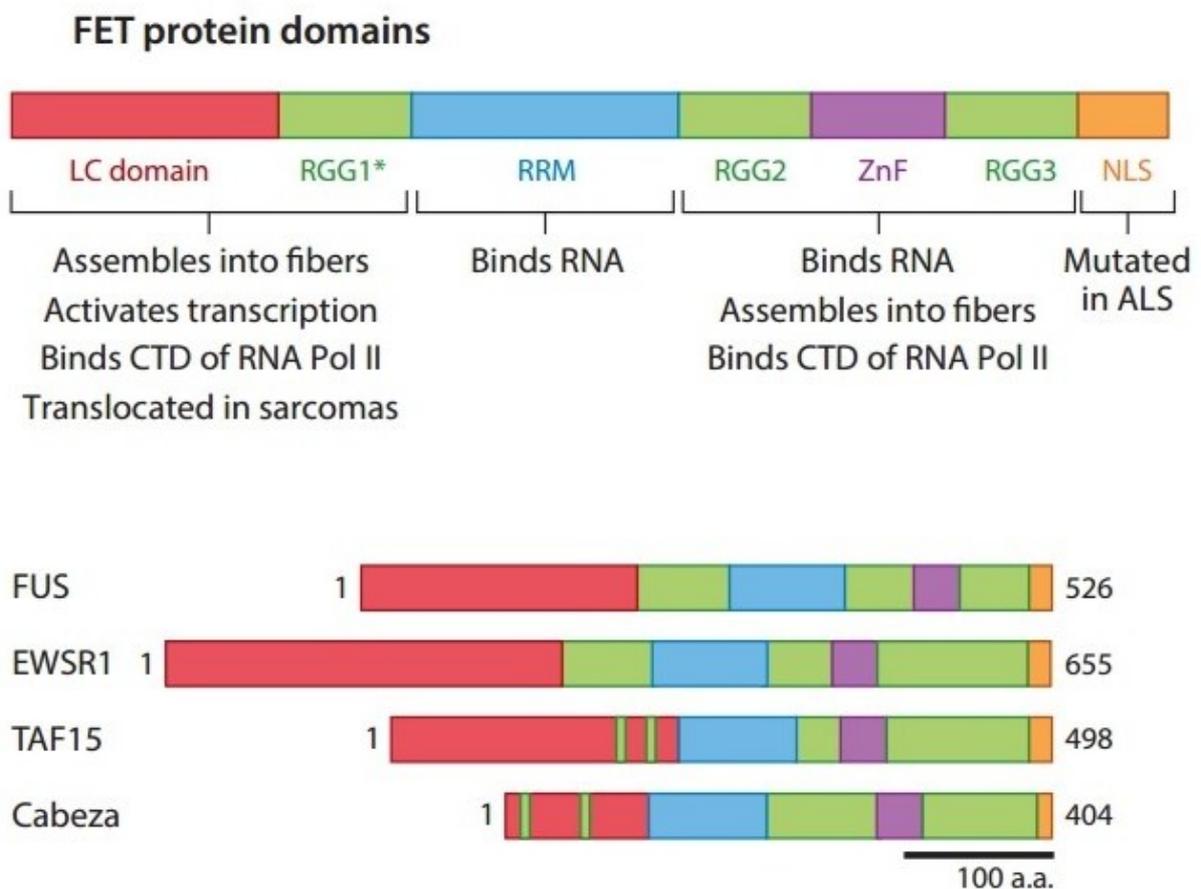


Figure 4-Protein domain organisation of FET proteins [87].

FET proteins bind to double as well as single stranded DNA through RGG domains while they bind to RNA through RRM and RGG-ZnF-RGG domains [22, 85, 86]. Studying DNA and RNA binding transcripts of FET family proteins and their interacting protein partners therefore sheds light on the complex functions regulated by these proteins.

1.7.1 Functions of FET family proteins

Since FET family proteins are RBPs, they were assumed to have roles in regulation of transcription. Evidence to confirm their role in transcription comes from three sources. The first one is that the knockdown of FET proteins had upregulated and downregulated numerous mRNAs, thus affecting (at least indirectly) transcription [83, 87, 88]. The second evidence comes from the fact that they physically interact with components involved in transcription including RNA pol II [85, 89-91], several hormone receptors including retinoid X receptor, estrogen receptor, thyroid receptor, glucocorticoid receptor [92] and NF- κ B factor p65 [93]. FUS also binds and inhibits CREB-binding protein (CBP) and p300 histone acetyltransferases. FUS binds to CCDN1 and inhibits histone acetylation and reduces transcription. Another FET member EWSR1 (gene is described as EWSR1 and proteins is called EWS) binds to transcription factors like Oct-4, CBP and HNF4 α [94, 95]. EWS was also shown to inhibit the activity of retinoic acid receptor [95].

FET family proteins role in transcription could be linked to their role in splicing and polyadenylation since these two process occur co-transcriptionally. Evidence for this comes from the fact that knockdown of FUS and EWSR1 have affected alternative splicing of many transcripts [96-98] as well as polyadenylation [91]. FUS also binds to several splicing regulating factors like SR proteins, U1snRNP and SMN proteins [99-102].

The other important post-transcriptional gene regulatory role played by FET proteins is RNA transport. Transcripts are bound by RBPs and are localized to their site of action either in the nucleus or cytoplasm or back and forth (nuclear cytoplasmic shuttling). Initial evidence into this function was established when FUS mutation in the nuclear localization signal caused ALS and FTLD [103]. Deletions or translocations of FUS, EWSR1 and TAF15 might also affect their trafficking mechanisms and therefore might lead to accumulation of proteins

either in the nucleus or cytoplasm. Not much is known about this yet, but several studies state that in some cell types or in response to environmental cues, FET protein levels are increased in the cytosol [104]. Another study states that asymmetric dimethylation of the RGG motifs, which is mediated through protein arginine methyltransferases 1 and 8 (PRMT1, PRMT8), effects the self association of intact EWS necessary for nuclear localisation [105, 106].

FET proteins are also implicated in DNA damage repair. Knockdown of FUS inhibits the recruitment of DNA repair factors histone deacetylases (HDAC1), γ HRAX, phosphorylated ATM and DNA-PK to sites of DNA damage [107].

1.8 Ewing Sarcoma break point region 1 (EWSR1)

We focussed our attention onto the second member of the FET family protein, EWS, which is –when translocated- known to cause Ewing Sarcoma. Ewing sarcoma is the second most common sarcoma in children and young adults. So far, research has focused on the resulting chimeric transcription factors, while the putative physiological function of heterozygous EWSR1 loss in these tumors has not been thoroughly investigated. We therefore focussed our study on the targets regulated by EWS and its role in post-transcriptional gene regulation.

1.8.1 Ewing Sarcoma

Ewing Sarcoma is an aggressive form of childhood cancer where tumors develop in bones or soft tissues. It occurs most frequently in children above the age of 5 years and adolescents. It is much more common in white populations and slightly predominant in males [108]. Common types of Ewing Sarcoma are Ewing Sarcoma of the bone, extraosseous Ewing Sarcoma, peripheral primitive neuroectodermal tumor, and Askin tumor (a primitive neuroectodermal tumors (PNETs) of the thoracopulmonary region), which are together referred to as Ewing Sarcoma family tumors (ESFTs). The most common translocation that

causes Ewing Sarcoma is a rearrangement (translocation) of genetic material between chromosome 22 and chromosome 11 (t(11;22)). This translocation fuses part of the EWSR1 gene on chromosome 22 with part of another gene on chromosome 11 called FLI1, creating an EWSR1/FLI1 fusion gene (Figure 5) [109]. This mutation accounts for almost 85% of all Ewing Sarcoma cases. The translocations that cause these tumors are acquired during a person's lifetime and are present only in the tumor cells.

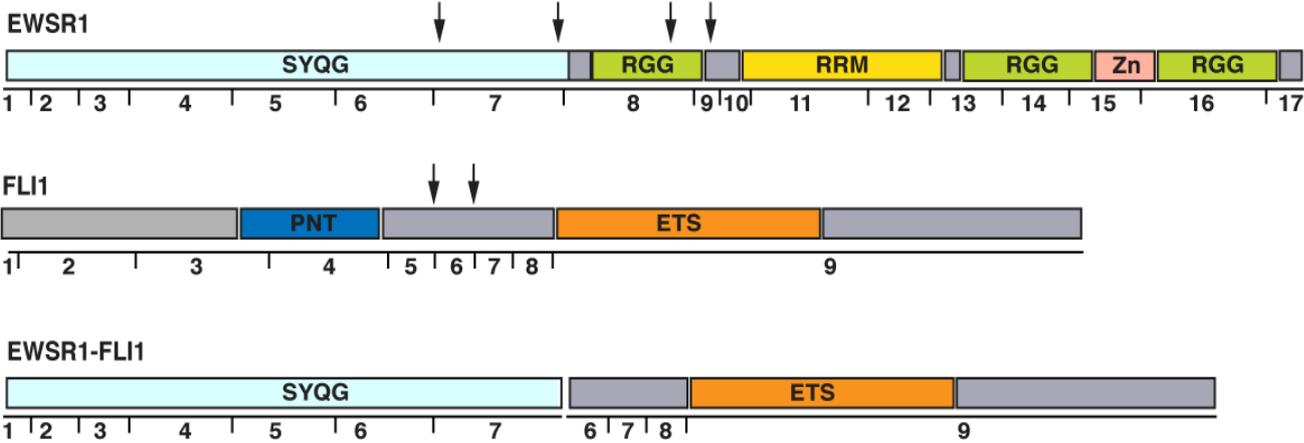


Figure 5-Protein domain organization of EWSR1 and FLI1.

The black vertical arrows indicate common breakpoints in Ewing Sarcoma. Numbers correspond to exons and a typical EWSR1-FLI1 fusion protein is also shown.

Other less common types of translocations include the fusion between EWSR1 and WT1, which also causes desmoplastic small round cell tumor, a type of soft tissue sarcoma that often occurs in the abdomen [110]. The other less common fusion is between EWSR1 and ATF1 which causes soft tissue clear cell sarcoma, where tumors usually develop in the tendons, especially in the knees, feet, and ankles [111]. Fusions in EWSR1 gene cause other sarcomas which are unrelated to ESFTs. Examples include EWSR1/NR4A3 fusion gene which causes extraskeletal myxoid chondrosarcoma, which is a rare type of soft tissue tumor that usually occurs in the lower body, such as the thighs or gluteal region. The least common

translocation is between EWSR1 and CHOP fusion gene, which is sometimes found in myxoid liposarcomas, which -as the name suggests- occur in fatty tissue in many parts of the body [112, 113].

1.8.2 Functions of EWS

1.8.2.1 Physiological role of EWS

Wildtype EWS, although implicated in multiple carcinomas and sarcomas- has not been thoroughly studied so far [114]. First insights into its functions were revealed by the knockout mouse model (EWS^{-/-}). Mice born with normal mendelian ratio but were smaller in size compared to their littermates. They also displayed a high postnatal mortality (90%) prior to weaning. These knockout mice revealed a role for EWS in pre-B lymphocyte development and meiosis [115, 116]. Knockdown of EWS in mouse embryonic fibroblasts revealed that EWS interacts with Lamin A/C [115]. EWS loss also resulted in abnormal changes in the hematopoietic compartment where the cells upon loss of EWS moved from their quiescent stage to cell cycle division state [117]. Immunological relevance of EWS was revealed when the EWS null mice displayed a specific decrease in the production of IFN-gamma and interleukin- (IL-) 2 production [118]. Absence of EWS renders the mice to be hypersensitive to ionizing radiation, which further leads to DNA breaks thus showing a role of EWS in cellular senescence [115].

1.8.2.2 Role of EWS in transcriptional gene regulation

EWS contains several domains capable of binding independently to nucleic acid sequences such as the RNA-binding domains. As discussed previously, all FET family proteins have roles in PTGR. EWS also plays a role in PTGR by associating with several transcription

factors like RNA polymerase II [90] and with subunits of Rpb3, Rpb4 and Rbp7 [119]. Additionally by directly binding to transcription factors, EWS regulates PTGR by indirectly binding to transcriptional activators and repressors. For example, EWS binds to various transcription factors containing the POU homeodomain which exert critical developmental functions. EWS also binds to OCT4, a transcriptional activator, which is essential to maintain an undifferentiated totipotent state of embryonic stem and germ cells [95]. EWS has also been shown to interact with the histone acetyl transferase CREB binding protein (CBP) and the p300 transcription activator, thus cotransactivating several promoters in a cell-type specific manner [94, 120]. EWS interacts with the transcriptional repressor ZFM1 [121]. Translocations in Ewing Sarcoma patients give rise to fusion transcripts (EWS-FLI1) which also act as transcriptional activators, triggering unwanted expression of several genes ultimately leading to the development of sarcomas.

1.8.2.3 EWS and splicing

EWS plays a critical role in splicing mechanism since the transcripts which were downregulated upon EWS knockdown showed defective alternate splicing. A yeast two-hybrid screen revealed that EWS interacts with U1C which is one of the three U1 small nuclear protein component of the U1snRNP [122], which binds to the 5' splice site on pre-mRNA. More recently, two studies have revealed that alternative splicing of the two cancer relevant genes MDM2 [123] and DNA damage response gene DDR [98] is regulated by EWS. EWS was also shown to interact with YB1, which functions as a shuttling splicing activator [100]. Human Cyclin D1 (CCND1) alternative splicing as well as transcriptional regulation was shown to be controlled by EWS [124].

EWS is therefore involved in the regulation of a wide range of cellular processes to ensure genomic integrity and proper cellular functions. EWS fuses to different DNA-binding proteins

like FL1, ERG, ETV1, ATF1 resulting in chimeric proteins. These chimeric transcription factors are expressed at high levels in the cells, and are generally believed to be the main cause of malignant transformation. However, recent studies suggest that factors unrelated to these chimeric transcription factors could also contribute to the oncogenic transformation in Ewing Sarcoma [125]. These other factors could be attributed to the fact that loss of one EWSR1 copy in Ewing Sarcoma causes a haploinsufficient state of wild-type EWS, with consequences for its RNA-binding activity and RNA-binding partners [126].

2. Results and discussion

First publication: The cell cycle regulator CCDC6 is a key target of RNA-binding protein EWS.

Sujitha Duggimpudi, Erik Larsson, Schafiq Nabhani, Arndt Borkhardt, Jessica I Hoell

Published in “PLoS One 10.1371, 2015

“PLoS One”, Impact factor- 3.5; First author

Second publication: RNA targets of wild type and mutant FET family proteins

Jessica I Hoell, Larsson E, Runge S, Nusbaum J, **Sujitha Duggimpudi**, Farazi T, Hafner M, Borkhardt A, Sander C, Tuschl T

Published in Nature structural and molecular biology, 2011,18:1428-31

Impact factor 11.63; Co-author

2.1 Publication I: The cell cycle regulator CCDC6 is a key target of RNA-binding protein EWS.

2.1.1 Abstract

Genetic translocation of EWSR1 to ETS transcription factor coding region is considered as primary cause for Ewing Sarcoma. Previous studies focused on the biology of chimeric transcription factors formed due to this translocation. However, the physiological consequences of heterozygous EWSR1 loss in these tumors have largely remained elusive. Previously, we have identified various mRNAs bound to EWS using PAR-CLIP. In this study, we demonstrate CCDC6, a known cell cycle regulator protein, as a novel target regulated by EWS. siRNA mediated down regulation of EWS caused an elevated apoptosis in cells in a CCDC6-dependant manner. This effect was rescued upon re-expression of CCDC6. This study provides evidence for a novel functional link through which wild-type EWS operates in a target-dependant manner in Ewing Sarcoma.

2.1.2 Results:

We have previously described the transcriptome-wide targets of FET family proteins and narrowed down our focus on to the EWSR1 targets which are not only bound but also regulated by the protein. We performed microarrays after knocking down EWS in HEK 293 T cells and analyzed the targets which were significantly down- or upregulated. In total, we found 116 regulated genes with a corrected p-value of <0.05 (32 at <0.01), which had more than one PAR-CLIP cluster and whose expression level changed (either up or down) by at least 50% upon knockdown of EWS compared to controls. Analysis of the PAR-CLIP cluster binding localization revealed that these 116 regulated mRNAs were preferentially bound in the 3'UTR (60%) and had less intronic clusters compared to those of all bound 4488 mRNAs (40%). For our initial analysis we focused on the four highly regulated targets *CCDC6* (log

fold change of -1.10), *MDM2* (-0.53), *FGF9* (-0.66) and *CBFB* (-0.58) which were all downregulated upon knockdown of EWS.

We next focused on *CCDC6*, which is known to have several cell cycle associated functions including DNA damage response [127, 128], cell cycle regulation by controlling the intra-S-Phase and G2/M checkpoints. Knockdown of *CCDC6* also showed increased apoptosis and decreased proliferation. As a first step, we wanted to explore whether EWS stabilizes *CCDC6* by binding to it. We performed mRNA stability assay by treating the cells with Actinomycin D to inhibit de novo RNA synthesis and found that the half-life of *CCDC6* was ~4.1 h in control cells, whereas in cells treated with EWS siRNA the half-life decreased to ~3.2 h. Together, these data suggest that EWS stabilizes the *CCDC6* mRNA. We also performed luciferase assays to confirm the regulation of *CCDC6* mRNA by EWS protein. For our further experiments, we chose MHH-ES-1, a Ewing Sarcoma cell line which carries the most frequently occurring translocation EWS-FLI. We knocked down EWS in this cell line using siRNA which is designed to target the C-terminal region of EWSR1 which is absent in EWS-FLI1 fusion transcript in MHH-ES-1. qRT-PCR results showed a clear downregulation of *CCDC6* mRNA levels upon knockdown of EWS. We repeated this in HEK293 T cells and observed the same effect. Also, Western blot showed decreased expression of *CCDC6* protein upon EWS knockdown in MHH-ES-1 and also in HEK293T cells thus showing that the regulation of *CCDC6* by EWS extends to the protein level.

It was previously shown that FET proteins show redundancy and that knockdown of EWS upregulates its two family members FUS and TAF15 in HEK293 cells as well as in liposarcoma cell lines [39]. To address whether this mechanism might also lessen the effects of EWS haploinsufficiency in Ewing Sarcoma, we measured FUS and TAF15 mRNA levels upon EWS knockdown in MHH-ES-1 cells. However, there was no increase in mRNA levels

for either of the two genes. This indicates that, probably due to a different cellular context, in Ewing Sarcoma there is no rescue mechanism for loss of EWS expression.

EWS and CCDC6 were previously reported to regulate cell cycle processes like apoptosis, proliferation and cell cycle. Therefore we assessed apoptosis rates following downregulation of EWS by Annexin V FITC and PI double staining of EWS knockdown and control HEK293T cells. We recorded an increase in the percentage of apoptotic cells in EWS siRNA treated cells (63% living cells, 27% apoptotic and 10% necrotic cells) compared to controls (74% living, 13% apoptotic and 13% necrotic cells) while the % necrotic cells remained unchanged. We next performed gain of function experiments to see if upregulation of CCDC6 after EWS knockdown will rescue this phenotype. To do this, we co-transfected the cells with EWS siRNA and CCDC6 overexpression vector and found that the apoptotic rate was 14% less in CCDC6 overexpressed cells compared to empty vector, confirming that overexpression of CCDC6 upon EWS knockdown indeed rescued the observed phenotype. Apoptotic signals are coupled to growth regulatory processes such as proliferation, cell cycle arrest, and cellular differentiation. Therefore, to examine if the observed apoptosis affected the proliferation rate of the cells, we measured the proliferation of EWS knockdown and control cells using CCK8 assay. We indeed noticed that the proliferation rate remarkably dropped upon downregulation of EWS compared to controls. Given that EWS regulates the expression of CCDC6, that EWS downregulation induces apoptosis, and that CCDC6 has been implicated in apoptosis coupled to S and G2/M phase cell cycle defects, we next tested cell cycle progression of the cells by propidium iodide staining 48 hours after transfection. We observed a decrease in the number of cells in S and G2/M phase upon EWS knockdown and an increased cell number with a sub G0/G1 DNA content typical of apoptotic cells. This indicated increased apoptosis of cells that had passed the G1 checkpoint. Taken together our results demonstrate that EWS controls apoptosis and cell cycle progression by regulating one of its key mRNA targets CCDC6. EWS

therefore regulates a wide range of cellular processes to ensure genome integrity and cellular homeostasis. Future studies will show whether EWS and/or its target mRNAs will offer new therapeutic approaches.

2.1.3 Discussion

Ewing Sarcoma represents the second most common bone and soft tissue malignancy in adolescents and young adults. Genetically, almost all of these tumors are characterized by translocations in EWSR1 gene. EWSR1 is a member of the FET family of nuclear proteins (FUS, EWSR1 and TAF15), which are ubiquitously expressed at high levels in all cell types and whose functions remain largely unknown. Ewing Sarcoma is characterized by translocation in EWSR1 gene whereby the N-terminal portion of the RNA-binding protein (RBP) EWSR1 is joined to a DNA-binding protein belonging to the so-called ETS family (e.g. FLI1). In around 85% of Ewing Sarcomas, the translocation *EWSR1-FLI1* / t(11;22) can be found, rarer examples include *EWSR1-ERG* / t(21;22) (10%) and *EWSR1-ETV1* / t(7;22) (<5%). Given that the resulting transcription factor is expressed at high levels in the cells, it was generally believed to be the main cause of malignant transformation and research focused almost exclusively on EWSR1-FLI1.

Meanwhile, little attention was paid to the “lost” RNA-binding capabilities of the translocated C-terminal portion of EWSR1 and a potential dysregulation of its targets upon translocation, despite the fact that it had been known for some time that -in addition to the DNA-binding-dependent activity of the fusion proteins- other, DNA-binding-independent factors contribute to the oncogenic transformation in Ewing Sarcoma. We applied PAR-CLIP to EWSR1 and found the mRNAs of 4488 genes to be directly targeted by this protein. We then performed knockdown of EWSR1 and identified 116 genes which had altered expression levels and are therefore regulated by this protein. The regulated genes included tumor-associated genes such

as LIN28B which regulates let-7 miRNA processing [129], PURB which controls both DNA replication and transcription and its deletion has been associated with acute myeloid leukemia [130]. CCDC6 which has roles in cell cycle and was also predicted to be a potential tumor suppressor [131], CDCA4 which regulates E2F-dependent transcriptional activation and cell proliferation, mainly through the E2F/retinoblastoma pathway [132], MDM2 which is an E3 ubiquitin protein ligase and repressor of transcription factor p53 [133], and FGF9 which has roles in tissue repair, tumor growth and invasion. Downregulation of these genes could further have an effect on genes they regulate and signaling pathways they are involved in. For example, MDM2 is a protein which inhibits activation of p53 [133], any downregulation in the levels of MDM2 disturbs the functional interactions of these networks which further affects tissue homeostasis [134] such as the notch induced p53 activity [135, 136]. Another target, FGF9, which was reported to be downregulated and mutated in several carcinomas plays an important role in activating FGFR signaling. This signaling pathway plays a role in differentiation and inhibits growth thus acting as a tumor suppressor. Therefore its downregulation upon EWS knockdown could repress its tumor suppressor activity [137, 138]. Similarly, CBFβ is also frequently mutated in AML and is known to play a role in hematopoietic development [139]. The genes regulated by CBFβ are important in chromatin deacetylation and promoter methylation and thus playing a role in gene transcriptional activation or repression. Therefore its downregulation could disrupt the normal pathways in hematopoiesis and also effect gene regulation. Future experiments are necessary to study in detail the interactions of EWS with these genes and the complex pathways they regulate.

Several of the EWS up-regulated targets and down-regulated targets had roles in cell cycle and G1/S transition like SKP2 (p45), BCAT1, CCDC6, CUL2, PPP1CB, RHOA etc. Genes with roles in proliferation and cell death included DNAJA2, MDM2, NOP2, FGF9, ID2 etc. Others like AEN, APTX, TOP2A and UBE3A had roles in DNA damage response. Functional

annotation using DAVID analysis [140, 141] showed that several genes are associated with diverse functions ranging from response to DNA damage, response to cellular stress, cell cycle process, translational initiation and cellular differentiation. Wild type EWS therefore seems to be essential gene controlling complex process for proper functioning of cell. Our focus then shifted to CCDC6, Coiled-Coil Domain Containing 6 gene, since it showed high fold change in the microarray analysis and a high number of PAR-CLIP clusters. CCDC6 is a ubiquitously expressed protein which is downregulated in several cancers and predicted to be a tumor suppressor [131]. It is known to be frequently rearranged with RET protein in papillary thyroid carcinomas [142, 143] and fuses to platelet derived growth factor receptor beta gene causing atypical chronic myeloid leukemia [144]. CCDC6 is also known to have several cell cycle associated functions including DNA damage response [127, 128], cell cycle regulation by controlling the intra-S-Phase and G2/M checkpoints. Knockdown of CCDC6 also showed increased apoptosis and decreased proliferation [145, 146]. In the current study we sought to check if CCDC6 is regulated by EWS and that loss of EWS effects CCDC6 and the downstream functional roles of CCDC6. Using Actinomycin D treatment we confirmed the regulation of CCDC6 by EWS by looking at the mRNA stability of the former after downregulating the latter. We further confirmed this regulation by employing luciferase assay technique. Next, we chose MHH-ES-1 cell line which is an Ewing Sarcoma cell line and carries EWS-FLI1 translocation which accounts for 85% of Ewing Sarcoma cases. Together with HEK 293T cell line and MHH-ES-1 cell line we sought to look at the interplay between both these proteins. Protein and mRNA levels of CCDC6 decreased upon EWS knockdown which emphasizes the point that steady state level of mRNA and proteins and under the control of the RNA binding proteins that bind to them and stabilize. This regulation of CCDC6 by EWS might further have an impact on several other downstream genes that CCDC6 itself regulates, we tested one of these published targets, namely FBXW7, which is

required for DNA damage response [147], and indeed, qRT-PCR on FBXW7 following knockdown of EWS showed reduced RNA levels (Fig 2A). Since FBXW7 is mutated in several cancers and shown to be a tumor suppressor, the pathway through which all these proteins inter-regulate might further play a role in sarcomagenesis [148]. Further studies are required to elucidate the role of FBXW7 in Ewing Sarcoma context. Taken together, our findings suggest that EWS regulates the expression levels of CCDC6 by stabilizing it and further exerting its effect on the expression of its downstream targets.

Further we performed gene set enrichment analysis (GSEA) and DAVID analysis and revealed that several targets regulated by EWS play a role in regulating cell proliferation and cell cycle. Also, EWS was reported to maintain mitotic integrity and proneural cell survival in zebra fish by regulating aurora B [149]. Earlier studies reported that CCDC6 silencing increased apoptosis, decreased cell proliferation and affected cell cycle division by controlling the intra S phase and G2/M checkpoints [145]. TAF15, another member of the FET family of proteins was also shown to regulate cell proliferation and apoptosis *in vivo*. With this data in the background we focused on the effect of the regulation between EWS and CCDC6 on Cell survival pathways like apoptosis, proliferation and cell cycle. Knockdown of EWS increased the apoptosis rate which might be due to its target CCDC6 downregulation. Therefore we tried to rescue the effect by putting back CCDC6 over expression plasmid, which drastically improved the viability of cells with significant decrease in the apoptotic rate. Apoptotic signals are coupled to growth regulatory processes such as proliferation, cell cycle arrest, and cellular differentiation therefore we next looked at the proliferation which as expected decreased upon EWS knockdown. Given that EWS regulates the expression of CCDC6, and that EWS downregulation induces apoptosis, and CCDC6 has been implicated in apoptosis coupled to S and G2/M phase cell cycle defects, we next tested cell cycle progression of the cells by propidium iodide staining. The number of cells decreased in S and G2/M phase upon

EWS knockdown and an increased cell number with a sub G0/G1 DNA content typical of apoptotic cells. This indicated increased apoptosis of cells that had passed the G1 checkpoint. Higher apoptotic rate might have likely affected the duration of cell cycle or the doubling time of the cells. The observed decrease of cell number in S and G2/M phase might be due to the fact that CCDC6 regulates cell cycle checkpoints like 14-3-3 σ and CDC25C [146] which are important for S phase duration, transition into G2 phase and activation of mitosis. Cell cycle checkpoints safeguard the cells from accumulating genetic errors and any deregulation will prepare the ground for increased mutagenesis and onset of several cancers. Therefore the aberrant entrance of cells into next phases by skipping the checkpoint controls regulated by CCDC6 might further triggers unwanted mutations leading to genetic aberrations.

To summarise our results demonstrate that EWS controls apoptosis and cell cycle progression by regulating one of its key mRNA target CCDC6. EWS therefore regulates a wide range of cellular processes to ensure genome integrity and cellular homeostasis. Recent studies have unveiled new roles of EWS in gene regulation and RNA metabolism. This suggests that highly complex roles are played by EWS and a thorough elucidation of the entire mRNA target network of native EWS is essential in understanding the molecular and cellular biology of Ewing Sarcoma. Since EWS regulates mRNA targets which hold promising roles in the proliferation of cells and its viability, its haploinsufficiency in Ewing Sarcoma could affect these processes and further trigger sarcomagenesis. Further studies on other EWS targets involved in these processes could unravel the underlying mechanisms of Ewing Sarcoma.

We therefore propose that future research on Ewing Sarcoma should not only focus on the translocated EWSR1 allele but also on the functions lost by the wild type protein. The combined knowledge from both areas of research will give us a better understanding of

sarcomagenesis and will further help us to establish new therapeutic approaches for this aggressive sarcoma of children and young adults.

2.2 Publication II: RNA targets of wild type and mutant FET family proteins

2.2.1 Abstract

The RNA-binding proteins FUS, EWSR1 and TAF15 form the FET protein family. FET translocations are diagnostic of certain cancers and FUS mutations were recently shown to cause amyotrophic lateral sclerosis (ALS). We defined the RNA-binding sites and consensus RNA recognition element (RRE) of wild-type FET proteins and two ALS-causing FUS mutants by PAR-CLIP. The RRE was confirmed biochemically, consisting of a stem-loop opened by a non-Watson-Crick U-pyrimidine base pair and an A immediately 3' to the mispaired U in the loop region. Nuclear localized wild-type FET proteins predominantly bind intronic RNA segments, in contrast to the cytoplasmic localized FUS mutants, which preferably bind 3' UTRs. Knockdown of wild-type FET proteins, despite their proposed function, did not significantly impact splice regulation nor alter mRNA stability in a binding-site-dependent manner. Instead, their abundant, ubiquitous expression suggests a more general function in supporting basic nuclear RNA processing functions.

2.2.2 Results

Cell lines expressing FLAG/HA-tagged wild-type FET and mutant FUS proteins were grown for 12 to 16 h in 4-thiouridine-supplemented medium to allow for its incorporation into nascent RNA transcripts, followed by PAR-CLIP [57, 60]. Crosslinked RNAs were recovered from SDS-PAGE-purified FET protein immunoprecipitates, converted into cDNA libraries, and then Solexa-sequenced. We obtained on average 17,500 mRNA-derived clusters with crosslinking evidence. Wild-type FET proteins crosslinked to predominantly intronic regions consistent with their nuclear localization (FUS: 76% / 74% [stable / inducible], EWS: 50% / 52%, TAF15: 47% / 58% intronic). In contrast, mutant FUS proteins crosslinked predominantly to 3' UTR (58 / 62%), indicating that the RNA-binding properties of mutant

FUS were not impaired, but result in a drastically altered target RNA distribution. Overall, transcripts bound by wild-type FET proteins were often bound at multiple positions; on average one FET CC every 5581 nucleotides. The top 50 transcripts ranked according to number of CCs had between 145 and 527 FET CCs with an average density of one CC every 1473 nucleotides. FUS targeted 8032 genes, EWSR1 targeted 7797 genes and TAF15 targeted 5591 genes. Overall, 9701 genes were targeted by at least one FET protein. Interestingly, FET proteins also bound to their own mRNAs, *FUS* had 35 CCs (66% intronic), *EWSR1* had 61 CCs (75% intronic) and *TAF15* had 90 CCs (72% intronic). Mutant FUS had 7 CCs on *FUS* (43% exonic), 8 CCs on *EWSR1* (75% exonic), and 15 CCs on *TAF15* (53% exonic). In summary, wild-type FET and mutant FUS had the expected intronic / exonic distribution when binding to their own mRNAs.

Next we investigated which transcripts were preferentially bound by FET proteins. Gene Ontology (GO) analysis of these 200 FET targets revealed an enrichment of DNA repair categories. Examples of mRNAs of DNA repair genes included *fanconi anemia*, *complementation group L (FANCL)* or *polymerase (DNA directed), alpha 1, catalytic subunit (POLA1)*. The top 200 genes that were preferably bound by wild-type FUS were enriched for DNA repair-related GO categories, whereas genes that were preferentially bound by the mutants were enriched for categories such as “membrane” and “prenylated protein catabolism”. We then tested whether ALS implicated genes were also affected by the differential binding between wild-type and mutant FUS. Only seven genes have been repeatedly described as causing familial ALS, although the list of genes mentioned in case reports only is larger. Interestingly we found endoplasmic reticulum and ubiquitin-proteasome-related target gene categories to be over represented among transcripts uniquely targeted by mutant FUS.

Next we elucidated the RRE of individual FET family members. Use of standard bioinformatic tools did not return a significant RRE motif for any of the FET family protein, indicating that RNA structure may play a role in RNA recognition. Searching for secondary structure within the CCs, we found that 70% of the above-mentioned CCs contained a conventional stem-loop structure. The loop opens with a U-U or U-C non-Watson-Crick base pair, and the invariant U at the 5' end of the loop is followed by an A in the loop. T-to-C changes were identified for Us placed in stem as well as the loop regions, suggesting that all Us are positioned in close proximity to aromatic amino acid side chains of the RBPs and are amenable to crosslinking. To confirm the RRE, we tested trinucleotide-repeat containing oligoribonucleotides (length of 36 nucleotides) as well as one crosslinked stem-loop sequence mapped to *SON* by electrophoretic mobility shift assays (EMSAs).

To examine the function of FET proteins in HEK293 cells, we knocked down each FET gene individually using siRNAs and monitored the differences in mRNA profiles by Affymetrix GeneChip U133 Plus 2.0 arrays. We identified numerous genes with significantly altered transcript levels (2085 genes for FUS, 5255 genes for EWSR1 and 628 genes for TAF15). Interestingly, not only did FET proteins bind to their own mRNAs as discussed earlier, silencing of a FET family gene generally upregulated the remaining FET family members. Several studies suggested that FET proteins play a role in mRNA splicing. To address if these binding events translated into splice regulatory functions, we investigated the positional distribution of CCs in relation to splice sites. Our analysis revealed an increased frequency of CCs near 3' splice sites, but not near 5' splice sites. However, compared to all CCs, the actual number of CCs localized 50 nt upstream to a 3' intron-5' exon junction only represented a small fraction of all CCs; 1.4% of FUS, 1.1% of EWSR1, and 2.7% of TAF15 CCs.

To further investigate the protein interactions of FET protein family members, we immunoprecipitated and analyzed FET-associated proteins from RNase-treated total cell lysates of HEK293 cells stably expressing FLAG/HA-tagged FET-proteins by mass spectrometry. The FLAG/HA-FUS immunoprecipitates contained TAF15 and the FLAG/HA-EWSR1 immunoprecipitates contained TAF15 and FUS as the most enriched identified proteins. Neither FUS nor EWSR1 were identified in FLAG/HA-TAF15 immunoprecipitates. To identify proteins that selectively interact with wild-type and mutant FUS protein, we examined mutant and wild-type FLAG/HA-FUS immunoprecipitates. We identified eleven proteins bound to wild-type FUS and 22 proteins bound to mutant FUS.

2.2.3 Discussion

In this study we used 4-thiouridine PAR-CLIP to identify the transcriptome-wide RNA targets of wild-type, nuclear FET proteins and ALS-causing cytoplasmic FUS mutants. Similar to previously studied RBPs and miRNPs, we identified thousands of RNA-binding sites and their exonic versus intronic distribution reflected the cellular localization of wild-type FET and mutant FUS proteins. Mutations of FUS did neither alter its general RNA-binding properties nor its RG/RGG dimethylation status. Analysis of the binding sites revealed a consensus RRE for the FET proteins which was confirmed biochemically using synthetic RNA and recombinant protein.

2.2.3.1 The FET family RRE

For both wild-type and mutant FET proteins the majority of CCs (70%) contained the RRE defined by a stem-loop where the first unpaired nucleotides emerge as 5' UA, followed by several residues devoid of G, and a pyrimidine residue closing the loop by a non-Watson-Crick U-U or U-C base pair involving the U of the conserved UA dinucleotide. The

importance of the stem for FUS protein binding was confirmed biochemically via EMSA using classical disruption and restoration of the stem. The sequence of the stem did not appear to be critical based on sequence analysis of stems predicted in CCs. Somewhat unexpectedly, when testing trinucleotide repeat sequences as RNA ligands, (AUU)₁₂ but not (GGU)₁₂ bound to recombinant FUS. AUU repeats can adopt the structure and sequence of the identified RRE, in contrast to GGU repeats. This was a particularly surprising finding in light of reports proposing GGUG as a consensus element in RNA ligands identified by *in vitro* selection for binding to recombinant FUS [150]. The reported binding constants for the artificial G-rich ligands were similar to both our natural ligand and the AUU repeat, and illustrate the need for methods directly identifying RNA binding sites in their cellular context. It is also unclear in as much the unmethylated state of RG/RGG domains in bacterially expressed FUS [150] affected its RNA-binding properties; in our studies we used baculovirus-expressed FUS, which was determined to be at least partially dimethylated.

The ability of RRM-containing proteins to bind to RNA stem-loops had been first reported in crystal structures of SNRPA/U1A protein and SNRPB2/U2B'' binding to stem-loop II of U1 snRNA [151] and stem-loop IV of U2 snRNA [152], respectively. The SNRPB2 protein recognizes a similar sequence motif involving a U-U mismatch and the UA dinucleotide step. Interestingly, SNRPB2 protein requires SNRNPA1, an additional U2 snRNA binding protein for its recruitment, possibly a mechanism to avoid interference by FET protein family members in U2 snRNP assembly [152].

2.2.3.2 FET family protein subcellular localization

DNA sequencing of ALS patients has so far revealed mutations at 24 amino acid positions in FUS, most of which are located in the C-terminal 17 amino acids; the others were located upstream, in the RG-rich regions [153]. ALS mutations in the C-terminus cause a striking

cytoplasmic relocation of the otherwise nuclear wild-type protein [154, 155]. It was recently proposed that the C-terminal 13 amino acids of FUS represent the nuclear localization signal for karyopherin-dependent nuclear import by TNPO1 and 2 [156]. The same region was also proposed to control EWSR1 nuclear localization signals [157], with additional contributions from the Zinc finger and RG/RGG domains [158]. Interestingly, placement of some of the same mutations into the FUS paralogous position of EWSR1 did not noticeably alter EWSR1 localization. This may explain why mutations of EWSR1 have not been identified in ALS patients, despite the fact that EWSR1 is as strongly expressed as FUS and is also frequently translocated in various tumors [159].

Arginine dimethylation of RG/RGG domains was shown to play a role for the localization of some proteins (reviewed in [160]). However, we found no alterations in the arginine methylation patterns of mutant versus wild-type FUS, thereby ruling out that the C-terminal domain controls RG/RGG dimethylation status. Additionally, we analyzed the composition of nuclease-treated immunoprecipitates of mutant and wild-type FUS using proteomics but were unable to identify proteins that might dictate FUS localization.

2.2.3.3 Functions of nuclear wild-type FET proteins

Members of the FET protein family are ubiquitously expressed. FUS knockout mice showed defective B-lymphocyte development and activation, perinatal death, and an increased radiation sensitivity leading to chromosomal instability, suggesting a role for FUS in genome maintenance after UV damage [116, 161]. A similar phenotype was observed for EWSR1 knockout mice [115], which may not be surprising given the structural similarities of the two proteins and their target overlap. Consistent with the increased UV-sensitivity in the knockout animals, we found that primary transcripts of DNA repair pathways genes were enriched among specific targets of nuclear FET proteins. It is also conceivable that FET proteins have

protective or repair functions, which are only revealed under conditions of DNA damage or other cellular stress situations, and that loss of this protective function upon mutation of FUS contributes to the increased death rate of motor neurons in ALS patients [162].

siRNA knockdown analysis of FET members or overexpression of mutant FUS did not reveal evidence to support a postulated function in splicing regulation [99, 163]. These initial observations might be explained by the high abundance of FUS and a general hnRNP-like function such as pre-organizing mRNAs for facilitating recognition of RNA by more specific RBPs like splice factors [164]. Overall, the changes in mRNA stability upon knockdown were small, but interestingly, knockdown of one FET mRNA was compensated by increased mRNA levels of the other two. The FET protein family therefore represents an extremely abundant nuclear protein family and, with its many target sites and transcripts, may play a much more global role in RNA metabolism or protection than previously recognized, such as mediating retention of un-spliced transcripts in the nucleus.

2.2.3.4 Molecular targets of cytoplasmic mutant FUS proteins

Transcripts that showed increased binding in mutant compared to wild-type FUS were enriched for GO terms unrelated to those of wild-type FUS targets indicative of an alteration of its gene target spectrum. Yet, with effects even less pronounced than for the FUS knockdown, induction of mutant FUS in stable HEK293 cells did not result in >2 fold gene expression changes. The absence of a correlation of gene expression changes with RNA-binding sites suggests that FUS does not directly engage in gene regulatory function but that the regulatory effects are a consequence of altered competition of FUS binding sites with many different positive and negative regulatory RBPs or miRNPs. Because ALS-related FUS mutations do not impair RNA-binding, but instead mislocalize the protein to the cytoplasm

where mature mRNA targets are bound, one has to assume a cytoplasmic gain-of-function contribution in ALS.

Another gene mutation in *TARDBP*, which encodes for the RBP TDP-43, was recently shown to also cause familial ALS [165]. TDP-43 contains two RRM domains and, like FUS, is nuclearly localized as wild-type, but also shuttles based on its cytoplasmic inclusions in ALS [166]. It is conceivable that TDP-43 and FUS share some targets whose dysregulation may be responsible for causing ALS. Our dataset for wild-type and mutant FUS targets provides the starting-point for such a future analysis. Identification of common molecular targets may lead to the development of new strategies to understand the long-term effects of controlling these genes. This will allow for the development of new animal models and possibly novel strategies in ALS treatment, such as the identification of small-molecule drugs modulating the activity of jointly bound, disease-causing targets.

3. References

1. Fechter P, Brownlee GG. Recognition of mRNA cap structures by viral and cellular proteins. *The Journal of general virology*. 2005;86(Pt 5):1239-49.
2. Mangus DA, Evans MC, Jacobson A. Poly(A)-binding proteins: multifunctional scaffolds for the post-transcriptional control of gene expression. *Genome biology*. 2003;4(7):223.
3. Skabkin MA, Liabin DN, Ovchinnikov LP. [Nonspecific and specific interaction of Y-box binding protein 1 (YB-1) with mRNA and posttranscriptional regulation of protein synthesis in animal cells]. *Molekuliarnaia biologiya*. 2006;40(4):620-33.
4. Long JC, Caceres JF. The SR protein family of splicing factors: master regulators of gene expression. *The Biochemical journal*. 2009;417(1):15-27.
5. Tange TO, Shibuya T, Jurica MS, Moore MJ. Biochemical analysis of the EJC reveals two new factors and a stable tetrameric protein core. *Rna*. 2005;11(12):1869-83.
6. Gall JG. On the submicroscopic structure of chromosomes. *Brookhaven symposia in biology*. 1956(8):17-32.
7. Costanzo MC, Crawford ME, Hirschman JE, Kranz JE, Olsen P, Robertson LS, et al. YPD, PombePD and WormPD: model organism volumes of the BioKnowledge library, an integrated resource for protein information. *Nucleic acids research*. 2001;29(1):75-9.
8. Moore MJ. From birth to death: the complex lives of eukaryotic mRNAs. *Science*. 2005;309(5740):1514-8.
9. Maris C, Dominguez C, Allain FH. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *The FEBS journal*. 2005;272(9):2118-31.
10. Messias AC, Sattler M. Structural basis of single-stranded RNA recognition. *Accounts of chemical research*. 2004;37(5):279-87.
11. Gerstberger S, Hafner M, Tuschl T. A census of human RNA-binding proteins. *Nature reviews Genetics*. 2014;15(12):829-45.
12. Glisovic T, Bachorik JL, Yong J, Dreyfuss G. RNA-binding proteins and post-transcriptional gene regulation. *FEBS letters*. 2008;582(14):1977-86.
13. Sanchez-Diaz P, Penalva LO. Post-transcription meets post-genomic: the saga of RNA binding proteins in a new era. *RNA biology*. 2006;3(3):101-9.
14. Topisirovic I, Svitkin YV, Sonenberg N, Shatkin AJ. Cap and cap-binding proteins in the control of gene expression. *Wiley interdisciplinary reviews RNA*. 2011;2(2):277-98.
15. Chen HC, Cheng SC. Functional roles of protein splicing factors. *Bioscience reports*. 2012;32(4):345-59.
16. Tomaselli S, Locatelli F, Gallo A. The RNA editing enzymes ADARs: mechanism of action and human disease. *Cell and tissue research*. 2014;356(3):527-32.
17. Hollams EM, Giles KM, Thomson AM, Leedman PJ. mRNA stability and the control of gene expression: implications for human disease. *Neurochemical research*. 2002;27(10):957-80.
18. Johnson JM, Castle J, Garrett-Engele P, Kan Z, Loerch PM, Armour CD, et al. Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science*. 2003;302(5653):2141-4.
19. Buckanovich RJ, Posner JB, Darnell RB. Nova, the paraneoplastic Ri antigen, is homologous to an RNA-binding protein and is specifically expressed in the developing motor system. *Neuron*. 1993;11(4):657-72.
20. Anderson MP, Gregory RJ, Thompson S, Souza DW, Paul S, Mulligan RC, et al. Demonstration that CFTR is a chloride channel by alteration of its anion selectivity. *Science*. 1991;253(5016):202-5.

21. Svitkin YV, Sonenberg N. [Translational control by the poly(A) binding protein: a check for mRNA integrity]. *Molekuliarnaia biologiiia*. 2006;40(4):684-93.
22. Ding J, Hayashi MK, Zhang Y, Manche L, Krainer AR, Xu RM. Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. *Genes & development*. 1999;13(9):1102-15.
23. Ernst RK, Bray M, Rekosh D, Hammarskjold ML. A structured retroviral RNA element that mediates nucleocytoplasmic export of intron-containing RNA. *Molecular and cellular biology*. 1997;17(1):135-44.
24. Gonsalvez GB, Urbinati CR, Long RM. RNA localization in yeast: moving towards a mechanism. *Biology of the cell / under the auspices of the European Cell Biology Organization*. 2005;97(1):75-86.
25. Oleynikov Y, Singer RH. Real-time visualization of ZBP1 association with beta-actin mRNA during transcription and localization. *Current biology : CB*. 2003;13(3):199-207.
26. Mootz D, Ho DM, Hunter CP. The STAR/Maxi-KH domain protein GLD-1 mediates a developmental switch in the translational control of *C. elegans* PAL-1. *Development*. 2004;131(14):3263-72.
27. Guil S, Caceres JF. The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. *Nature structural & molecular biology*. 2007;14(7):591-6.
28. Nyquist KB, Thorsen J, Zeller B, Haaland A, Troen G, Heim S, et al. Identification of the TAF15-ZNF384 fusion gene in two new cases of acute lymphoblastic leukemia with a t(12;17)(p13;q12). *Cancer genetics*. 2011;204(3):147-52.
29. Barbouti A, Hoglund M, Johansson B, Lassen C, Nilsson PG, Hagemeijer A, et al. A novel gene, MSI2, encoding a putative RNA-binding protein is recurrently rearranged at disease progression of chronic myeloid leukemia and forms a fusion gene with HOXA9 as a result of the cryptic t(7;17)(p15;q23). *Cancer research*. 2003;63(6):1202-6.
30. Lukong KE, Chang KW, Khandjian EW, Richard S. RNA-binding proteins in human genetic disease. *Trends in genetics : TIG*. 2008;24(8):416-25.
31. Penagarikano O, Mulle JG, Warren ST. The pathophysiology of fragile x syndrome. *Annual review of genomics and human genetics*. 2007;8:109-29.
32. Jacquemont S, Hagerman RJ, Hagerman PJ, Leehey MA. Fragile-X syndrome and fragile X-associated tremor/ataxia syndrome: two faces of FMR1. *The Lancet Neurology*. 2007;6(1):45-55.
33. Lorson CL, Strasswimmer J, Yao JM, Baleja JD, Hahnen E, Wirth B, et al. SMN oligomerization defect correlates with spinal muscular atrophy severity. *Nature genetics*. 1998;19(1):63-6.
34. Darnell RB, Posner JB. Paraneoplastic syndromes involving the nervous system. *The New England journal of medicine*. 2003;349(16):1543-54.
35. Budde BS, Namavar Y, Barth PG, Poll-The BT, Nurnberg G, Becker C, et al. tRNA splicing endonuclease mutations cause pontocerebellar hypoplasia. *Nature genetics*. 2008;40(9):1113-8.
36. Wheeler TM, Thornton CA. Myotonic dystrophy: RNA-mediated muscle disease. *Current opinion in neurology*. 2007;20(5):572-6.
37. Hagerman PJ, Hagerman RJ. Fragile X-associated tremor/ataxia syndrome (FXTAS). *Mental retardation and developmental disabilities research reviews*. 2004;10(1):25-30.
38. Berry-Kravis E, Abrams L, Coffey SM, Hall DA, Greco C, Gane LW, et al. Fragile X-associated tremor/ataxia syndrome: clinical features, genetics, and testing guidelines. *Movement disorders : official journal of the Movement Disorder Society*. 2007;22(14):2018-30, quiz 140.
39. Spitzer JI, Ugras S, Runge S, Decarolis P, Antonescu C, Tuschl T, et al. mRNA and protein levels of FUS, EWSR1, and TAF15 are upregulated in liposarcoma. *Genes, chromosomes & cancer*. 2011;50(5):338-47.
40. Ito T, Kwon HY, Zimdahl B, Congdon KL, Blum J, Lento WE, et al. Regulation of myeloid leukaemia by the cell-fate determinant Musashi. *Nature*. 2010;466(7307):765-8.
41. Wang S, Li N, Yousefi M, Nakauka-Ddamba A, Li F, Parada K, et al. Transformation of the intestinal epithelium by the MSI2 RNA-binding protein. *Nature communications*. 2015;6:6517.
42. Sonenberg N, Hinnebusch AG. New modes of translational control in development, behavior, and disease. *Molecular cell*. 2007;28(5):721-9.

43. Lukong KE, Larocque D, Tyner AL, Richard S. Tyrosine phosphorylation of sam68 by breast tumor kinase regulates intranuclear localization and cell cycle progression. *The Journal of biological chemistry*. 2005;280(46):38639-47.
44. Busa R, Paronetto MP, Farini D, Pierantozzi E, Botti F, Angelini DF, et al. The RNA-binding protein Sam68 contributes to proliferation and survival of human prostate cancer cells. *Oncogene*. 2007;26(30):4372-82.
45. Chenard CA, Richard S. New implications for the QUAKING RNA binding protein in human disease. *Journal of neuroscience research*. 2008;86(2):233-42.
46. Hogan DJ, Riordan DP, Gerber AP, Herschlag D, Brown PO. Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system. *PLoS biology*. 2008;6(10):e255.
47. Schneider D, Gold L, Platt T. Selective enrichment of RNA species for tight binding to Escherichia coli rho factor. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*. 1993;7(1):201-7.
48. Jolma A, Kivioja T, Toivonen J, Cheng L, Wei G, Enge M, et al. Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. *Genome research*. 2010;20(6):861-73.
49. Philippakis AA, Qureshi AM, Berger MF, Bulyk ML. Design of compact, universal DNA microarrays for protein binding microarray experiments. *Journal of computational biology : a journal of computational molecular cell biology*. 2008;15(7):655-65.
50. Li X, Kazan H, Lipshitz HD, Morris QD. Finding the target sites of RNA-binding proteins. *Wiley interdisciplinary reviews RNA*. 2014;5(1):111-30.
51. Sengupta DJ, Wickens M, Fields S. Identification of RNAs that bind to a specific protein using the yeast three-hybrid system. *Rna*. 1999;5(4):596-601.
52. Tenenbaum SA, Carson CC, Lager PJ, Keene JD. Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. *Proceedings of the National Academy of Sciences of the United States of America*. 2000;97(26):14085-90.
53. Sanchez M, Galy B, Hentze MW, Muckenthaler MU. Identification of target mRNAs of regulatory RNA-binding proteins using mRNP immunopurification and microarrays. *Nature protocols*. 2007;2(8):2033-42.
54. Ule J, Jensen KB, Ruggiu M, Mele A, Ule A, Darnell RB. CLIP identifies Nova-regulated RNA networks in the brain. *Science*. 2003;302(5648):1212-5.
55. Licatalosi DD, Mele A, Fak JJ, Ule J, Kayikci M, Chi SW, et al. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature*. 2008;456(7221):464-9.
56. Taniguchi Y, Choi PJ, Li GW, Chen H, Babu M, Hearn J, et al. Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science*. 2010;329(5991):533-8.
57. Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*. 2010;141(1):129-41.
58. Mili S, Steitz JA. Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. *Rna*. 2004;10(11):1692-4.
59. Ascano M, Hafner M, Cekan P, Gerstberger S, Tuschl T. Identification of RNA-protein interaction networks using PAR-CLIP. *Wiley interdisciplinary reviews RNA*. 2012;3(2):159-77.
60. Spitzer J, Hafner M, Landthaler M, Ascano M, Farazi T, Wardle G, et al. PAR-CLIP (Photoactivatable Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation): a step-by-step protocol to the transcriptome-wide identification of binding sites of RNA-binding proteins. *Methods in enzymology*. 2014;539:113-61.
61. Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, et al. PAR-CLIP--a method to identify transcriptome-wide the binding sites of RNA binding proteins. *Journal of visualized experiments : JoVE*. 2010(41).
62. Kloetgen A, Munch PC, Borkhardt A, Hoell JI, McHardy AC. Biochemical and bioinformatic methods for elucidating the role of RNA-protein interactions in posttranscriptional regulation. *Briefings in functional genomics*. 2014.

63. Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends in genetics : TIG*. 2000;16(6):276-7.
64. Schroder J, Schroder H, Puglisi SJ, Sinha R, Schmidt B. SHREC: a short-read error correction method. *Bioinformatics*. 2009;25(17):2157-63.
65. Kelley DR, Schatz MC, Salzberg SL. Quake: quality-aware detection and correction of sequencing errors. *Genome biology*. 2010;11(11):R116.
66. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001;409(6822):860-921.
67. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754-60.
68. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology*. 2009;10(3):R25.
69. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature methods*. 2012;9(4):357-9.
70. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell*. 2010;38(4):576-89.
71. Georgiev S, Boyle AP, Jayasurya K, Ding X, Mukherjee S, Ohler U. Evidence-ranked motif identification. *Genome biology*. 2010;11(2):R19.
72. Zhang C, Lee KY, Swanson MS, Darnell RB. Prediction of clustered RNA-binding protein motif sites in the mammalian genome. *Nucleic acids research*. 2013;41(14):6793-807.
73. Bailey TL, Noble WS. Searching for statistically significant regulatory modules. *Bioinformatics*. 2003;19 Suppl 2:ii16-25.
74. Siddharthan R, Siggia ED, van Nimwegen E. PhyloGibbs: a Gibbs sampling motif finder that incorporates phylogeny. *PLoS computational biology*. 2005;1(7):e67.
75. Kazan H, Ray D, Chan ET, Hughes TR, Morris Q. RNAcontext: a new method for learning the sequence and structure binding preferences of RNA-binding proteins. *PLoS computational biology*. 2010;6:e1000832.
76. Jungkamp AC, Stoeckius M, Mecnas D, Grun D, Mastrobuoni G, Kempa S, et al. In vivo and transcriptome-wide identification of RNA binding protein target sites. *Molecular cell*. 2011;44(5):828-40.
77. Miller MR, Robinson KJ, Cleary MD, Doe CQ. TU-tagging: cell type-specific RNA isolation from intact complex tissues. *Nature methods*. 2009;6(6):439-41.
78. Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann BM, Strein C, et al. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell*. 2012;149(6):1393-406.
79. Pinol-Roma S, Dreyfuss G. hnRNP proteins: localization and transport between the nucleus and the cytoplasm. *Trends in cell biology*. 1993;3(5):151-5.
80. Tan AY, Manley JL. The TET family of proteins: functions and roles in disease. *Journal of molecular cell biology*. 2009;1(2):82-92.
81. Pahlich S, Quero L, Roschitzki B, Leemann-Zakaryan RP, Gehring H. Analysis of Ewing sarcoma (EWS)-binding proteins: interaction with hnRNP M, U, and RNA-helicases p68/72 within protein-RNA complexes. *Journal of proteome research*. 2009;8(10):4455-65.
82. Riggi N, Cironi L, Suva ML, Stamenkovic I. Sarcomas: genetics, signalling, and cellular origins. Part 1: The fellowship of TET. *The Journal of pathology*. 2007;213(1):4-20.
83. Lagier-Tourenne C, Cleveland DW. Rethinking ALS: the FUS about TDP-43. *Cell*. 2009;136(6):1001-4.
84. Schwartz JC, Cech TR, Parker RR. Biochemical Properties and Biological Functions of FET Proteins. *Annual review of biochemistry*. 2014.
85. Bertolotti A, Lutz Y, Heard DJ, Chambon P, Tora L. hTAF(II)68, a novel RNA/ssDNA-binding protein with homology to the pro-oncoproteins TLS/FUS and EWS is associated with both TFIID and RNA polymerase II. *The EMBO journal*. 1996;15(18):5022-31.
86. Iko Y, Kodama TS, Kasai N, Oyama T, Morita EH, Muto T, et al. Domain architectures and characterization of an RNA-binding protein, TLS. *The Journal of biological chemistry*. 2004;279(43):44834-40.

87. Hoell JI, Larsson E, Runge S, Nusbaum JD, Duggimpudi S, Farazi TA, et al. RNA targets of wild-type and mutant FET family proteins. *Nature structural & molecular biology*. 2011;18(12):1428-31.
88. Colombrita C, Onesto E, Megiorni F, Pizzuti A, Baralle FE, Buratti E, et al. TDP-43 and FUS RNA-binding proteins bind distinct sets of cytoplasmic messenger RNAs and differently regulate their post-transcriptional fate in motoneuron-like cells. *The Journal of biological chemistry*. 2012;287(19):15635-47.
89. Kwon I, Kato M, Xiang S, Wu L, Theodoropoulos P, Mirzaei H, et al. Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. *Cell*. 2013;155(5):1049-60.
90. Bertolotti A, Melot T, Acker J, Vigneron M, Delattre O, Tora L. EWS, but not EWS-FLI-1, is associated with both TFIID and RNA polymerase II: interactions between two members of the TET family, EWS and hTAFII68, and subunits of TFIID and RNA polymerase II complexes. *Molecular and cellular biology*. 1998;18(3):1489-97.
91. Schwartz JC, Ebmeier CC, Podell ER, Heimiller J, Taatjes DJ, Cech TR. FUS binds the CTD of RNA polymerase II and regulates its phosphorylation at Ser2. *Genes & development*. 2012;26(24):2690-5.
92. Powers CA, Mathur M, Raaka BM, Ron D, Samuels HH. TLS (translocated-in-liposarcoma) is a high-affinity interactor for steroid, thyroid hormone, and retinoid receptors. *Molecular endocrinology*. 1998;12(1):4-18.
93. Uranishi H, Tetsuka T, Yamashita M, Asamitsu K, Shimizu M, Itoh M, et al. Involvement of the pro-oncoprotein TLS (translocated in liposarcoma) in nuclear factor-kappa B p65-mediated transcription as a coactivator. *The Journal of biological chemistry*. 2001;276(16):13395-401.
94. Araya N, Hirota K, Shimamoto Y, Miyagishi M, Yoshida E, Ishida J, et al. Cooperative interaction of EWS with CREB-binding protein selectively activates hepatocyte nuclear factor 4-mediated transcription. *The Journal of biological chemistry*. 2003;278(7):5427-32.
95. Lee J, Rhee BK, Bae GY, Han YM, Kim J. Stimulation of Oct-4 activity by Ewing's sarcoma protein. *Stem cells*. 2005;23(6):738-51.
96. Ishigaki S, Masuda A, Fujioka Y, Iguchi Y, Katsuno M, Shibata A, et al. Position-dependent FUS-RNA interactions regulate alternative splicing events and transcriptions. *Scientific reports*. 2012;2:529.
97. Rogelj B, Easton LE, Bogu GK, Stanton LW, Rot G, Curk T, et al. Widespread binding of FUS along nascent RNA regulates alternative splicing in the brain. *Scientific reports*. 2012;2:603.
98. Paronetto MP, Minana B, Valcarcel J. The Ewing sarcoma protein regulates DNA damage-induced alternative splicing. *Molecular cell*. 2011;43(3):353-68.
99. Yang L, Embree LJ, Hickstein DD. TLS-ERG leukemia fusion protein inhibits RNA splicing mediated by serine-arginine proteins. *Molecular and cellular biology*. 2000;20(10):3345-54.
100. Chansky HA, Hu M, Hickstein DD, Yang L. Oncogenic TLS/ERG and EWS/Fli-1 fusion proteins inhibit RNA splicing mediated by YB-1 protein. *Cancer research*. 2001;61(9):3586-90.
101. Hackl W, Luhrmann R. Molecular cloning and subcellular localisation of the snRNP-associated protein 69KD, a structural homologue of the proto-oncoproteins TLS and EWS with RNA and DNA-binding properties. *Journal of molecular biology*. 1996;264(5):843-51.
102. Tsuiji H, Iguchi Y, Furuya A, Kataoka A, Hatsuta H, Atsuta N, et al. Spliceosome integrity is defective in the motor neuron diseases ALS and SMA. *EMBO molecular medicine*. 2013;5(2):221-34.
103. Mackenzie IR, Neumann M. FET proteins in frontotemporal dementia and amyotrophic lateral sclerosis. *Brain research*. 2012;1462:40-3.
104. Yamaguchi A, Kitajo K. The effect of PRMT1-mediated arginine methylation on the subcellular localization, stress granules, and detergent-insoluble aggregates of FUS/TLS. *PloS one*. 2012;7(11):e49267.
105. Araya N, Hiraga H, Kako K, Arao Y, Kato S, Fukamizu A. Transcriptional down-regulation through nuclear exclusion of EWS methylated by PRMT1. *Biochemical and biophysical research communications*. 2005;329(2):653-60.
106. Kim JD, Kako K, Kakiuchi M, Park GG, Fukamizu A. EWS is a substrate of type I protein arginine methyltransferase, PRMT8. *International journal of molecular medicine*. 2008;22(3):309-15.

107. Wang WY, Pan L, Su SC, Quinn EJ, Sasaki M, Jimenez JC, et al. Interaction of FUS and HDAC1 regulates DNA damage response and repair in neurons. *Nature neuroscience*. 2013;16(10):1383-91.
108. Esiashvili N, Goodman M, Marcus RB, Jr. Changes in incidence and survival of Ewing sarcoma patients over the past 3 decades: Surveillance Epidemiology and End Results data. *Journal of pediatric hematology/oncology*. 2008;30(6):425-30.
109. Aurias A, Rimbaut C, Buffe D, Zucker JM, Mazabraud A. Translocation involving chromosome 22 in Ewing's sarcoma. A cytogenetic study of four fresh tumors. *Cancer genetics and cytogenetics*. 1984;12(1):21-5.
110. Lee SB, Kolquist KA, Nichols K, Englert C, Maheswaran S, Ladanyi M, et al. The EWS-WT1 translocation product induces PDGFA in desmoplastic small round-cell tumour. *Nature genetics*. 1997;17(3):309-13.
111. Pellin A, Monteagudo C, Lopez-Gines C, Carda C, Boix J, Llombart-Bosch A. New type of chimeric fusion product between the EWS and ATF1 genes in clear cell sarcoma (malignant melanoma of soft parts). *Genes, chromosomes & cancer*. 1998;23(4):358-60.
112. Crozat A, Aman P, Mandahl N, Ron D. Fusion of CHOP to a novel RNA-binding protein in human myxoid liposarcoma. *Nature*. 1993;363(6430):640-4.
113. Hosaka T, Nakashima Y, Kusuzaki K, Murata H, Nakayama T, Nakamata T, et al. A novel type of EWS-CHOP fusion gene in two cases of myxoid liposarcoma. *The Journal of molecular diagnostics : JMD*. 2002;4(3):164-71.
114. Paronetto MP. Ewing sarcoma protein: a key player in human cancer. *International journal of cell biology*. 2013;2013:642853.
115. Li H, Watford W, Li C, Parmelee A, Bryant MA, Deng C, et al. Ewing sarcoma gene EWS is essential for meiosis and B lymphocyte development. *The Journal of clinical investigation*. 2007;117(5):1314-23.
116. Hicks GG, Singh N, Nashabi A, Mai S, Bozek G, Klewes L, et al. Fus deficiency in mice results in defective B-lymphocyte development and activation, high levels of chromosomal instability and perinatal death. *Nature genetics*. 2000;24(2):175-9.
117. Cho J, Shen H, Yu H, Li H, Cheng T, Lee SB, et al. Ewing sarcoma gene Ews regulates hematopoietic stem cell senescence. *Blood*. 2011;117(4):1156-66.
118. Mosmann TR, Coffman RL. TH1 and TH2 cells: different patterns of lymphokine secretion lead to different functional properties. *Annual review of immunology*. 1989;7:145-73.
119. Zhou H, Lee KA. An hsRBP4/7-dependent yeast assay for trans-activation by the EWS oncogene. *Oncogene*. 2001;20(12):1519-24.
120. Fujimura Y, Siddique H, Lee L, Rao VN, Reddy ES. EWS-ATF-1 chimeric protein in soft tissue clear cell sarcoma associates with CREB-binding protein and interferes with p53-mediated trans-activation function. *Oncogene*. 2001;20(46):6653-9.
121. Zhang D, Paley AJ, Childs G. The transcriptional repressor ZFM1 interacts with and modulates the ability of EWS to activate transcription. *The Journal of biological chemistry*. 1998;273(29):18086-91.
122. Knoop LL, Baker SJ. The splicing factor U1C represses EWS/FLI-mediated transactivation. *The Journal of biological chemistry*. 2000;275(32):24865-71.
123. Dutertre M, Sanchez G, De Cian MC, Barbier J, Dardenne E, Gratadou L, et al. Cotranscriptional exon skipping in the genotoxic stress response. *Nature structural & molecular biology*. 2010;17(11):1358-66.
124. Sanchez G, Bittencourt D, Laud K, Barbier J, Delattre O, Auboeuf D, et al. Alteration of cyclin D1 transcript elongation by a mutated transcription factor up-regulates the oncogenic D1b splice isoform in cancer. *Proceedings of the National Academy of Sciences of the United States of America*. 2008;105(16):6004-9.
125. Welford SM, Hebert SP, Deneen B, Arvand A, Denny CT. DNA binding domain-independent pathways are involved in EWS/FLI1-mediated oncogenesis. *The Journal of biological chemistry*. 2001;276(45):41977-84.
126. Sankar S, Lessnick SL. Promiscuous partnerships in Ewing's sarcoma. *Cancer genetics*. 2011;204(7):351-65.

127. Merolla F, Luise C, Muller MT, Pacelli R, Fusco A, Celetti A. Loss of CCDC6, the first identified RET partner gene, affects p_{H2AX} S139 levels and accelerates mitotic entry upon DNA damage. *PloS one*. 2012;7(5):e36177.
128. Leone V, Mansueto G, Pierantoni GM, Tornincasa M, Merolla F, Cerrato A, et al. CCDC6 represses CREB1 activity by recruiting histone deacetylase 1 and protein phosphatase 1. *Oncogene*. 2010;29(30):4341-51.
129. Viswanathan SR, Daley GQ, Gregory RI. Selective blockade of microRNA processing by Lin28. *Science*. 2008;320(5872):97-100.
130. Lezon-Geyda K, Najfeld V, Johnson EM. Deletions of PURA, at 5q31, and PURB, at 7p13, in myelodysplastic syndrome and progression to acute myelogenous leukemia. *Leukemia*. 2001;15(6):954-62.
131. Puxeddu E, Zhao G, Stringer JR, Medvedovic M, Moretti S, Fagin JA. Characterization of novel non-clonal intrachromosomal rearrangements between the H4 and PTEN genes (H4/PTEN) in human thyroid cell lines and papillary thyroid cancer specimens. *Mutation research*. 2005;570(1):17-32.
132. Hayashi R, Goto Y, Ikeda R, Yokoyama KK, Yoshida K. CDCA4 is an E2F transcription factor family-induced nuclear factor that regulates E2F-dependent transcriptional activation and cell proliferation. *The Journal of biological chemistry*. 2006;281(47):35633-48.
133. Haupt Y, Maya R, Kazaz A, Oren M. Mdm2 promotes the rapid degradation of p53. *Nature*. 1997;387(6630):296-9.
134. Wade M, Li YC, Wahl GM. MDM2, MDMX and p53 in oncogenesis and cancer therapy. *Nature reviews Cancer*. 2013;13(2):83-96.
135. Pishas KI, Al-Ejeh F, Zinonos I, Kumar R, Evdokiou A, Brown MP, et al. Nutlin-3a is a potential therapeutic for ewing sarcoma. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2011;17(3):494-504.
136. Li Y, Tanaka K, Fan X, Nakatani F, Li X, Nakamura T, et al. Inhibition of the transcriptional function of p53 by EWS-Fli1 chimeric protein in Ewing Family Tumors. *Cancer letters*. 2010;294(1):57-65.
137. Krejci P, Prochazkova J, Bryja V, Kozubik A, Wilcox WR. Molecular pathology of the fibroblast growth factor family. *Human mutation*. 2009;30(9):1245-55.
138. Abdel-Rahman WM, Kalinina J, Shoman S, Eissa S, Ollikainen M, Elomaa O, et al. Somatic FGF9 mutations in colorectal and endometrial carcinomas associated with membranous beta-catenin. *Human mutation*. 2008;29(3):390-7.
139. Sangle NA, Perkins SL. Core-binding factor acute myeloid leukemia. *Archives of pathology & laboratory medicine*. 2011;135(11):1504-9.
140. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols*. 2009;4(1):44-57.
141. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids research*. 2009;37(1):1-13.
142. Kondo T, Ezzat S, Asa SL. Pathogenetic mechanisms in thyroid follicular-cell neoplasia. *Nature reviews Cancer*. 2006;6(4):292-306.
143. Celetti A, Cerrato A, Merolla F, Vitagliano D, Vecchio G, Grieco M. H4(D10S170), a gene frequently rearranged with RET in papillary thyroid carcinomas: functional characterization. *Oncogene*. 2004;23(1):109-21.
144. Schwaller J, Anastasiadou E, Cain D, Kutok J, Wojiski S, Williams IR, et al. H4(D10S170), a gene frequently rearranged in papillary thyroid carcinoma, is fused to the platelet-derived growth factor receptor beta gene in atypical chronic myeloid leukemia with t(5;10)(q33;q22). *Blood*. 2001;97(12):3910-8.
145. Thanasopoulou A, Stravopodis DJ, Dimas KS, Schwaller J, Anastasiadou E. Loss of CCDC6 affects cell cycle through impaired intra-S-phase checkpoint control. *PloS one*. 2012;7(2):e31007.
146. Thanasopoulou A, Xanthopoulou AG, Anagnostopoulos AK, Konstantakou EG, Margaritis LH, Papassideri IS, et al. Silencing of CCDC6 reduces the expression of 14-3-3sigma in colorectal carcinoma cells. *Anticancer research*. 2012;32(3):907-13.

147. Zhao J, Tang J, Men W, Ren K. FBXW7-mediated degradation of CCDC6 is impaired by ATM during DNA damage response in lung cancer cells. *FEBS Lett.* 2012;586(24):4257-63.
148. Wang Z, Inuzuka H, Zhong J, Wan L, Fukushima H, Sarkar FH, et al. Tumor suppressor functions of FBW7 in cancer development and progression. *FEBS Lett.* 2012;586(10):1409-18.
149. Azuma M, Embree LJ, Sabaawy H, Hickstein DD. Ewing sarcoma protein *ewsr1* maintains mitotic integrity and proneural cell survival in the zebrafish embryo. *PloS one.* 2007;2(10):e979.
150. Lerga A, Hallier M, Delva L, Orvain C, Gallais I, Marie J, et al. Identification of an RNA binding specificity for the potential splicing factor TLS. *The Journal of biological chemistry.* 2001;276(9):6807-16.
151. Oubridge C, Ito N, Evans PR, Teo CH, Nagai K. Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature.* 1994;372(6505):432-8.
152. Price SR, Evans PR, Nagai K. Crystal structure of the spliceosomal U2B'-U2A' protein complex bound to a fragment of U2 small nuclear RNA. *Nature.* 1998;394(6694):645-50.
153. Lagier-Tourenne C, Cleveland DW. Neurodegeneration: An expansion in ALS genetics. *Nature.* 2010;466(7310):1052-3.
154. Kwiatkowski TJ, Jr., Bosco DA, Leclerc AL, Tamrazian E, Vanderburg CR, Russ C, et al. Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. *Science.* 2009;323(5918):1205-8.
155. Vance C, Rogelj B, Hortobagyi T, De Vos KJ, Nishimura AL, Sreedharan J, et al. Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. *Science.* 2009;323(5918):1208-11.
156. Dormann D, Rodde R, Edbauer D, Bentmann E, Fischer I, Hruscha A, et al. ALS-associated fused in sarcoma (FUS) mutations disrupt Transportin-mediated nuclear import. *The EMBO journal.* 2010;29(16):2841-57.
157. Zakaryan RP, Gehring H. Identification and characterization of the nuclear localization/retention signal in the EWS proto-oncoprotein. *Journal of molecular biology.* 2006;363(1):27-38.
158. Shaw DJ, Morse R, Todd AG, Eggleton P, Lorson CL, Young PJ. Identification of a tripartite import signal in the Ewing Sarcoma protein (EWS). *Biochemical and biophysical research communications.* 2009;390(4):1197-201.
159. Romeo S, Dei Tos AP. Soft tissue tumors associated with EWSR1 translocation. *Virchows Archiv : an international journal of pathology.* 2010;456(2):219-34.
160. Bedford MT, Clarke SG. Protein arginine methylation in mammals: who, what, and why. *Molecular cell.* 2009;33(1):1-13.
161. Kuroda M, Sok J, Webb L, Baechtold H, Urano F, Yin Y, et al. Male sterility and enhanced radiation sensitivity in TLS(-/-) mice. *The EMBO journal.* 2000;19(3):453-62.
162. Wijesekera LC, Leigh PN. Amyotrophic lateral sclerosis. *Orphanet journal of rare diseases.* 2009;4:3.
163. Hallier M, Lerga A, Barnache S, Tavitian A, Moreau-Gachelin F. The transcription factor Spi-1/PU.1 interacts with the potential splicing factor TLS. *The Journal of biological chemistry.* 1998;273(9):4838-42.
164. Krecic AM, Swanson MS. hnRNP complexes: composition, structure, and function. *Current opinion in cell biology.* 1999;11(3):363-71.
165. Sreedharan J, Blair IP, Tripathi VB, Hu X, Vance C, Rogelj B, et al. TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. *Science.* 2008;319(5870):1668-72.
166. Geser F, Martinez-Lage M, Kwong LK, Lee VM, Trojanowski JQ. Amyotrophic lateral sclerosis, frontotemporal dementia and beyond: the TDP-43 diseases. *Journal of neurology.* 2009;256(8):1205-14.

4. Abbreviations

4SU- 4-Thio-Uridine

6SG-6-Thio-Guanidine

ADAR- Adenosine Deaminase Acting on RNA

ALS- amyotrophic lateral sclerosis

AML- Acute myeloid leukemia

APA- Alternative cleavage and Polyadenylation

ARE-(AU)-rich elements

ASH-absent, small or homeotic discs

ATF1-Activating transcription factor 1

BCAT1-Branched chain amino-acid transaminase 1

BWA- Burrows-Wheeler Aligner

CBC- cap binding protein complex

CC-Coupled cluster

CCDC6-Coiled Coil Domain containing protein 6,

CCDN1-Cyclin D1

CCK8-Cell counting kit-8

cDNA- Complimentary DNA

cERMIT-Evidence ranked motif identification

CFTR- Cystic fibrosis transmembrane receptor

CML-Chronic myeloid leukemia

CPSF73- Cleavage and polyadenylation specificity factor 73

CREB-cAMP response element-bindin protein

C-terminal-Carboxy terminal

CUGBP1- CUG-binding protein 1
CUL2-Cullin2
DAVID- Database for Annotation, visualization and integrated discovery
DDR-DNA damage response
DM1- myotonic dystrophy type 1
DM2- myotonic dystrophy type 2
DMPK- myotonic dystrophy protein kinase
EJC-Exon junction complexes
eIF4E- elongation factor 4 E
EMSA-Electrophoretic mobility shift assay
ERG-v-ets Avian erythroblastosis virus E26 oncogene homolog
ESFT-Ewing Sarcoma family tumors
ETV-1-ets variant 1
EWS- Ewing Sarcoma
EWSR1-Ewing Sarcoma breakpoint 1
FANCL- fanconi anemia, complementation group L
FGF9-Fibroblast growth factor 9
FITC- Flourescein isothiocyanate
FMR1-Fragile X Mental Retardation 1
FTLD- Frontotemporal lobar dementia
FUS-Fused in Sarcoma
FXTAS- Fragile-X-associated tremor/ataxia syndrome
G0-Gap 0 phase
G1-Gap 1 phase
G2-Gap 2 phase
Gld1-Glycerol dehydrogenase 1
GO- Gene ontology

HA- Human Influenza Hemagglutinin
HDAC1-Histone deacetylase 1
HEK-Human Embryonic Kidney cells
HITS-CLIP-High throughput Sequencing of RNA isolated by crosslinking IP
hnRNP- Heterogeneous nuclear ribonucleoprotein particle
HOMER- Hypergeometric optimization of Motif Enrichment
HOXA9-Homeobox A cluster 9
i-CLIP- individual nucleotide resolution Crosslinking and Immunoprecipitation
IFN-Interferon
IL-Interleukin
IP-Immunoprecipitation
LC- Low complexity
lncRNA-long noncoding RNA
MBNL1- muscleblind-like protein 1
mCarts - Hidden Markov model to predict clustered RNA motif sites
MCAST - Motif cluster alignment and search tool
MEME - Multiple Em for Motif Elicitation
miRNA - microRNA
M - Mitosis phase
mRNA - messenger ribonucleic acid
MSI2 - Musashi 2
NF- κ B -Nuclear factor kappa light chain enhancer of activated B cells
NGS -Next generation sequencing
NLS - Nuclear localization signal
OPMD - Oculopharyngeal muscular dystrophy
PABP - Poly (A)binding protein
Pal1 - Phenylalanine ammonia lyase 1

PAR-CLIP - PhotoActivatable Ribonucleoside enhanced CrossLinking and IP

PCR - Polymerase chain reaction

PEM/SN - Paraneoplastic encephalomyelitis/sensory neuropathy

PI - Propidium Iodide

piRNA - PIWI interacting RNAs

PNET - Primitive neuroectodermal tumors

POLA1 - Polymerase (DNA directed), alpha 1, catalytic subunit

POMA - Paraneoplastic opsoclonus-myoclonus ataxia

PPP1CB - Protein phosphatase 1, catalytic subunit beta isozyyme

PRMT - Protein arginine methyltransferases

PTGR – Post-transcriptional gene regulation

QKI - Quaking

qRT-PCR - quantitative real time PCR

RBD - RNA binding domain

RBP - RNA binding protein

RHOU - Ras homolog family member U

RIP-CHIP - RNA immunoprecipitation- microarray

RNA pol II - RNA polymerase II

RRM - RNA recognition motif

rRNA - ribosomal RNA

S - Synthesis phase of cell cycle

SDS - Sodium dodecyl sulphate

SELEX - Systematic evolution of ligands by exponential enrichment

SF2/ASF - Serine/arginine-rich splicing factor 1

SHREC - Short read error correction

SKP2 - S-Phase kinase-associated protein 2

SMN1 - Survival of motor neuron 1

snoRNA - small nucleolar RNA

snRNA- small nuclear RNA

SNRNPA1- small nuclear ribonucleoprotein 1

SNRPB2-small nuclear ribonucleoprotein polypeptide B 2

SR-Serine arginine proteins

STAR-Steroidogenic acute regulatory protein

TAF15-TATA box binding factor (TBP)-associated factor 15

TDP43- TAR DNA binding protein 43

TLS-Translocated in liposarcoma

TNPO1-Transportin 1

tRNA-transfer RNA

TU- ThioUracil

UTR-Untranslated region

UV-Ultraviolet

ZBP1-Z-DNA binding protein 1

ZNF384-Zinc Finger 384

ZNF9- Zinc finger 9

RESEARCH ARTICLE

The Cell Cycle Regulator CCDC6 Is a Key Target of RNA-Binding Protein EWS

Sujitha Duggimpudi¹, Erik Larsson², Schafiq Nabhani¹, Arndt Borkhardt¹, Jessica I Hoell^{1*}

1 Department of Pediatric Oncology, Hematology and Clinical Immunology, Center for Child and Adolescent Health, Heinrich Heine University, Medical Faculty, Duesseldorf, Germany, **2** Department of Medical Biochemistry and Cell biology, Institute of Biomedicine, The Sahlgrenska Academy, University of Gothenburg, Sweden

* Jessica.hoell@med.uni-duesseldorf.de


 OPEN ACCESS

Citation: Duggimpudi S, Larsson E, Nabhani S, Borkhardt A, Hoell JI (2015) The Cell Cycle Regulator CCDC6 Is a Key Target of RNA-Binding Protein EWS. *PLoS ONE* 10(3): e0119066. doi:10.1371/journal.pone.0119066

Academic Editor: Sean Bong Lee, Tulane University School of Medicine, UNITED STATES

Received: May 8, 2014

Accepted: January 27, 2015

Published: March 9, 2015

Copyright: © 2015 Duggimpudi et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its supporting Information files.

Funding: This work was supported by a "Forschungsstipendium der Vereinigung rheinisch-westfälischer Kinder- und Jugendärzte und Kinderchirurgen" fellowship and Elterninitiative Kinderkrebsklinik e.V. Düsseldorf. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Genetic translocation of EWSR1 to ETS transcription factor coding region is considered as primary cause for Ewing sarcoma. Previous studies focused on the biology of chimeric transcription factors formed due to this translocation. However, the physiological consequences of heterozygous EWSR1 loss in these tumors have largely remained elusive. Previously, we have identified various mRNAs bound to EWS using PAR-CLIP. In this study, we demonstrate CCDC6, a known cell cycle regulator protein, as a novel target regulated by EWS. siRNA mediated down regulation of EWS caused an elevated apoptosis in cells in a CCDC6-dependant manner. This effect was rescued upon re-expression of CCDC6. This study provides evidence for a novel functional link through which wild-type EWS operates in a target-dependant manner in Ewing sarcoma.

Introduction

Ewing sarcoma which was first reported by James Ewing in 1921 is the second most common bone and soft tissue malignancy in adolescents and young adults [1, 2]. Genetically, 90% of these tumors are characterized by a translocation whereby the N-terminal portion of the RNA-binding protein (RBP) EWSR1 is joined to a DNA-binding protein belonging to the ETS family of transcription factors (e.g. FLI1, ERG, and ETV1) (Fig. 1A). Additionally, EWSR1 fusions to ATF1 cause soft tissue clear cell sarcoma while fusions to CHOP cause myxoid liposarcoma [3, 4]. Considering that the resulting chimeric transcription factors such as EWS-FLI are under the control of the strong FET promoter, they are expressed at high levels in the cell and generally believed to be the main cause of malignant transformation. Additionally, the loss of the EWSR1 allele creates haploinsufficiency of EWS protein which affects its RNA-binding activity and also its mRNA targets suggesting that EWS and its targets have important roles in the development of disease [5]. It has also been shown that EWS/FLI alone is not sufficient to induce sarcomagenesis in a transgenic mouse model suggesting that factors unrelated to the aberrant transcription factors also contribute to the development of Ewing sarcoma family tumours (ESFT) [6].

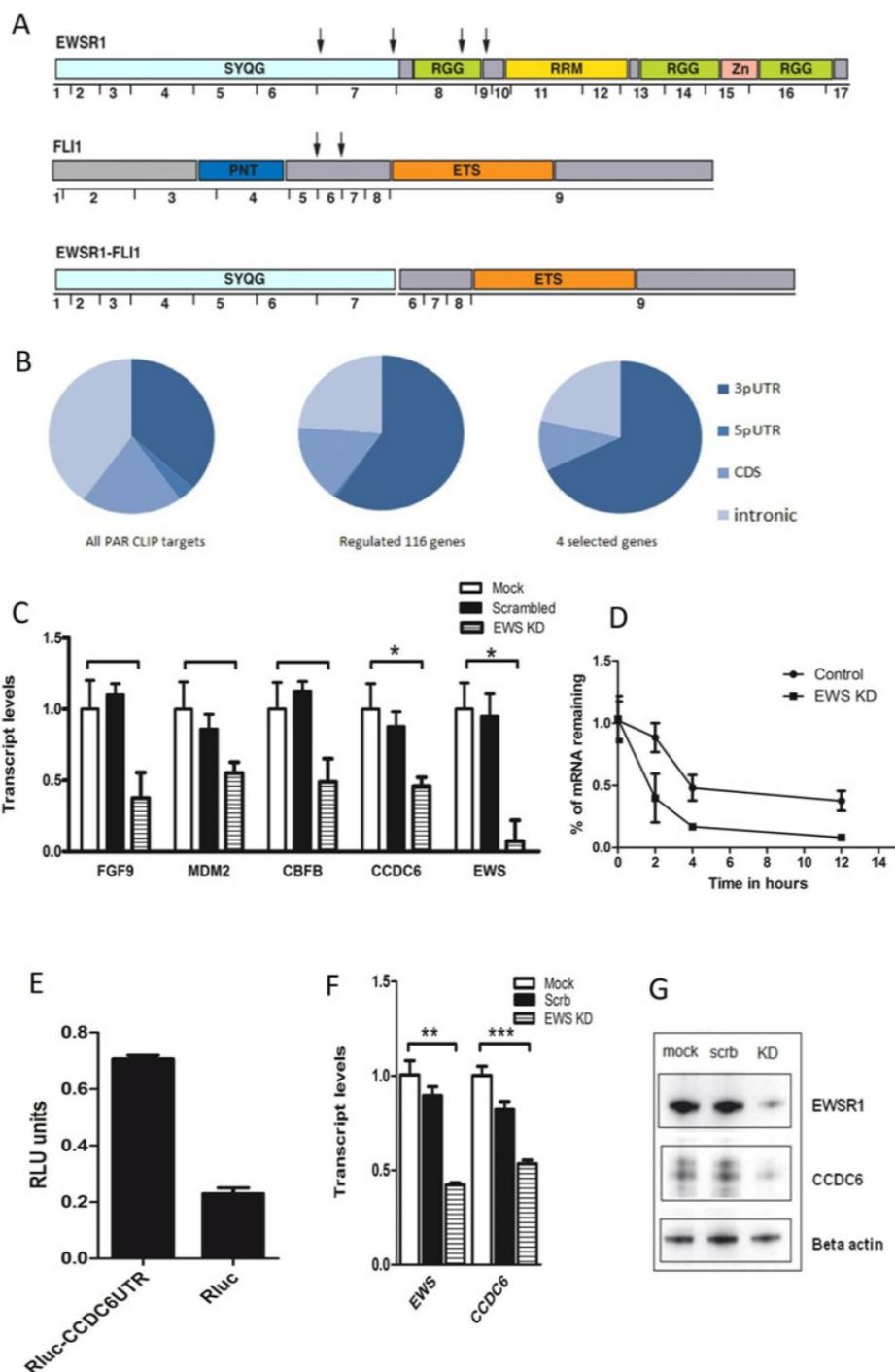


Fig 1. Regulation of targets by EWS in vivo. **A.** Protein domain organization of EWS and FLI1. The black vertical arrows indicate common breakpoints in Ewing sarcoma. Numbers correspond to exons and a typical EWS-FLI1 fusion protein is also shown. Note that the RNA-binding domain of EWS is lost in the process of translocation. **B.** Pie diagram showing the distribution of PAR-CLIP clusters across 3'UTR, 5'UTR, intronic and coding regions of Refseq RNAs. The three diagrams give the cluster distribution of all sequenced EWS

Antibodies

Monoclonal anti-EWS (Abcam, AB54708), monoclonal anti-CCDC6 (Abcam, AB56353), monoclonal anti- β actin antibody (Sigma Aldrich, A2228) were used as primary antibodies at 1:3000, 1:1000 and 1:10000 dilutions respectively. Secondary goat anti-mouse IgG-HRP antibody (Santa Cruz biotechnology, sc-2005) was used at 1:10000 dilution.

siRNA transfections

HEK293T and MHH-ES-1 cell lines were reversely transfected with siRNAs using Lipofectamine RNAiMAX (Invitrogen). siRNAs targeting EWS were designed and obtained from Eurofins MWG operon. AllStars Negative Control siRNA was obtained from Qiagen. HEK293T and MHH-ES-1 cell lines were transfected with siRNAs at 50 nM concentration in 12 well plates for EWS, CCDC6 mRNA and protein analysis. 72h post transfection, cells were harvested for analysis. Transfection rates were in the range of 60–70% for MHH-ES-1 and 90% for HEK293T cell lines. For the sequence of the siRNA and the figure showing the region being targeted by the siRNA see the [S1 Table](#) and [S1A Fig](#).

Quantitative RT-PCR

Total RNA was extracted using the RNeasy Mini kit (Qiagen). cDNA was synthesized using Superscript Reverse Transcriptase Kit III (Invitrogen) according to the manufacturer's instructions (oligo dT). Quantitative PCR was carried out using Power SYBR green kit (Applied Biosystems). All reactions were run on an ABI 7500 Real time PCR machine (Applied Biosystems) in triplicate. Data was acquired using the ABI SDS 2.0.1 software package. RNA isolated from the samples was tested for the expression levels of the chosen targets and their Δ ct values were subtracted from the respective beta actin expression levels.

Western blot

Cells were harvested, resuspended in NP40 lysis buffer and lysed. A 10% SDS-PAGE gel was run in Tris-glycine-SDS buffer. A semi-dry transfer procedure was carried out onto cellulose membrane. After transfer, the membrane was blocked in TBS with Tween 20 and 5% milk. The membrane was probed with mouse monoclonal antibodies detecting CCDC6, EWS and beta actin. Horseradish peroxidase (HRP)-conjugated goat anti-mouse antibody was used as a secondary antibody. Chemiluminescence was used to detect EWS, CCDC6 and beta actin using the Super Signal West Pico chemiluminescent Substrate (Thermoscientific).

Reporter assay

HEK293T cells were co-transfected with 25ng of psiCHECK-2 plasmid containing the target mRNA sequences and 50ng of EWS expression plasmid using Lipofectamine 2000 (Invitrogen) transfection reagent. 25ng of empty vector psiCHECK-2 along with 50ng of EWS expression plasmid was used as negative control. A total of 5×10^4 cells were plated in 24 well plates 24h prior to transfection. 48h after transfection cell lysates were prepared according to the manufacturer's instructions with the Dual luciferase assay system (Promega). Renilla and firefly luminescence was read using Luminoskan ascent microplate luminometer (Thermoscientific). Renilla/ firefly luciferase ratios were calculated from the mean values of triplicates. Data is represented as mean SEM (Standard error of mean) values.

PAR-CLIP targets, all targets regulated by EWS and the four targets we validated (FGF9, MDM2, CBFβ, CCDC6). **C.** Relative mRNA levels of target genes FGF9, MDM2, CBFβ, CCDC6 and EWS in HEK293T cells following EWS knockdown assayed by qRT-PCR (mock: only transfection reagent used; scrambled: AllStars Negative Control siRNA; EWS: siRNA targeting EWS). Relative mRNA levels were normalized to beta actin and quantified relative to the mock and scrambled control levels. Results are shown as mean SEM values (* $P < 0.05$; $n = 3$ per group). **D.** Amount of CCDC6 mRNA transcript percentage is measured upon knocking down of EWS as compared to control. The level of transcript was measured by qRT-PCR after knocking down for 24 hours followed by treatment with actinomycin D. The linear regression and slopes were calculated and the data is presented as Mean and SEM on a linear scale. **E.** Luciferase activity of CCDC6 upon EWS transfection (normalized to empty psiCHECK-2 plasmid). Data is shown as the fold increase in luciferase activity (RLU units) relative to control. Results are shown as the mean SEM values (* $P < 0.05$; $n = 3$ per group). **F.** Relative mRNA levels of CCDC6 and EWS in mock, control and EWS knockdown in MHH-ES-1 cells. Knockdown of EWS decreased the expression of CCDC6. The mRNA levels were normalized to beta actin. Data is represented as mean SEM values (* $P < 0.05$; $n = 3$ per group). **G.** Western blot showing the downregulation of CCDC6 upon EWS knockdown in MHH-ES-1 cells. Antibodies are indicated.

doi:10.1371/journal.pone.0119066.g001

EWS is one of roughly 600 RNA-binding proteins (RBPs) which play important roles in mRNA stability, transport and cellular localization [7]. It belongs to the FET family of RBPs which includes FUS, EWS and TAF15. These proteins bind to RNA and DNA and are implicated in the regulation of gene expression and cellular signalling. Altered protein expression of these ubiquitously expressed proteins has been shown to cause various human cancers [3, 8]. EWS affects cellular growth mechanism like proliferation, migration and invasion by regulating AKT substrate PRAS40 [5] and FAS dependent apoptosis by regulating the exon skipping of FAS/CD90 [9]. EWS plays a major role in mitosis during spindle formation by regulating Aurora B kinase [10, 11]. EWS knockout mice showed defects in B-lymphocyte development, meiosis, spermatocyte development, interferon signalling and HSC dynamics suggesting an important role of EWS in DNA damage response (DDR) [12]. Together, EWS showcases a promising role in several cell survival pathways.

To get a deeper insight into the physiological function of EWS, PAR-CLIP (Photo Activatable Ribonucleoside-enhanced CrossLinking and Immuno-Precipitation), a technique to study the RNA interactome of any RBP of interest [13], was previously applied to EWS revealing its transcriptome-wide RNA targets [14]. Using PAR-CLIP combined with stringent bioinformatic quality criteria we could show that the mRNAs of 4488 genes were directly bound by EWS. We now sought to further investigate the regulated among the many bound targets to uncover the physiological function of EWS. We further identified 116 genes whose expression altered upon EWS down regulation and show that a cell cycle regulator CCDC6 is regulated by EWS by binding to its 3' untranslated region (3'UTR).

Material and Methods

Cell lines and cell culture

For culture conditions of T-REx HEK293 Flp-In cells (Invitrogen), HEK293T cells and cells stably expressing FLAG/HA-tagged EWS please see [14]. Ewing sarcoma cell line MHH-ES-1 was purchased from DSMZ (Germany) [15–17] and grown in RPMI-1640 supplemented with 10% FBS. All of the above-mentioned cell lines were incubated at 37°C and 5% CO₂.

Plasmid constructs

For detailed descriptions of the EWS expression plasmids please refer to [14]. The EWS binding region sequences of the selected mRNA targets were cloned into pSI-CHECK2 vector (Promega) under the control of Renilla luciferase. Primer sequences are given in the [S1 Table](#).

Cell cycle and cell death analysis using flow cytometry

Cells were prepared for cell cycle analysis and viability using propidium iodide staining of the DNA content. Cells were harvested by washing with PBS, trypsinized and pelleted via centrifugation at 1500 rpm for 5 minutes at RT and resuspended directly in the propidium iodide staining solution or Nicoletti buffer (0.1% sodium citrate, TritonX-100 and 50mg/ml of propidium iodide) The cells were incubated in the dark for 15 minutes at 4°C and about 50000 cells were analyzed on the BD FACScanto Flow cytometry machine (BD biosciences). To further distinguish apoptotic cells from necrotic cells, cells were stained with Annexin-V-FITC and counterstained with propidium iodide (PI). The cells were washed in cold PBS and resuspended in Annexin binding buffer at 1×10^6 cells per ml. 2 μ l of Annexin-V-FITC (BD Pharmingen) and 4 μ l of PI (20 μ g/ml) were added to 100 μ l of cell suspension. The cells were incubated in the dark for 15 minutes at RT. After the incubation 100 μ l of Annexin binding buffer was added and the cells were subsequently analyzed.

CCK-8 assay to measure cell proliferation

Cell proliferation assay was performed using CCK-8 kit (Sigma-Aldrich) as per the manufacturer's protocol. Cells were transfected with siRNAs against EWS as well as controls as previously described and seeded into 96 wells. Absorbance at 450nm was measured 4h after the addition of 10 μ l of CCK-8 reagent per well with 1×10^5 cells. Readings were taken at 36, 60 and 84 hours using Tecan Photometer.

mRNA stability assays

The mRNA half-life was determined by treating the cells with 3 μ g/ μ l Actinomycin D (ActD Sigma-Aldrich). Cells were collected at different time points (0–12) and total RNA was isolated followed by cDNA synthesis. Real time PCR, to measure the percentage of RNA remaining for EWS, CCDC6 along with β -actin as an internal control was performed. A linear plot of the percentage RNA remaining and time of ActD treatment was plotted to calculate the mRNA decay constant.

Statistical analysis

All experiments were performed at least in triplicates. Numerical data were expressed as mean \pm SEMs. Group comparisons were analyzed by two way ANOVA or unpaired *t* test. P values were calculated and a p value of <0.05 was considered significant.

Results and Discussion

RNA-binding protein EWS binds and regulates the expression of 116 target mRNAs in HEK293T cells

In order to study the transcriptome of FET family of RNA binding proteins, we previously performed PAR-CLIP on all the three proteins and showed that mRNAs of 4488 genes were bound by EWS [14] To now identify the regulated among all bound mRNAs, we performed siRNA mediated knockdown experiments of EWS in HEK293T cells and analyzed mRNA expression changes using microarrays (S1B Fig.) In total, we found 116 (S2 Table) regulated genes with a corrected p-value of <0.05 (32 at <0.01), which had more than one PAR-CLIP cluster and whose expression level changed (either up or down) by at least 50% upon knock-down of EWS compared to controls.

The regulated genes included tumor-associated genes such as *LIN28B* (regulates let-7 miRNA processing) [18], *PURB* (controls both DNA replication and transcription, its deletion has been associated with acute myeloid leukemia) [19], *CCDC6* (a potential tumor suppressor) [20], *CDCA4* (regulates E2F-dependent transcriptional activation and cell proliferation, mainly through the E2F/retinoblastoma pathway) [21], *MDM2* (E3 ubiquitin protein ligase and repressor of transcription factor p53) [22–24] and *FGF9* (roles in tissue repair, tumor growth and invasion) [25, 26]. Functional annotation using DAVID analysis [27, 28] showed that several genes are associated with diverse functions ranging from response to DNA damage, response to cellular stress, cell cycle process, translational initiation and cellular differentiation.

For our initial analysis we focused on the four highly regulated targets *CCDC6* (log fold change of -1.10), *MDM2* (-0.53), *FGF9* (-0.66) and *CBFB* (-0.58) which were all downregulated upon knockdown of EWS.

Analysis of the PAR-CLIP cluster binding localization revealed that these 116 regulated mRNAs were preferentially bound in the 3'UTR (60%) and had less intronic clusters compared to those of all bound 4488 mRNAs (40%). This effect was even more pronounced in the four chosen targets (75% 3'UTR binding) (Fig. 1B).

EWS regulates a subset of genes with roles in several human cancers and in other cell survival mechanisms

The regulation of the four selected targets was further confirmed using qRT-PCR analysis. HEK293T cells which were treated with siRNA targeting EWS showed clear reduction in the expression of *CBFB*, *CCDC6*, *MDM2* and *FGF9* (Fig. 1C). Downregulation of these genes could further have an effect on genes they regulate and signaling pathways they are involved in. For example, *MDM2* is a protein which inhibits activation of p53 [24], any downregulation in the levels of *MDM2* disturbs the functional interactions of these networks which further affects tissue homeostasis [22] such as the notch induced p53 activity [29, 30]. Another target, *FGF9*, which was reported to be downregulated and mutated in several carcinomas plays an important role in activating *FGFR* signaling. This signaling pathway plays a role in differentiation and inhibits growth thus acting as a tumor suppressor. Therefore its downregulation upon EWS knockdown could repress its tumor suppressor activity [25, 26]. Similarly, *CBFB* is also frequently mutated in AML and is known to play a role in hematopoietic development [31]. The genes regulated by *CBFB* are important in chromatin deacetylation and promoter methylation and thus playing a role in gene transcriptional activation or repression. Therefore its downregulation could disrupt the normal pathways in hematopoiesis and also effect gene regulation. Future experiments are necessary to study in detail the interactions of EWS with these genes and the complex pathways they regulate.

Since ectopic expression of EWS/FLI1 resulted in growth arrest and apoptosis rather than promoting cellular transformation in cells [32] as well as in mice [33], we next sought to look at the genes regulated by EWS with roles in cellular proliferation, cell cycle and cell death. Several of the EWS up-regulated targets and down-regulated targets had roles in cell cycle and G1/S transition like *SKP2* (p45), *BCAT1*, *CCDC6*, *CUL2*, *PPP1CB*, *RHOA* etc. Genes with roles in proliferation and cell death included *DNAJA2*, *MDM2*, *NOP2*, *FGF9*, *ID2* etc. Others like *AEN*, *APTX*, *TOP2A* and *UBE3E* had roles in DNA damage response. We further narrowed down our focus onto *CCDC6*, Coiled-Coil Domain Containing 6 gene since it showed high fold change in the microarray analysis and a high number of PAR-CLIP clusters. *CCDC6* is also known to have several cell cycle associated functions including DNA damage response [34, 35], cell cycle regulation by controlling the intra-S-Phase and G2/M checkpoints. Knockdown of *CCDC6* also showed increased apoptosis and decreased proliferation [36, 37].

CCDC6 is regulated by EWS on RNA and protein level

CCDC6 is a ubiquitously expressed protein which is downregulated in several cancers and predicted to be a tumor suppressor [20]. It is known to be frequently rearranged with RET protein in papillary thyroid carcinomas [38, 39] and fuses to platelet derived growth factor receptor beta gene causing atypical chronic myeloid leukemia [40].

RNA-binding proteins are known to regulate gene expression via post transcriptional mechanisms by stabilizing the mRNA they bind to. We wanted to explore whether EWS stabilizes CCDC6 by binding to it. We performed mRNA stability assay by treating the cells with Actinomycin D to inhibit de novo RNA synthesis and found that the half-life of CCDC6 was ~4.1 h in control cells, whereas in cells treated with EWS siRNA the half-life decreased to ~3.2 h. Together, these data suggest that EWS stabilizes CCDC6 transcript (Fig. 1D).

We also performed luciferase assays to confirm the regulation of CCDC6 mRNA by EWS protein. We cloned the mRNA target sequence carrying the PAR-CLIP clusters downstream of the 3'UTR of Renilla luciferase, and used firefly luciferase as an internal control in psiCHECK-2 plasmid. Upon co-transfection of both plasmids we observed an increase of luciferase activity compared to the empty vector control due to the increase in the expression of CCDC6 UTR (Fig. 1E). We then performed the luciferase assays with different doses of EWS and a clear dose-dependent effect was observed which further confirmed the specificity of EWS mediated regulation of CCDC6 (S1C Fig.).

For our further experiments, we chose MHH-ES-1, a Ewing sarcoma cell line which carries the most frequently occurring translocation EWS-FLI. qRT-PCR results showed downregulation of CCDC6 mRNA levels upon knockdown of EWS. The siRNA is designed to target the c-terminal region of EWSR1 which is absent in EWS-FLI1 fusion transcript in MHH-ES-1 (Fig. 1F). The same effect was also observed in HEK293T cells (S1D Fig.). Also, Western blot showed decreased expression of CCDC6 protein upon EWS knockdown in MHH-ES-1 (Fig. 1G) and also in HEK293T cells (S1D Fig.) thus showing that the regulation of CCDC6 by EWS extends to the protein level. Since this regulation might further have an impact on several other downstream genes that CCDC6 itself regulates, we tested one of these published targets, namely FBXW7, which is required for DNA damage response [41], and indeed, qRT-PCR on FBXW7 following knockdown of EWS showed reduced RNA levels (Fig. 2A). Since FBXW7 is mutated in several cancers and shown to be a tumor suppressor, the pathway through which all these proteins inter-regulate might further play a role in sarcomagenesis [42]. Further studies are required to elucidate the role of FBXW7 in Ewing sarcoma context.

Taken together, these results suggest that EWS regulates the expression levels of CCDC6 by stabilizing it and further exerting its effect on the expression of its downstream targets.

No rescue effect by FET family members in Ewing sarcoma cell line

It was previously shown that FET proteins show redundancy and that knockdown of EWS upregulates its two family members FUS and TAF15 in HEK293 cells as well as in liposarcoma cell lines [43]. To address whether this mechanism might also lessen the effects of EWS haploinsufficiency in Ewing sarcoma, we measured FUS and TAF15 mRNA levels upon EWS knockdown in MHH-ES-1 cells. However, there was no increase in mRNA levels for either of the two genes (Fig. 2B). This indicates that, probably due to a different cellular context, in Ewing sarcoma there is no rescue mechanism for loss of EWS expression.

Downregulation of EWS affects viability, proliferation and cell cycle

Gene set enrichment analysis and DAVID analysis revealed that several targets regulated by EWS play a role in regulating cell proliferation and cell cycle. Also, EWS was reported to

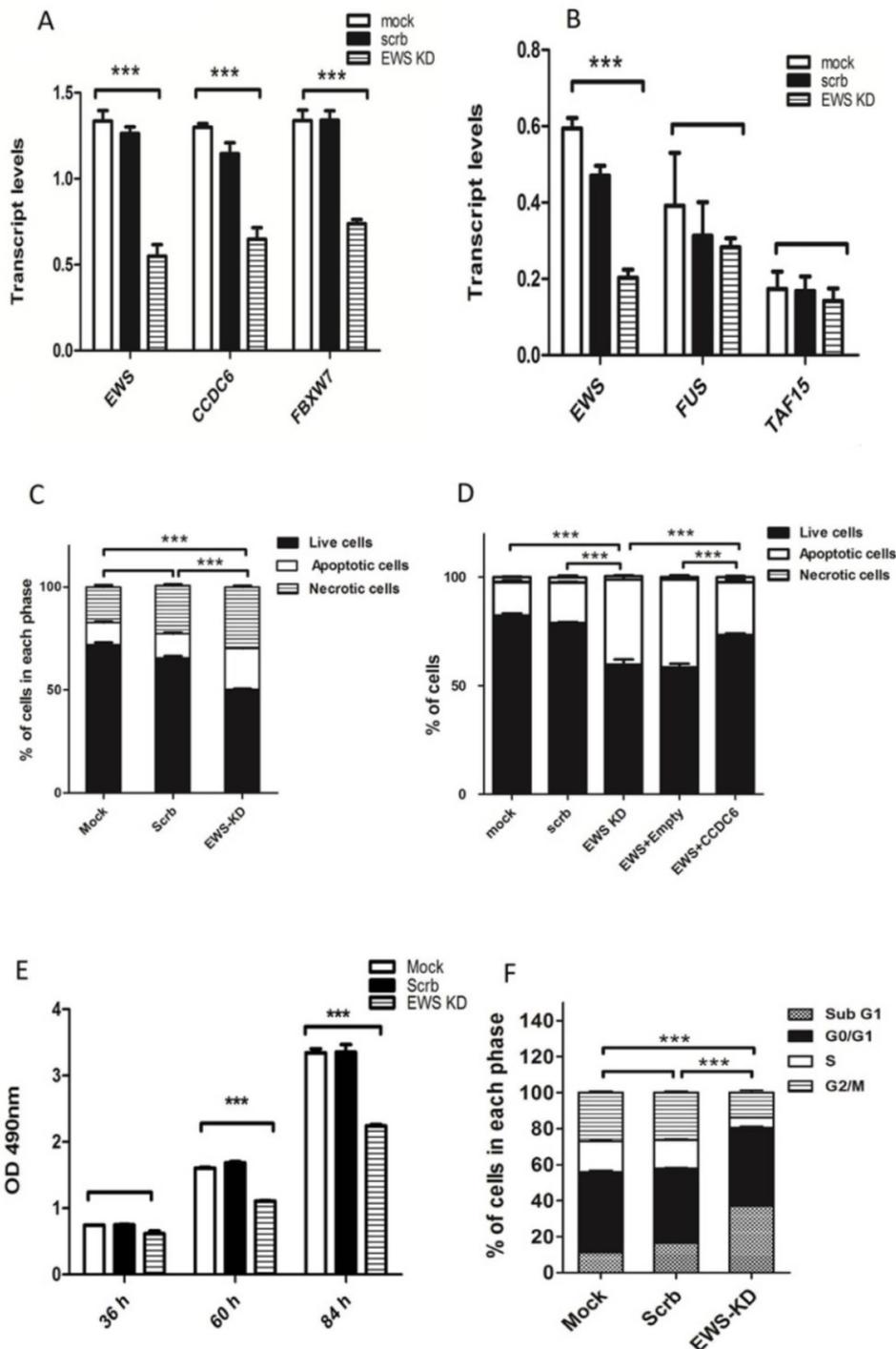


Fig 2. EWS downregulation affects apoptosis, cell cycle and proliferation. **A.** Decrease in the relative mRNA levels of FBXW7 upon knocking down EWS. **B.** Relative mRNA levels of FUS, EWS and TAF15 (FET family proteins) upon EWS knock down. **C.** Total percentage of living, necrotic and apoptotic cells after EWS and scrambled siRNA knockdown are represented on the bars. Apoptotic cells are defined by the sum of population of cells in early apoptosis and late apoptosis. Mock, scrambled and EWS KD had 10%, 11.8% and 20% apoptotic cells and 17%, 22% and 30% of necrotic cells respectively. The P values refer to the apoptotic cell population. **D.** Rescue effect upon co-transfection of EWS siRNA and CCDC6 overexpression. Bars represent total

percentage of living, apoptotic and necrotic cells. 50nM siRNA, 100ng of empty vector and 100ng of CCDC6 expression vector were transfected. Mock and scrambled had 15% and 18.6% of apoptotic cells, EWS KD and EWS KD+ Empty vector had 39% and 40.5% apoptotic cells respectively and EWS KD+ CCDC6 vector had only 24% apoptotic cells. The P values refer to the apoptotic cell population. **E.** Proliferation rates on three consecutive days using CCK8 assay was calculated by measuring the absorbance which is proportional to the amount of living cells. **F.** Quantification of cell cycle distribution. The bars indicate the % of cells in each cell cycle phase subG1, G0/G1, S and G2/M phase. Mock has 11.3%, 44.7%, 16%, 28% of cells in subG1, G0/G1, S and G2/M phase respectively. 16%, 42%, 15%, 27% of cells for scrambled and 37%, 44%, 5%, 14% of cells for EWS KD in each phase respectively. The calculated P values refer to the cells in S phase. Data in all the above figures (A-F) is presented as mean SEM values (* $P < 0.05$; where $n = 3$ per group).

doi:10.1371/journal.pone.0119066.g002

maintain mitotic integrity and proneural cell survival in zebra fish by regulating aurora B [10]. Earlier studies reported that CCDC6 silencing increased apoptosis, decreased cell proliferation and affected cell cycle division by controlling the intra S phase and G2/M checkpoints [36]. TAF15, another member of the FET family of proteins was also shown to regulate cell proliferation and apoptosis *in vivo* [44].

To confirm whether downregulation of EWS also affects those processes, we assessed apoptosis using Annexin V FITC and PI double staining of EWS knockdown and control HEK293T cells. We recorded an increase in the percentage of apoptotic cells in EWS siRNA treated cells (63% living cells, 27% apoptotic and 10% necrotic cells) compared to controls (74% living, 13% apoptotic and 13% necrotic cells) (Fig. 2C) while the % necrotic cells remained unchanged. We also confirmed increased cell death using a trypan blue dye exclusion assay by counting the cell number of dead and vital cells on three consecutive days (data not shown). These results further confirmed the previous finding that EWS knockdown induces apoptosis [10]. We next performed gain of function experiments to see if up regulation of CCDC6 after EWS knockdown will rescue this phenotype. To do this, we co-transfected the cells with EWS siRNA and CCDC6 overexpression vector and found that the apoptotic rate were 14% less in CCDC6 overexpressed cells compared to empty vector (Fig. 2D), confirming that overexpression of CCDC6 upon EWS knockdown indeed rescued the observed phenotype.

Apoptotic signals are coupled to growth regulatory processes such as proliferation, cell cycle arrest, and cellular differentiation. Therefore to examine if the observed apoptosis affected the proliferation rate of the cells, we measured the proliferation of EWS knockdown and control cells using CCK8 assay. We indeed noticed that the proliferation rate was remarkably declined upon downregulation of EWS compared to those transfected with no siRNA transfection and scrambled siRNA transfection (Fig. 2E).

Given that EWS regulates the expression of CCDC6, and that EWS downregulation induces apoptosis, and CCDC6 has been implicated in apoptosis coupled to S and G2/M phase cell cycle defects, we next tested cell cycle progression of the cells by propidium iodide staining after 48 hours of transfection. As shown in Fig. 2F, we observed a decrease of cells in S and G2/M phase upon EWS knockdown and an increased cell number with a sub G0/G1 DNA content typical of apoptotic cells. This indicated increased apoptosis of cells that had passed the G1 checkpoint. Higher apoptotic rate might have likely affected the duration of cell cycle or the doubling time of the cells. The observed decrease of cell number in S and G2/M phase might be due to the fact that CCDC6 regulates cell cycle checkpoints like 14-3-3 σ and CDC25C [37] which are important for S phase duration, transition into G2 phase and activation of mitosis. Cell cycle checkpoints safeguard the cells from accumulating genetic errors and any deregulation will prepare the ground for increased mutagenesis and onset of several cancers. Therefore the aberrant entrance of cells into next phases by skipping the checkpoint controls regulated by CCDC6 might further triggers unwanted mutations leading to genetic aberrations.

Taken together our results demonstrate that EWS controls apoptosis and cell cycle progression by regulating one of its key mRNA target CCDC6. EWS therefore regulates a wide range

of cellular processes to ensure genome integrity and cellular homeostasis. Recent studies have unveiled new roles of EWS in gene regulation and RNA metabolism. This suggests that highly complex roles are played by EWS and a thorough elucidation of the entire mRNA target network of native EWS is essential in understanding the molecular and cellular biology of Ewing sarcoma. Future studies will show whether EWS and/or its target mRNAs will offer new therapeutic approaches.

Supporting Information

S1 Fig. Location of siRNA on EWS (A). The arrow indicates the RRM region of EWSR1 targeted by siRNA thus exclusively targeting only the non translocated allele. **siRNA mediated knockdown of EWS for microarray analysis (B).** Bars indicate transcript levels as measured using Affymetrix U133 Plus 2.0 arrays showing efficient EWS knockdown on mRNA level. Signal intensities were calculated by averaging redundant probe sets for the same gene. Error bars indicate standard error of mean (SEM). *: $P < 0.001$. β actin was included for comparison. a.u., arbitrary units; ctrl., transfection of control siRNA. **Dose dependent regulation of CCDC6 by EWS in luciferase assay (C).** Regulation of CCDC6 by EWS was further tested with increasing concentrations of EWS by measuring the relative luciferase activity. On the x axis the alphabets indicate the plasmids that were co transfected along with 50 ng psiCHECK-2-CCDC6 and the following co-plasmids accordingly. a) 25 ng of pDEST-EWS b) 50 ng of pDEST-EWS, c) 75 ng of pDEST-EWS d) 75 ng of empty pDEST e) 25 ng of empty pDEST f) 50 ng of empty pDEST g) 75 ng of empty pDEST h) 75 empty psiCHECK-2. **Downregulation of CCDC6 following EWS knockdown in HEK 293T cells (D).** Relative mRNA levels of CCDC6 and EWS in wild type, control and EWS knockdown in HEK293T cells. Relative mRNA levels were normalized to beta actin. Western blot showing the downregulation of CCDC6 upon EWS knockdown in HEK293T cells. Antibodies are indicated. (DOC)

S1 Table. Sequences for primers and siRNA.
(XLS)

S2 Table. List of genes regulated by EWS.
(XLS)

Acknowledgments

We would like to thank Thomas Tuschl and Ute Fischer for helpful comments on the manuscript.

Author Contributions

Conceived and designed the experiments: JIH SD. Performed the experiments: SD JIH SN. Analyzed the data: EL SD. Wrote the paper: SD. Supervised the project: AB JIH.

References

- Rodríguez-Galindo C, Spunt SL, Pappo AS. Treatment of Ewing sarcoma family of tumors: current status and outlook for the future. *Med Pediatr Oncol.* 2003; 40(5):276–87. PMID: [12652615](#)
- Esiashvili N, Goodman M, Marcus RB Jr. Changes in incidence and survival of Ewing sarcoma patients over the past 3 decades: Surveillance Epidemiology and End Results data. *J Pediatr Hematol Oncol.* 2008; 30(6):425–30. doi: [10.1097/MPH.0b013e31816e22f3](#) PMID: [18525458](#)
- Xia SJ, Barr FG. Chromosome translocations in sarcomas and the emergence of oncogenic transcription factors. *Eur J Cancer.* 2005; 41(16):2513–27. PMID: [16213703](#)

4. Kovar H. Dr. Jekyll and Mr. Hyde: The Two Faces of the FUS/EWS/TAF15 Protein Family. *Sarcoma*. 2011; 2011:837474. doi: [10.1155/2011/837474](https://doi.org/10.1155/2011/837474) PMID: [21197473](https://pubmed.ncbi.nlm.nih.gov/21197473/)
5. Huang L, Nakai Y, Kuwahara I, Matsumoto K. PRAS40 is a functionally critical target for EWS repression in Ewing sarcoma. *Cancer research*. 2012; 72(5):1260–9. doi: [10.1158/0008-5472.CAN-11-2254](https://doi.org/10.1158/0008-5472.CAN-11-2254) PMID: [22241085](https://pubmed.ncbi.nlm.nih.gov/22241085/)
6. Lin PP, Pandey MK, Jin F, Xiong S, Deavers M, Parant JM, et al. EWS-FLI1 induces developmental abnormalities and accelerates sarcoma formation in a transgenic mouse model. *Cancer research*. 2008; 68(21):8968–75. doi: [10.1158/0008-5472.CAN-08-0573](https://doi.org/10.1158/0008-5472.CAN-08-0573) PMID: [18974141](https://pubmed.ncbi.nlm.nih.gov/18974141/)
7. Glisovic T, Bachorik JL, Yong J, Dreyfuss G. RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett*. 2008; 582(14):1977–86. doi: [10.1016/j.febslet.2008.03.004](https://doi.org/10.1016/j.febslet.2008.03.004) PMID: [18342629](https://pubmed.ncbi.nlm.nih.gov/18342629/)
8. Tan AY, Manley JL. The TET family of proteins: functions and roles in disease. *J Mol Cell Biol*. 2009; 1(2):82–92. doi: [10.1093/jmcb/mjp025](https://doi.org/10.1093/jmcb/mjp025) PMID: [19783543](https://pubmed.ncbi.nlm.nih.gov/19783543/)
9. Paronetto MP, Bernardis I, Volpe E, Bechara E, Sebestyen E, Eyraes E, et al. Regulation of FAS exon definition and apoptosis by the Ewing sarcoma protein. *Cell reports*. 2014; 7(4):1211–26. doi: [10.1016/j.celrep.2014.03.077](https://doi.org/10.1016/j.celrep.2014.03.077) PMID: [24813895](https://pubmed.ncbi.nlm.nih.gov/24813895/)
10. Azuma M, Embree LJ, Sabaawy H, Hickstein DD. Ewing sarcoma protein *ewsr1* maintains mitotic integrity and proneural cell survival in the zebrafish embryo. *PLoS one*. 2007; 2(10):e979. PMID: [17912356](https://pubmed.ncbi.nlm.nih.gov/17912356/)
11. Park H, Turkalo TK, Nelson K, Folmsbee SS, Robb C, Roper B, et al. Ewing sarcoma EWS protein regulates midzone formation by recruiting Aurora B kinase to the midzone. *Cell cycle*. 2014; 13(15).
12. Paronetto MP. Ewing sarcoma protein: a key player in human cancer. *International journal of cell biology*. 2013; 2013:642853. doi: [10.1155/2013/642853](https://doi.org/10.1155/2013/642853) PMID: [24082883](https://pubmed.ncbi.nlm.nih.gov/24082883/)
13. Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*. 2010; 141(1):129–41. doi: [10.1016/j.cell.2010.03.009](https://doi.org/10.1016/j.cell.2010.03.009) PMID: [20371350](https://pubmed.ncbi.nlm.nih.gov/20371350/)
14. Hoell JI, Larsson E, Runge S, Nusbaum JD, Duggimpudi S, Farazi TA, et al. RNA targets of wild-type and mutant FET family proteins. *Nat Struct Mol Biol*. 18(12):1428–31. doi: [10.1038/nsmb.2163](https://doi.org/10.1038/nsmb.2163) PMID: [22081015](https://pubmed.ncbi.nlm.nih.gov/22081015/)
15. Kaufmann B, Muller S, Hanisch FG, Hartmann U, Paulsson M, Maurer P, et al. Structural variability of BM-40/SPARC/osteonectin glycosylation: implications for collagen affinity. *Glycobiology*. 2004; 14(7):609–19. PMID: [15044389](https://pubmed.ncbi.nlm.nih.gov/15044389/)
16. Garcia Gimenez D, Garcia Prado E, Saenz Rodriguez T, Fernandez Arche A, De la Puerta R. Cytotoxic effect of the pentacyclic oxindole alkaloid mitraphylline isolated from *Uncaria tomentosa* bark on human Ewing's sarcoma and breast cancer cell lines. *Planta medica*. 2010; 76(2):133–6. doi: [10.1055/s-0029-1186048](https://doi.org/10.1055/s-0029-1186048) PMID: [19724995](https://pubmed.ncbi.nlm.nih.gov/19724995/)
17. Ulaner GA, Hoffman AR, Otero J, Huang HY, Zhao Z, Mazumdar M, et al. Divergent patterns of telomere maintenance mechanisms among human sarcomas: sharply contrasting prevalence of the alternative lengthening of telomeres mechanism in Ewing's sarcomas and osteosarcomas. *Genes, chromosomes & cancer*. 2004; 41(2):155–62.
18. Viswanathan SR, Daley GQ, Gregory RI. Selective blockade of microRNA processing by Lin28. *Science*. 2008; 320(5872):97–100. doi: [10.1126/science.1154040](https://doi.org/10.1126/science.1154040) PMID: [18292307](https://pubmed.ncbi.nlm.nih.gov/18292307/)
19. Lezon-Geyda K, Najfeld V, Johnson EM. Deletions of PURA, at 5q31, and PURB, at 7p13, in myelodysplastic syndrome and progression to acute myelogenous leukemia. *Leukemia*. 2001; 15(6):954–62. PMID: [11417483](https://pubmed.ncbi.nlm.nih.gov/11417483/)
20. Puxeddu E, Zhao G, Stringer JR, Medvedovic M, Moretti S, Fagin JA. Characterization of novel non-clonal intrachromosomal rearrangements between the H4 and PTEN genes (H4/PTEN) in human thyroid cell lines and papillary thyroid cancer specimens. *Mutation research*. 2005; 570(1):17–32. PMID: [15680400](https://pubmed.ncbi.nlm.nih.gov/15680400/)
21. Hayashi R, Goto Y, Ikeda R, Yokoyama KK, Yoshida K. CDCA4 is an E2F transcription factor family-induced nuclear factor that regulates E2F-dependent transcriptional activation and cell proliferation. *J Biol Chem*. 2006; 281(47):35633–48. PMID: [16984923](https://pubmed.ncbi.nlm.nih.gov/16984923/)
22. Wade M, Li YC, Wahl GM. MDM2, MDMX and p53 in oncogenesis and cancer therapy. *Nature reviews Cancer*. 2013; 13(2):83–96. doi: [10.1038/nrc3430](https://doi.org/10.1038/nrc3430) PMID: [23303139](https://pubmed.ncbi.nlm.nih.gov/23303139/)
23. Li Q, Lozano G. Molecular pathways: targeting Mdm2 and Mdm4 in cancer therapy. *Clinical cancer research: an official journal of the American Association for Cancer Research*. 2013; 19(1):34–41. doi: [10.1158/1078-0432.CCR-12-0053](https://doi.org/10.1158/1078-0432.CCR-12-0053) PMID: [23262034](https://pubmed.ncbi.nlm.nih.gov/23262034/)
24. Haupt Y, Maya R, Kazaz A, Oren M. Mdm2 promotes the rapid degradation of p53. *Nature*. 1997; 387(6630):296–9. PMID: [9153395](https://pubmed.ncbi.nlm.nih.gov/9153395/)
25. Krejci P, Prochazkova J, Bryja V, Kozubik A, Wilcox WR. Molecular pathology of the fibroblast growth factor family. *Human mutation*. 2009; 30(9):1245–55. doi: [10.1002/humu.21067](https://doi.org/10.1002/humu.21067) PMID: [19621416](https://pubmed.ncbi.nlm.nih.gov/19621416/)

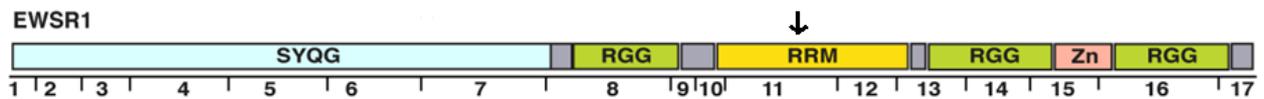
26. Abdel-Rahman WM, Kalinina J, Shoman S, Eissa S, Ollikainen M, Elomaa O, et al. Somatic FGF9 mutations in colorectal and endometrial carcinomas associated with membranous beta-catenin. *Human mutation*. 2008; 29(3):390–7. doi: [10.1002/humu.20653](https://doi.org/10.1002/humu.20653) PMID: [18165946](https://pubmed.ncbi.nlm.nih.gov/18165946/)
27. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols*. 2009; 4(1):44–57. doi: [10.1038/nprot.2008.211](https://doi.org/10.1038/nprot.2008.211) PMID: [19131956](https://pubmed.ncbi.nlm.nih.gov/19131956/)
28. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids research*. 2009; 37(1):1–13. doi: [10.1093/nar/gkn923](https://doi.org/10.1093/nar/gkn923) PMID: [19033363](https://pubmed.ncbi.nlm.nih.gov/19033363/)
29. Pishas KI, Al-Ejeh F, Zinonos I, Kumar R, Evdokiou A, Brown MP, et al. Nutlin-3a is a potential therapeutic for ewing sarcoma. *Clinical cancer research: an official journal of the American Association for Cancer Research*. 2011; 17(3):494–504. doi: [10.1158/1078-0432.CCR-10-1587](https://doi.org/10.1158/1078-0432.CCR-10-1587) PMID: [21098696](https://pubmed.ncbi.nlm.nih.gov/21098696/)
30. Li Y, Tanaka K, Fan X, Nakatani F, Li X, Nakamura T, et al. Inhibition of the transcriptional function of p53 by EWS-Fli1 chimeric protein in Ewing Family Tumors. *Cancer letters*. 2010; 294(1):57–65. doi: [10.1016/j.canlet.2010.01.022](https://doi.org/10.1016/j.canlet.2010.01.022) PMID: [20153576](https://pubmed.ncbi.nlm.nih.gov/20153576/)
31. Sangle NA, Perkins SL. Core-binding factor acute myeloid leukemia. *Archives of pathology & laboratory medicine*. 2011; 135(11):1504–9.
32. Deneen B, Denny CT. Loss of p16 pathways stabilizes EWS/FLI1 expression and complements EWS/FLI1 mediated transformation. *Oncogene*. 2001; 20(46):6731–41. PMID: [11709708](https://pubmed.ncbi.nlm.nih.gov/11709708/)
33. Sohn EJ, Li H, Reidy K, Beers LF, Christensen BL, Lee SB. EWS/FLI1 oncogene activates caspase 3 transcription and triggers apoptosis in vivo. *Cancer research*. 2010; 70(3):1154–63. doi: [10.1158/0008-5472.CAN-09-1993](https://doi.org/10.1158/0008-5472.CAN-09-1993) PMID: [20103643](https://pubmed.ncbi.nlm.nih.gov/20103643/)
34. Merolla F, Luise C, Muller MT, Pacelli R, Fusco A, Celetti A. Loss of CCDC6, the first identified RET partner gene, affects p53 levels and accelerates mitotic entry upon DNA damage. *PloS one*. 2012; 7(5):e36177. doi: [10.1371/journal.pone.0036177](https://doi.org/10.1371/journal.pone.0036177) PMID: [22655027](https://pubmed.ncbi.nlm.nih.gov/22655027/)
35. Leone V, Mansueto G, Pierantoni GM, Tomincasa M, Merolla F, Cerrato A, et al. CCDC6 represses CREB1 activity by recruiting histone deacetylase 1 and protein phosphatase 1. *Oncogene*. 2010; 29(30):4341–51. doi: [10.1038/onc.2010.179](https://doi.org/10.1038/onc.2010.179) PMID: [20498639](https://pubmed.ncbi.nlm.nih.gov/20498639/)
36. Thanasopoulou A, Stravopodis DJ, Dimas KS, Schwaller J, Anastasiadou E. Loss of CCDC6 affects cell cycle through impaired intra-S-phase checkpoint control. *PloS one*. 2012; 7(2):e31007. doi: [10.1371/journal.pone.0031007](https://doi.org/10.1371/journal.pone.0031007) PMID: [22363533](https://pubmed.ncbi.nlm.nih.gov/22363533/)
37. Thanasopoulou A, Xanthopoulou AG, Anagnostopoulos AK, Konstantakou EG, Margaritis LH, Papsideri IS, et al. Silencing of CCDC6 reduces the expression of 14–3–3sigma in colorectal carcinoma cells. *Anticancer research*. 2012; 32(3):907–13. PMID: [22399611](https://pubmed.ncbi.nlm.nih.gov/22399611/)
38. Kondo T, Ezzat S, Asa SL. Pathogenetic mechanisms in thyroid follicular-cell neoplasia. *Nature reviews Cancer*. 2006; 6(4):292–306. PMID: [16557281](https://pubmed.ncbi.nlm.nih.gov/16557281/)
39. Celetti A, Cerrato A, Merolla F, Vitagliano D, Vecchio G, Grieco M. H4(D10S170), a gene frequently rearranged with RET in papillary thyroid carcinomas: functional characterization. *Oncogene*. 2004; 23(1):109–21. PMID: [14712216](https://pubmed.ncbi.nlm.nih.gov/14712216/)
40. Schwaller J, Anastasiadou E, Cain D, Kutok J, Wojcicki S, Williams IR, et al. H4(D10S170), a gene frequently rearranged in papillary thyroid carcinoma, is fused to the platelet-derived growth factor receptor beta gene in atypical chronic myeloid leukemia with t(5;10)(q33;q22). *Blood*. 2001; 97(12):3910–8. PMID: [11389034](https://pubmed.ncbi.nlm.nih.gov/11389034/)
41. Zhao J, Tang J, Men W, Ren K. FBXW7-mediated degradation of CCDC6 is impaired by ATM during DNA damage response in lung cancer cells. *FEBS Lett*. 2012; 586(24):4257–63. doi: [10.1016/j.febslet.2012.10.029](https://doi.org/10.1016/j.febslet.2012.10.029) PMID: [23108047](https://pubmed.ncbi.nlm.nih.gov/23108047/)
42. Wang Z, Inuzuka H, Zhong J, Wan L, Fukushima H, Sarkar FH, et al. Tumor suppressor functions of FBW7 in cancer development and progression. *FEBS Lett*. 2012; 586(10):1409–18. doi: [10.1016/j.febslet.2012.03.017](https://doi.org/10.1016/j.febslet.2012.03.017) PMID: [22673505](https://pubmed.ncbi.nlm.nih.gov/22673505/)
43. Spitzer JL, Ugras S, Runge S, Decarolis P, Antonescu C, Tuschl T, et al. mRNA and protein levels of FUS, EWSR1, and TAF15 are upregulated in liposarcoma. *Genes, chromosomes & cancer*. 2011; 50(5):338–47.
44. Ballarino M, Jobert L, Dembele D, de la Grange P, Auboeuf D, Tora L. TAF15 is important for cellular proliferation and regulates the expression of a subset of cell cycle genes through miRNAs. *Oncogene*. 2013; 32(39):4646–55. doi: [10.1038/onc.2012.490](https://doi.org/10.1038/onc.2012.490) PMID: [23128393](https://pubmed.ncbi.nlm.nih.gov/23128393/)

6. Supplementary information for publication I

6.1 Supplementary figures

S1.A Fig.

Location of siRNA on EWSR1.

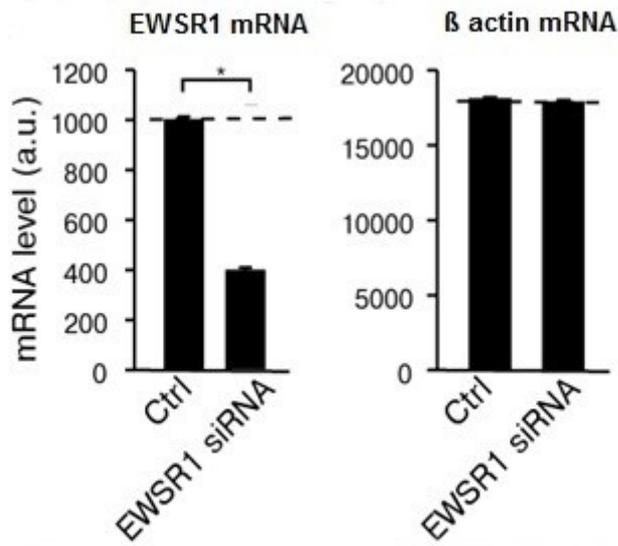


The arrow indicates the RRM region of EWSR1 targeted by siRNA thus exclusively targeting only the non translocated allele.

S1.B Fig.

siRNA mediated knockdown of EWS for microarray analysis.

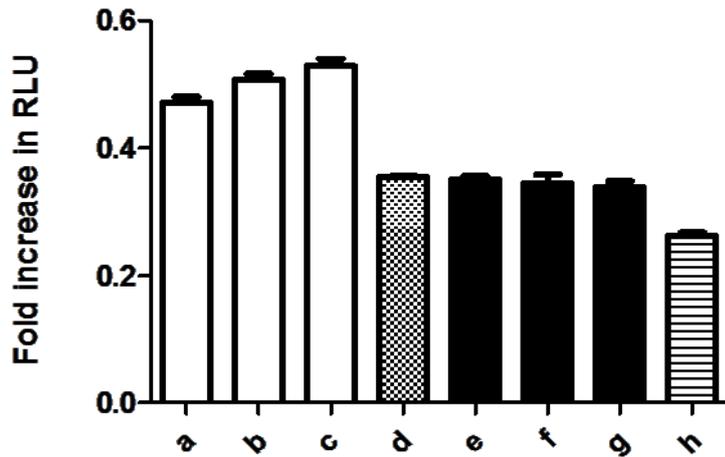
Bars indicate transcript levels as measured using Affymetrix U133 Plus 2.0 arrays showing efficient EWS knockdown on mRNA level. Signal intensities were calculated by averaging redundant probe sets for the same gene. Error bars indicate standard error of mean (SEM). *: $P < 0.001$. β actin was included for comparison. a.u., arbitrary units; ctrl., transfection of control siRNA.



S1.C Fig.

Dose dependent regulation of CCDC6 by EWS in luciferase assay.

Regulation of CCDC6 by EWS was further tested with increasing concentrations of EWS by measuring the relative luciferase activity. On the x axis the alphabets indicate the plasmids that were co transfected along with 50 ng psiCHECK-2-CCDC6 and the following co-plasmids accordingly. a) 25 ng of pDEST-EWS b) 50 ng of pDEST-EWS, c) 75 ng of pDEST-EWS d) 75 ng of empty pDEST e) 25 ng of empty pDEST f) 50 ng of empty pDEST g) 75 ng of empty pDEST h) 75 empty psiCHECK-2.

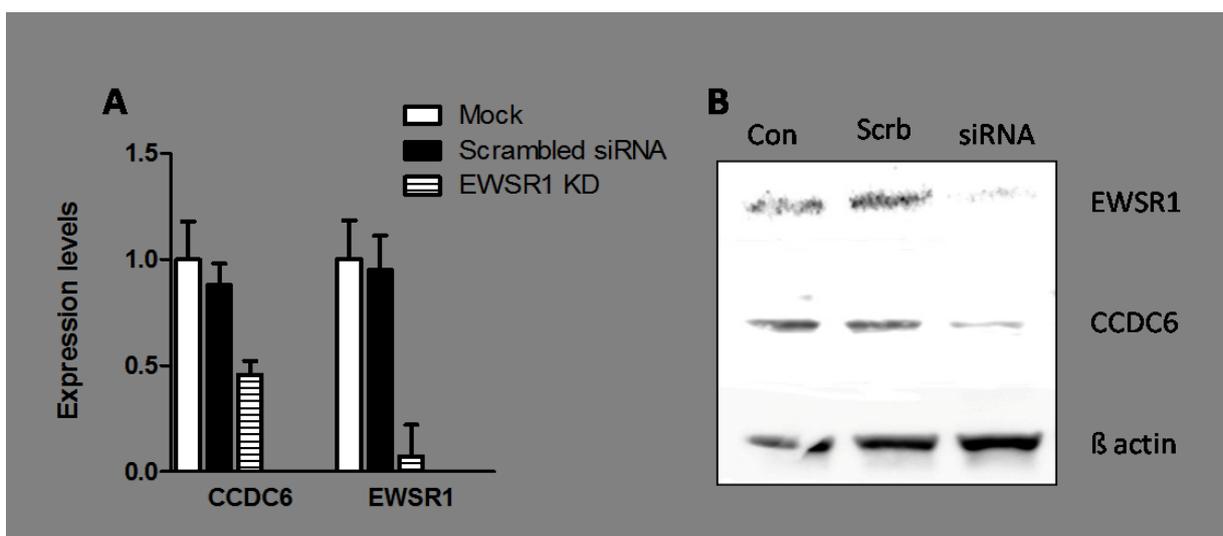


S1.D Fig.

Downregulation of CCDC6 following EWS knockdown in HEK 293T cells.

A) Relative mRNA levels of CCDC6 and EWS in wild type, control and EWS knockdown in HEK293T cells. Relative mRNA levels were normalized to beta actin.

B) Western blot showing the downregulation of CCDC6 upon EWS knockdown in HEK293T cells. Antibodies are indicated.



6.2 Supplementary tables

Supplementary table 1

Primers	Sequence
CCDC6 RT PCR primer Fwd	ACCATCCAAGCCAGGGCTGA
CCDC6 RT PCR Primer Rev	CAAGATGCTGTTCTAGTTCGGC
CBFB RT PCR primer Fwd	TGTGAGATTAAGTACACGG
CBFB RT PCR primer Rev	TAATGCATCCTCCTGCTGGGCT
FGF9 RT PCR primer Fwd	CGATTTGGCATTCTGGA
FGF9 RT PCR primer Rev	TAGTCCTAGTCCCTTCTCTCGG
MDM2 RT PCR primer Fwd	AGCAGGAATCATCGGAC
MDM2 RT PCR primer Rev	AGCATCAAGATCCGGATTTCGATGGCG
EWSR1 RT PCR Primer Fwd	CAGCCTCCCCTGGTTATACTACTCCA
EWSR1 RT PCR Primer Rev	CTGCGGTCTTGTAGGTGCAGT
Beta actin RT PCR Primer Fwd	GCACTCTTCCAGCCTTCC
Beta actin RT PCR Primer Rev	CTAGAAGCATTGCGGTG
CCDC6-pSI CHECK2 Fwd	ACGCAGTCGACTCGAGAACTCTTAAATATGCATTCTGA
CCDC6-pSICHECK2 Rev	GCGGCCGCTAAAGGAAAGTAGCACATTAG
CBFB-1pSICHECK2 Fwd	ACGCAGTCGACTCGAGTAGCCTGTTCATTAGAA
CBFB-1 pSICHECK2 Rev	GCGGCCGCTTCTTAAATCATAAAAAGTGTGTA
FGF9 pSICHECK2 Fwd	ACGCAGTCGACTCGAGCCGGTTTTGTTAAGTGACCAC
FGF9-pSICHECK2 Rev	GCGGCCGCTCAACATGTCATCTCATGGAC
MDM2-pSICHECK2 Fwd	ACGCAGTCGACTCGAGGTGCAACTGTGTGTTTTAACCTAG
MDM2 pSICHECK2 Rev	GCGGCCGCCATATATACCAAGGCCACGTAT
siRNA targeting EWSR1 EWSR1-siRNA	GCAGGAGTCTGGAGGATTT

Supplementary table 2

Gene symbol	fold change	p value	PAR-CLIP clusters
ADAM23	-0.56	4.48E-02	2
AEN	-0.69	7.39E-03	2
AKIRIN1	-0.83	1.69E-02	5
AMOT	-1.57	2.63E-03	2
APOLD1	-0.69	4.37E-02	2
APTX	-0.65	2.96E-02	6
ARHGAP11A	-0.6	1.54E-02	7
ARHGAP19	-0.78	1.05E-02	4
ARHGAP19	-0.79	2.75E-03	4
AZIN1	-0.63	1.03E-02	2
B4GALT6	-1.09	2.63E-02	2
BCAT1	-0.81	4.13E-03	12
C10orf18	-0.78	1.37E-04	37
C16orf63	-0.98	1.79E-02	2
C19orf2	-1.01	7.71E-03	8
C22orf30	-0.69	2.23E-02	10
CBFB	-0.58	2.04E-02	2
CCDC6	-1.1	1.23E-02	12
CDCA4	-0.65	6.59E-04	3
CMPK1	-0.84	9.66E-03	2
CPNE3	-0.8	1.61E-02	11
CSNK1E	-1.01	3.32E-03	3
CSNK1G3	-1.22	3.05E-02	6
CUL2	-0.62	2.79E-02	3
DCUN1D5	-0.58	1.54E-02	13
DDX21	-0.59	1.48E-02	8
DNAJA2	-0.59	2.26E-02	3
DPY19L3	-0.91	2.94E-02	4
DR1	-0.59	2.94E-04	3
EIF1AX	-0.79	2.15E-02	3
EIF2S3	-0.81	1.18E-03	2
EIF3J	-0.52	3.86E-02	3
FAM60A	-0.57	3.96E-02	4
FGF9	-0.66	3.21E-02	2
FLNA	-0.57	1.37E-02	5
FUT10	-0.52	2.44E-02	2
GALNT1	-0.9	1.57E-02	3
GATAD2A	-0.59	4.62E-02	2
GNG12	-0.86	1.33E-02	5
ID2	-1.4	2.93E-02	2
IDE	-0.52	1.04E-02	3
IMPAD1	-0.99	1.80E-02	5
INSIG1	-0.55	1.82E-02	2

KBTBD8	-0.64	4.35E-02	3
KCTD12	-1.23	1.30E-02	9
KCTD15	-0.65	1.82E-02	3
KLHL11	-0.64	1.03E-02	6
LIN28B	-1.14	1.44E-02	17
LIN7C	-0.59	2.09E-02	9
LPGAT1	-1.41	1.21E-02	16
LRP12	-0.92	5.28E-03	5
MCART1	-1.04	3.72E-02	2
MDM2	-0.53	2.71E-02	8
MID1	-0.69	5.04E-03	5
MIR17HG	-0.81	8.19E-03	2
NEDD1	-0.83	3.31E-02	4
NIP7	-1.04	1.94E-02	2
NOP2	-0.54	1.45E-02	2
NUFIP2	-0.63	1.72E-02	31
PCGF5	-1.33	1.69E-03	12
PHF20	-0.62	1.54E-02	6
PNMA2	-0.71	4.46E-02	2
PPP1CB	-0.68	6.40E-03	5
PRC1	-0.57	2.72E-03	4
PRNP	-1.36	1.23E-02	3
PSIP1	-0.6	2.66E-02	4
PURB	-0.53	3.51E-02	30
RAB21	-1.05	1.73E-02	6
RAP2A	-0.64	3.10E-02	5
REEP3	-0.65	4.61E-03	2
RHOA	-1.22	4.46E-03	9
RPS4X	-1.13	4.56E-03	2
SACS	-0.68	6.53E-03	29
SCARB1	-1.53	1.53E-04	2
SCML1	-0.72	3.47E-03	8
SF3B3	-0.89	6.91E-03	14
SKP2	-0.81	1.72E-02	2
SLC16A14	-1.21	1.63E-02	5
SLC25A13	-0.73	1.38E-02	2
SLC25A21	-0.67	4.15E-02	2
SLC7A5	-0.96	1.04E-02	2
SLC9A7	-0.65	2.04E-02	2
SOCS6	-0.57	2.48E-04	2
SOX5	-0.62	2.49E-02	2
SSR3	-0.63	4.26E-02	7
STC2	-0.61	5.39E-03	2
SUV39H2	-0.71	3.45E-02	9
SYAP1	-0.53	1.11E-02	3

TBC1D5	-0.93	3.44E-02	3
TCFL5	-0.67	2.33E-02	3
TMED10	-0.8	1.19E-02	4
TMEM33	-0.59	1.01E-02	16
TMEM65	-0.8	2.76E-02	2
TNFRSF10B	-0.68	2.52E-02	4
TOP2A	-0.64	2.52E-02	13
UBE2D1	-1.12	3.90E-04	3
UBE2V2	-1.43	2.69E-02	4
USP28	-0.53	2.54E-02	3
ZFAND1	-0.9	2.06E-02	2
ZNF711	-0.95	9.11E-03	12
CAMTA1	0.52	1.85E-02	6
CENPK	0.66	3.12E-02	6
CEP170	0.64	2.54E-03	5
DNAJC9	0.65	4.32E-03	2
ENSA	1.18	2.22E-02	2
FAM126B	1.03	3.64E-02	5
H1FO	0.62	1.90E-03	2
HNRNPH2	0.61	4.18E-02	2
NAPEPLD	0.52	3.53E-02	3
NSMCE2	0.53	3.10E-02	3
RIT1	0.59	9.02E-03	3
RPL27A	1.06	1.25E-02	5
SSFA2	0.82	4.66E-02	3
UBE3B	0.74	2.74E-02	2
UTRN	0.7	2.41E-02	3

RNA targets of wild-type and mutant FET family proteins

Jessica I Hoell^{1,2,4}, Erik Larsson^{3,4}, Simon Runge¹, Jeffrey D Nusbaum¹, Sujitha Duggimpudi², Thalia A Farazi¹, Markus Hafner¹, Arndt Borkhardt², Chris Sander³ & Thomas Tuschl¹

FUS, EWSR1 and TAF15, constituting the FET protein family, are abundant, highly conserved RNA-binding proteins with important roles in oncogenesis and neuronal disease, yet their RNA targets and recognition elements are unknown. Using PAR-CLIP, we defined global RNA targets for all human FET proteins and two ALS-causing human FUS mutants. FET members showed similar binding profiles, whereas FUS mutants showed a drastically altered binding pattern, consistent with changes in subcellular localization.

Post-transcriptional regulatory networks controlled by microRNAs (miRNAs) and RNA-binding proteins (RBPs) have important roles in mRNA maturation and gene regulation^{1–5}. Dysregulation of these networks by mutation, deletion or overexpression of ribonucleoprotein complex components may result in disease⁶. We used photoactivatable ribonucleoside-enhanced cross-linking and immunoprecipitation (PAR-CLIP)⁷ to determine global protein-RNA interactions for all three members of the FET family of RBPs as well as two mutant forms associated with familial amyotrophic lateral sclerosis (ALS), and we provide large-scale and high-resolution target data for the entire family.

FUS, together with *EWSR1* and *TAF15*, form a gene family (*FET*) that encodes a set of abundant, ubiquitously expressed RBPs⁸. *FET* genes are directly involved in deleterious genomic rearrangements, primarily in sarcomas and leukemia⁹. Given their predominantly nuclear localization, FET family proteins have been implicated in various nuclear processes. All three proteins associate with the transcription factor II D complex, as well as directly with RNA polymerase II⁹. Moreover, *FUS* has a role in splicing⁹. Recently, mutations in *FUS* have been described as causing familial ALS^{10,11}, an adult-onset, rapidly progressing neurodegenerative disorder. The first reported mutations include the C-terminally located *FUS-R521G* and *FUS-R521H*, which both cause mislocalization of the physiologically mostly nuclear *FUS* protein to the cytoplasm^{10,11}. Despite numerous biochemical studies addressing the function of FET proteins in various nuclear processes, the RNA recognition elements (RRE) and the molecular targets have remained unknown. The impact of *FUS* mutations on RNA-binding capability and effective binding spectrum has also been unclear.

We generated six stable Flp-In T-REx HEK293 cell lines with either stable or inducible expression of N-terminally Flag-hemagglutinin (Flag-HA)-tagged human *FUS*, *EWSR1* or *TAF15*. Additionally, we generated two cell lines stably expressing Flag-HA-tagged disease-causing mutant forms of *FUS* (*FUS-R521G* or *FUS-R521H*). As previously reported, wild-type FET proteins localized primarily to the nucleus and mutant *FUS* localized to the cytoplasm (**Supplementary Fig. 1a**). Cell lines were grown for 12 to 16 h in 4-thiouridine (4SU)-supplemented medium to allow for 4SU incorporation into nascent RNA transcripts, as required by the PAR-CLIP protocol⁷. Cross-linked RNAs were recovered from SDS-PAGE-purified FET protein immunoprecipitates (**Fig. 1a**; see **Supplementary Fig. 1** for additional PAR-CLIP controls), converted into cDNA libraries and then Solexa-sequenced. The raw data was deposited in the DDBJ Sequence Read Archive (DRA), accession SRA025082 (http://trace.ddbj.nig.ac.jp/dra/index_e.shtml). Sequence reads were preprocessed, aligned against the human genome—allowing up to one mismatch—and annotated essentially as previously described⁷ (**Supplementary Table 1**).

For the initial quality control, we computed quantitative binding profiles for each dataset based on the total number of uniquely mapped sequence reads per RefSeq gene, before introducing thresholds to identify individual binding sites. Clustering revealed a high degree of similarity between replicates ($R_s = 0.84$ to 0.87 , Spearman's rank correlation) as well as gross similarities in the binding spectra of the FET proteins relative to three unrelated reference RBPs⁷ (**Fig. 1b**). The binding patterns of the two mutant *FUS* proteins closely resembled each other ($R_s = 0.91$), but they were also similar to—yet distinguishable from—those of the wild-type FET proteins ($R_s = 0.63$ to 0.71).

To define individual top target sites, we combined replicate reads for each of the three FET proteins and of the mutant forms. Using PAR-CLIP, we identified sites of cross-linking between protein and RNA by scoring for T-to-C mutations in clusters of cDNA sequence reads⁷. Top target sites were defined as clusters of ten or more overlapping reads, of which at least 25% contained T-to-C mutations (referred to below as ‘cross-linked clusters’). We obtained 39,984, 19,020, 8,678 and 14,953 such cross-linked clusters for *FUS*, *EWSR1*, *TAF15* and mutant *FUS*, respectively. Eighty-two percent of these were within RefSeq transcripts (cross-linked clusters are provided in **Supplementary Data 1** in BED format for display in, for example, the University of California, Santa Cruz (UCSC) Genome Browser¹²). Power analysis showed that we did not reach saturation of sites at our depth of Solexa sequencing (**Supplementary Fig. 2**). Intersection of the 1,000 most highly ranked cross-linked clusters in each dataset revealed large site-level overlaps between FET proteins (215–332

¹Howard Hughes Medical Institute, Laboratory of RNA Molecular Biology, The Rockefeller University, New York, New York, USA. ²Department of Pediatric Oncology, Hematology and Clinical Immunology, Center for Child and Adolescent Health, Heinrich Heine University, Duesseldorf, Germany. ³Computational Biology Center, Memorial Sloan-Kettering Cancer Center, New York, New York, USA. ⁴These authors contributed equally to this work. Correspondence should be addressed to T.T. (ttuschl@rockefeller.edu) or C.S. (sander@cbio.mskcc.org).

Received 14 April; accepted 15 September; published online 13 November 2011; doi:10.1038/nsmb.2163

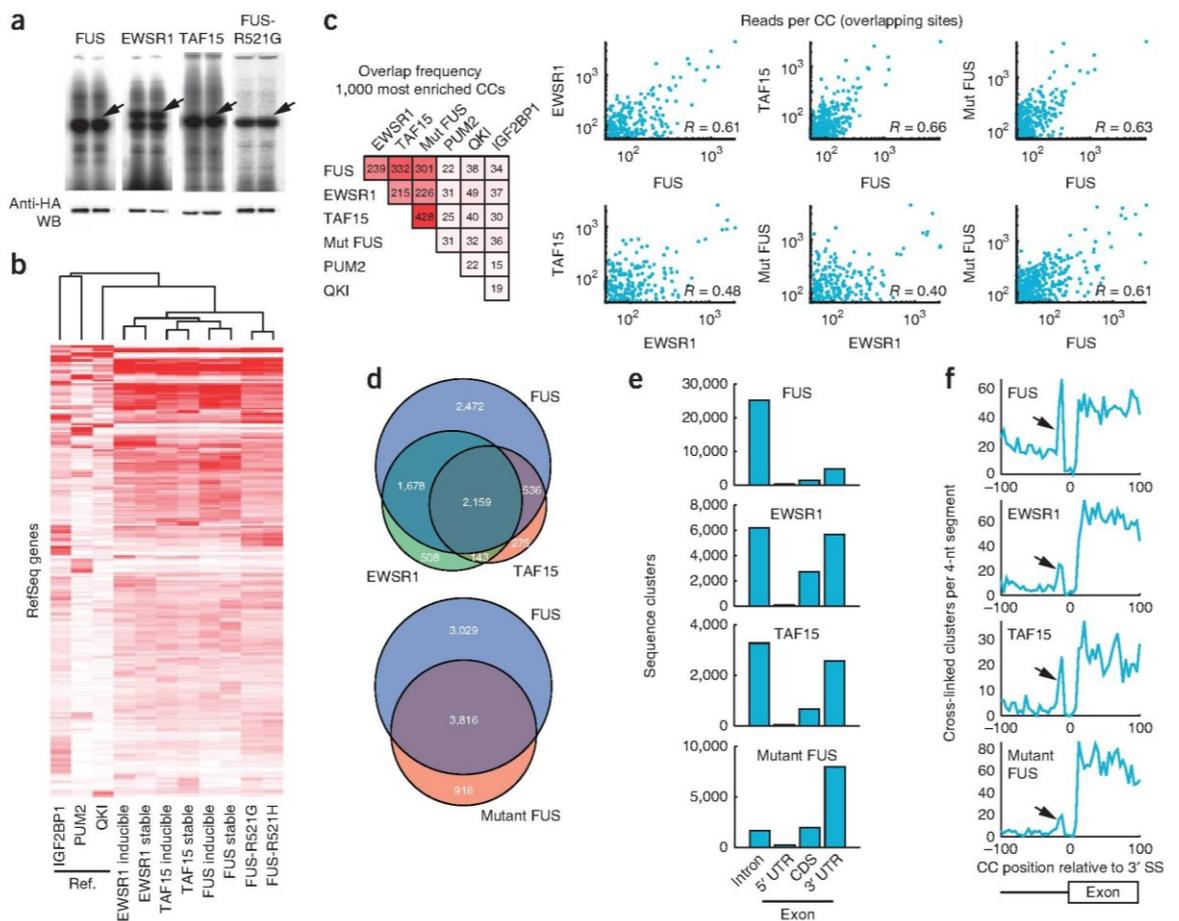


Figure 1 Protein-RNA interaction maps of FET proteins. (a) Phosphorimages of SDS-PAGE gels that resolved ^{32}P -labeled RNA-Flag-HA-FUS or EWSR1 or TAF15 PAR-CLIP immunoprecipitates. Arrows, excised regions. Protein identities of these bands were confirmed by MS (not shown). Western blots (WB) were probed with an anti-HA antibody. (b) Hierarchical clustering diagram of binding patterns based on the number of reads per gene and the Spearman correlation. Three unrelated reference (Ref.) datasets were included for comparison⁷. Binding profiles were mean-intensity normalized. Similar results were obtained for size-normalized datasets (data not shown). ‘Stable’, constitutive expression of the indicated protein; ‘inducible’, inducible expression of the indicated protein. (c) Overlap frequencies based on the top 1,000 cross-linked clusters of each protein, based on the number of sequence reads. Cross-linked clusters (CCs) were considered overlapping when center positions were within 10 nt. Scatter plots show the reproducibility in number of reads per overlapping site. Correlations (Pearson’s R) were calculated based on log-transformed values. Mut, mutant. (d) Overlaps between genes targeted by FET proteins, as well as between FUS and mutant FUS. (e) Distribution of cross-linked clusters across introns and exonic regions of RefSeq mRNAs. (f) Positional distribution of cross-linked clusters near intron-exon junctions show enriched binding upstream of the 3’ end splice site (arrows, 3’ SS). The P value for observing a peak of similar magnitude or higher anywhere in a 10,000-nt region upstream of the splice site was in all cases <0.025 (based on randomization of cross-linked cluster positions within introns).

cross-linked clusters) as well as between FET proteins and mutant FUS (226–428 cross-linked clusters), whereas overlaps with unrelated reference RBPs were small (22–49 cross-linked clusters, Fig. 1c). The number of reads per overlapping site was positively correlated between different FET proteins (Fig. 1c).

To complement the site-level analysis, cross-linked clusters were summarized on a per-gene basis (Supplementary Data 2). FUS, EWSR1 and TAF15 each targeted 6,845, 4,488 and 3,113 different genes, respectively, and these gene sets were largely overlapping (Fig. 1d). Mutant FUS, which targeted 4,732 genes, had an elevated fraction of unique targets compared to EWSR1 and TAF15, pointing toward an altered, rather than disrupted, binding profile (Fig. 1d). We also resampled library sequence reads to compare datasets with similar target gene numbers and obtained similar results (Supplementary Table 2). Overall, transcripts bound by wild-type FET proteins were

often bound at multiple positions, with one cross-linked cluster every 13,379 nucleotides, on average. A large fraction of cross-linked clusters for wild-type FET proteins fell within intronic regions, consistent with the nuclear localization of these proteins (FUS: 78% of mRNA clusters, EWS: 39% and TAF15: 47%; see Fig. 1e). By contrast, mutant FUS proteins had few intronic sites (13%) and bound predominantly to 3’ UTRs (61%). The distribution of cross-linked clusters across mRNA regions (5’ UTR, 3’ UTR and coding sequence) was markedly different from distributions of reference RBPs⁷ and also deviated from the relative sizes of these regions in RefSeq (Supplementary Fig. 3). Taken together, gene-level and site-level analyses showed similarities in the binding patterns of FET proteins and indicated that the RNA-binding properties of mutant FUS were not impaired or altered, but a different spectrum of target RNAs was accessed as a result of altered subcellular localization.

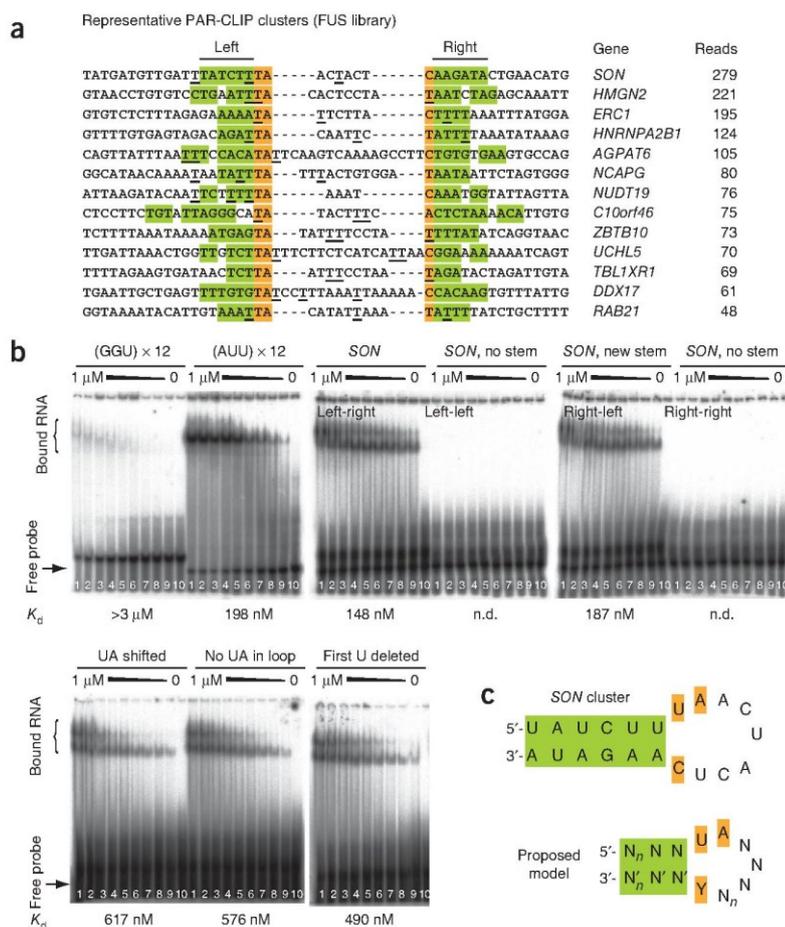
BRIEF COMMUNICATIONS

Figure 2 RNA-binding preferences of wild-type FET and mutant FUS proteins. (a) Genomic sequences of representative FET cross-linked clusters. All clusters are present in all wild-type FET and mutant FUS datasets. Green shading indicates stems ('left' and 'right'); orange shading indicates the nucleotides opening and closing the loop. The most frequent cross-linking positions are underscored. Gene name and number of reads are indicated.

(b) Phosphorimage of native PAGE gels that resolved complexes of recombinant FUS protein with different RNA oligoribonucleotides (all at 1 nM): GGU × 12, AUU × 12, SON (stem in natural left-right configuration as indicated in panel A), altered SON (non-complementary left-left stem), altered SON (non-complementary right-right stem) and altered SON (reconstituted right-left stem). Additionally, the effects of changing the 'UA' that opens the loop are shown (UA shifted, no UA in loop, or first U deleted). Concentrations of FUS protein ranged from 1,000 nM to 0 nM (lanes 1 to 10; fractions of bound versus unbound protein can be found in **Supplementary Data 3**). Dissociation constants (K_d) are indicated; n.d., non-determinable. (c) Proposed model of the FET protein RRE. The variable n is an integer ≥ 1 indicating stem and loop lengths.

By investigating the positional distribution of FET cross-linked clusters in relation to splice sites, we observed an increased frequency of intronic binding immediately upstream of 3' end splice sites (**Fig. 1f**) but not downstream of 5' end splice sites (not shown). This pattern was not observed in the three reference datasets (**Supplementary Fig. 4**). The G-rich intron-exon junction was essentially void of cross-linked clusters, which may be due to specific cleavage downstream of G residues by RNase T1 during the PAR-CLIP protocol.

Use of standard bioinformatic tools (see **Supplementary Methods**) did not return a significant RRE motif for any of the FET proteins, indicating that the RNA structure may have a role in recognition. Many of the FUS cross-linked clusters in our study contained a conventional stem-loop structure (**Fig. 2a**) that frequently opened with a U•U or U•C non-Watson-Crick base pair, in which the U at the 5' end of the loop is followed by an A in the loop (67% in a 60-nucleotide (nt) window around cluster centers). This pattern was less frequent in randomly selected intronic and 3' UTR regions (14% and 39%) but similar in shuffled sequences. Furthermore, the FUS cross-linked clusters had low G and high AU content, which could not only reflect the binding preferences of the protein but also, at least in part, a methodological preference^{7,13}. We therefore experimentally tested the ability of FUS protein to bind to AU-rich stem-loop structures, and we evaluated 37-nt oligoribonucleotides corresponding to a predicted stem-loop sequence within a cross-linked cluster of the SON transcript (**Supplementary Fig. 5**) by electrophoretic mobility shift assays. We additionally evaluated a GGU repeat, as FUS had previously been shown to bind to GGUG-containing RNAs in a G-rich context¹⁴. The dissociation constant of the SON stem-loop (148 nM) was at least 15-fold higher than that of the GGU repeat RNA (**Fig. 2b**). Disruption of the SON stem-loop abolished FUS binding, whereas the compensatory sequence change restoring the disrupted



stem-loop also restored binding. Altering the UA residues opening the loop also decreased binding (**Fig. 2b**). An AUU trinucleotide repeat sequence, which approximates the nucleotide distribution of our cross-linked clusters and is predicted to accommodate a stem-loop structure comprising non-Watson-Crick U•U base pairs, had a binding constant similar to the SON stem-loop (198 nM). Together, our results suggest that FUS protein binds AU-rich stem-loops, and whereas the stem especially contributes to binding (**Fig. 2c**), specific loop residues also contribute. Structural studies using natural and non-natural high-affinity target RNAs can now be undertaken to test our interaction model.

Our PAR-CLIP results show that FET proteins bind RNA, including most cell-expressed mRNAs, at high frequency. We detected preferential binding near splice acceptors, in support of their proposed role in pre-mRNA splicing⁹, but these events represented a minor fraction of all cross-linked clusters. In addition, global changes in transcript abundance in response to FUS silencing in HEK293 cells were weak and did not correlate with FUS binding (**Supplementary Fig. 6**). Mutant FUS proteins showed drastically altered distribution of cross-linked clusters across transcript regions, consistent with translocation to the cytoplasm, while still maintaining their RNA-binding capability and specificity. This result supports a model in which the deleterious effect of dominant ALS-causing FUS mutations such as R521G and R521H may be partly caused by a gain-of-function effect due to either increased interaction with cytoplasmic RNA targets or independent cellular stress from mislocalization of an abundant

nuclear protein. Notably, we found endoplasmic reticulum and ubiquitin-proteasome-related target gene categories to be overrepresented among transcripts uniquely targeted by mutant FUS proteins (Fig. 1d and Supplementary Table 3), which further supports the idea that protein synthesis and degradation represent major pathways perturbed in ALS¹⁵. Comprehensive mapping of the molecular targets of FET proteins and disease-causing mutant FUS forms (see Supplementary Data 1 and 2) might facilitate future studies related to ALS and the evaluation of models for this disease.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

J.I.H. is supported by the Deutsche Forschungsgemeinschaft, E.L. is supported by the Swedish Research Council and M.H. is supported by the Charles Revson Jr. Foundation. T.T. is a Howard Hughes Medical Institute investigator, and work in his laboratory was supported by US National Institutes of Health grant MH08442 and by the Starr Foundation.

AUTHOR CONTRIBUTIONS

J.I.H. carried out stable cell line generation (FUS, EWSR1 and FUS mutants), PAR-CLIP experiments (FUS, EWSR1 and FUS mutants), short interfering RNA (siRNA) knockdowns and gel shift experiments; E.L. did all the computational analyses; S.R. carried out the generation of TAF15 stable cell lines and the TAF15

PAR-CLIP; J.D.N. did the recombinant protein purification and conducted gel shift experiments; S.D. conducted the western blot experiments; T.A.F. and M.H. did the siRNA knockdowns; A.B., C.S. and T.T. supervised the project; and E.L., J.I.H. and T.T. wrote the paper.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at <http://www.nature.com/nsmb/>.

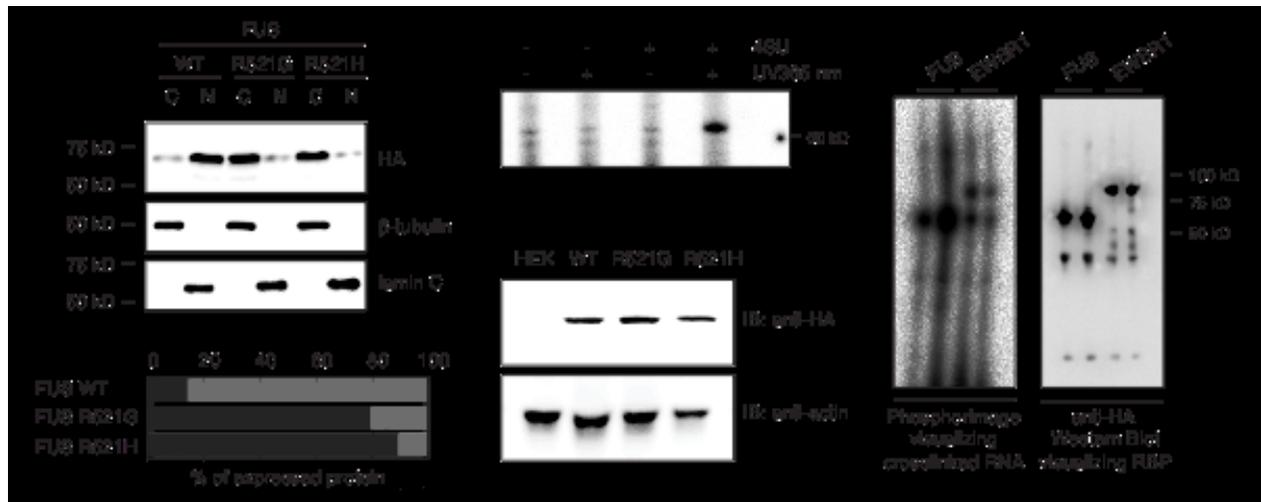
Published online at <http://www.nature.com/nsmb/>.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

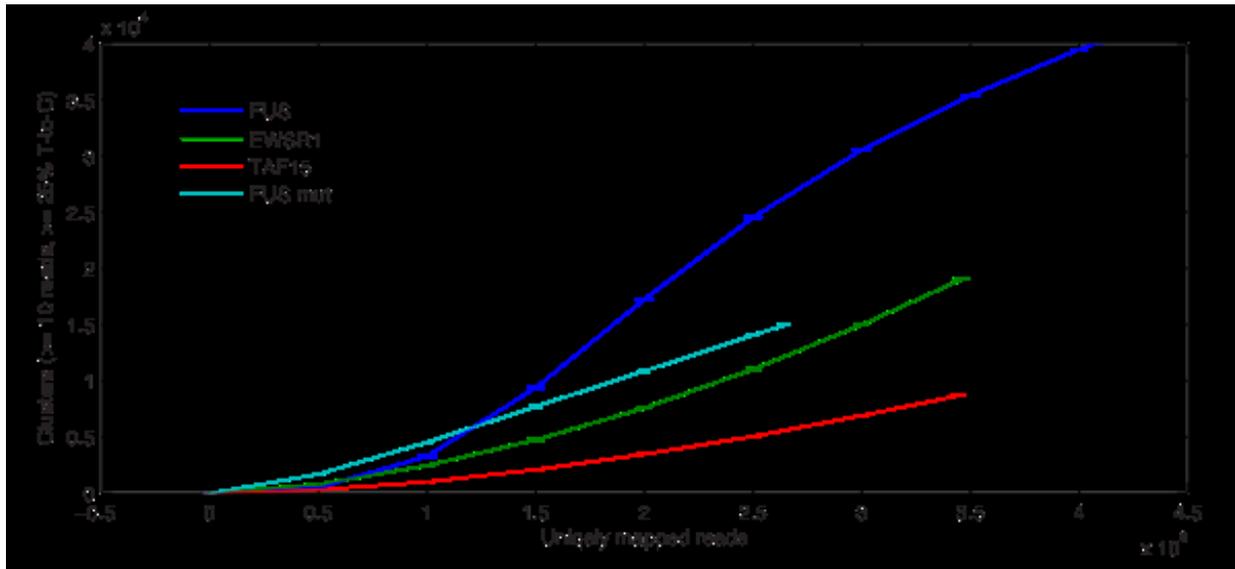
- Keene, J.D. *Nat. Rev. Genet.* **8**, 533–543 (2007).
- Licalosi, D.D. & Darnell, R.B. *Nat. Rev. Genet.* **11**, 75–87 (2010).
- Martin, K.C. & Ephrussi, A. *Cell* **136**, 719–730 (2009).
- Moore, M.J. & Proudfoot, N.J. *Cell* **136**, 688–700 (2009).
- Sonenberg, N. & Hinnebusch, A.G. *Cell* **136**, 731–745 (2009).
- Cooper, T.A., Wan, L. & Dreyfuss, G. *Cell* **136**, 777–793 (2009).
- Hafner, M. *et al. Cell* **141**, 129–141 (2010).
- Wu, C. *et al. Genome Biol.* **10**, R130 (2009).
- Tan, A.Y. & Manley, J.L. *J. Mol. Cell Biol.* **1**, 82–92 (2009).
- Kwiatkowski, T.J. Jr. *et al. Science* **323**, 1205–1208 (2009).
- Vance, C. *et al. Science* **323**, 1208–1211 (2009).
- Kent, W.J. *et al. Genome Res.* **12**, 996–1006 (2002).
- Kishore, S. *et al. Nat. Methods* **8**, 559–564 (2011).
- Lerga, A. *et al. J. Biol. Chem.* **276**, 6807–6816 (2001).
- Okamoto, K., Fujita, Y. & Mizuno, Y. *Neuropathology* **30**, 189–193 (2010).

8. Supplementary information for Publication II

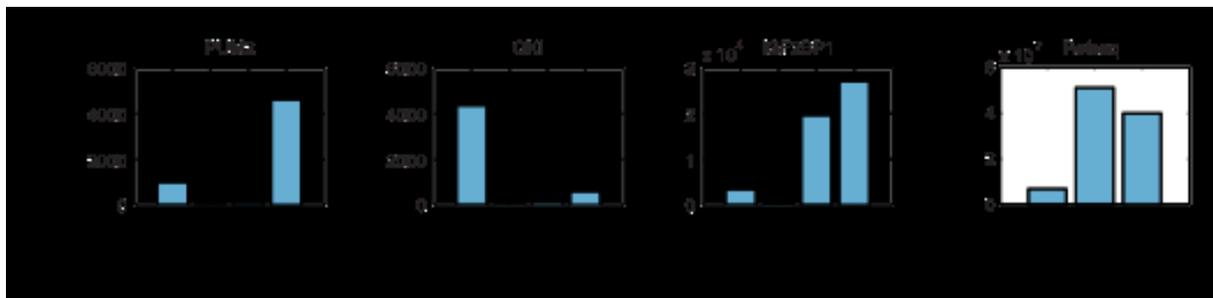
8.1 Supplementary Figures



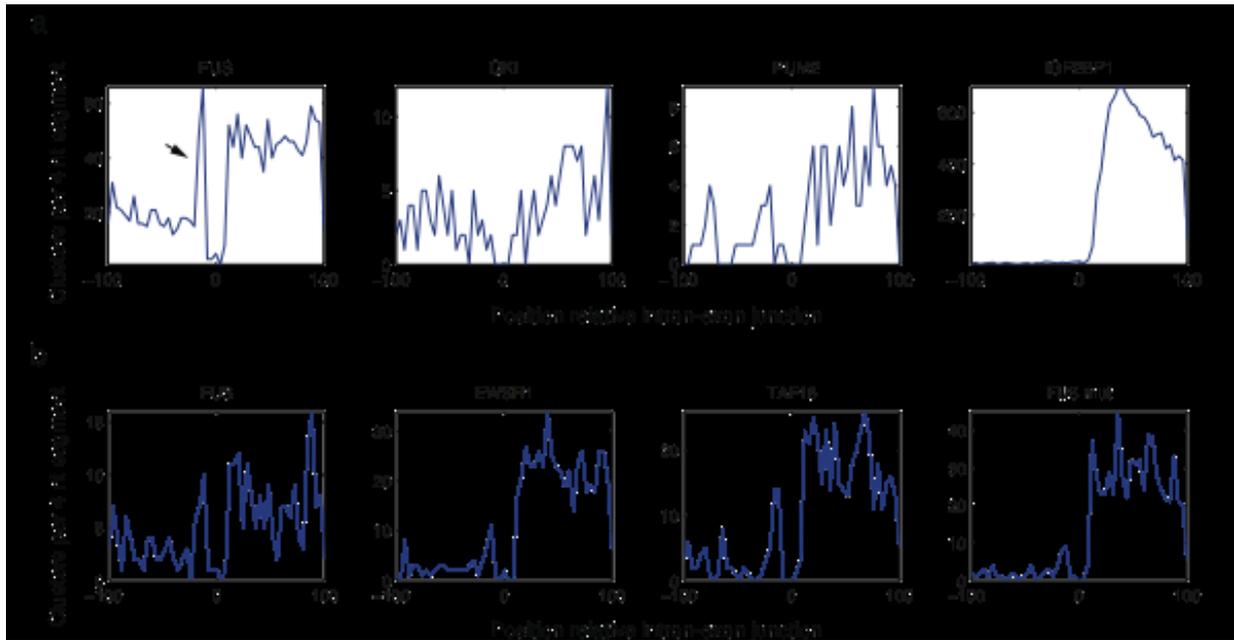
Supplementary Figure 1 Additional PAR-CLIP quality controls. **(a)** Subcellular localization of wild-type and mutant FUS. C, cytoplasmic; N, nuclear; protein targets of antibodies are indicated. Quantification of subcellular distribution of FUS mutants is shown in the bar graph; dark grey, cytoplasmic localized protein; light grey, nuclear localized protein. Band intensities were measured using ImageJ and the sum of each set (C+N) was set to 100%. Tagged wild-type FUS is mostly nuclear whereas tagged mutant FUS is mostly cytoplasmic, similar to reports for untagged protein^{1,2}. **(b)** Crosslinking of RNA to FUS in PAR-CLIP experiments is dependent on the incorporation of 4SU and UV 365 nm irradiation. Four similar-sized cell pellets obtained from FLAGHA-tagged FUS expressing cell lines grown in presence or absence of 4SU and treated with or without UV 365 nm light were carried through the PAR-CLIP protocol up to the SDS-PAGE gel. 4SU, 4-thiouridine; UV 365 nm, samples crosslinked at 365 nm wavelength. **(c)** FLAGHA-tagged wild-type and mutant FUS expressing HEK293 T-REx Fip-In cells show similar protein abundance. Cell lysates of obtained from the same numbers of HEK293 T-REx Fip-In cells expressing the indicated FLAGHA-tagged protein were separated by SDS-PAGE and Western blot analysis was performed using anti-HA and anti-beta actin antibody. **(d)** Denaturing SDS-PAGE demonstrates that crosslinked RNA and FLAGHA-tagged RBP co-migrate. Aliquots of FUS and EWSR1 PAR-CLIP samples (see Figure 1a) were separated by SDS-PAGE. The blot was first analyzed by phosphorimaging visualizing radiolabeled RNA and subsequently by immunoblotting with anti-HA antibody. The lower of the double bands for EWSR1 was found by mass spectrometry analysis (data not shown) to additionally contain FUS and TAF15, suggesting that FET proteins may heterodimerize or bind in close proximity.



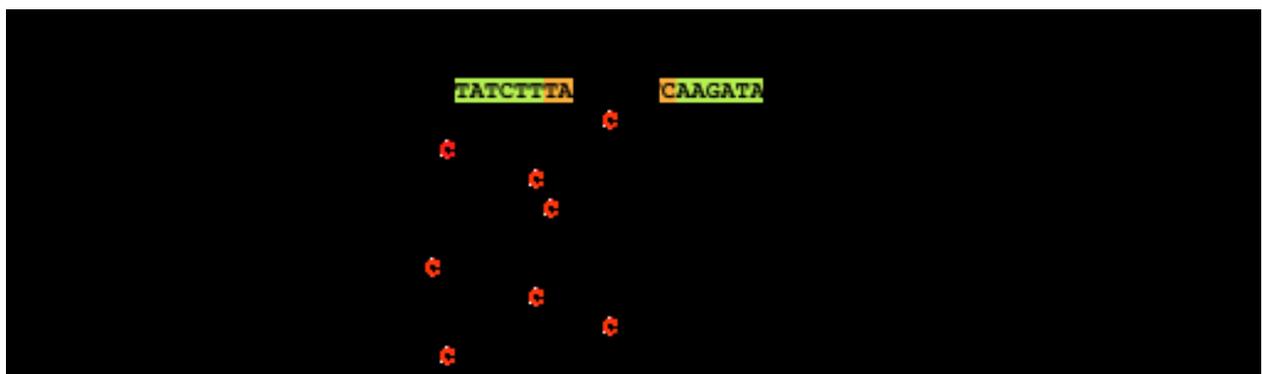
Supplementary Figure 2 Power analysis shows subsaturation in discovery of CCs. We performed simulations where sequence reads were randomly removed to emulate smaller sequence datasets. The number of CCs is plotted as a function of the number of uniquely mapped reads, in steps of 500,000. The plot shows the mean number of CCs obtained in three independent simulations. Error bars indicate the maximum and minimum obtained value.



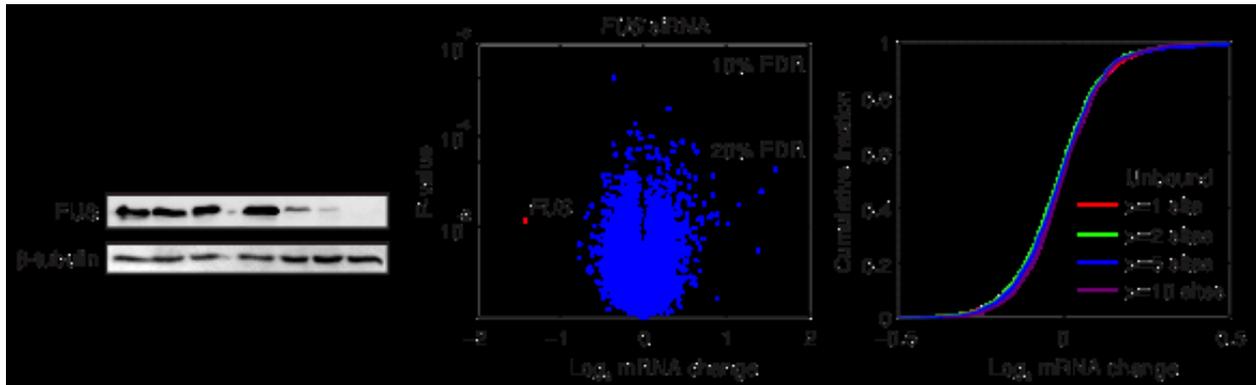
Supplementary Figure 3 Binding frequencies across transcript regions for reference RBPs. **(a)** Distribution of CCs across transcript regions for PUM2, QKI and IGF2BP1. These are included for comparison with Figure 1e. **(b)** The total number of nucleotides in 5' UTRs, CDS and 3' UTRs in RefSeq. The numbers were calculated by summarizing over all coding RefSeq transcripts.



Supplementary Figure 4 Extended analysis of positional distribution of CCs at intron-exon junctions for various RBPs. **(a)** Enriched binding near splice acceptors was not observed for QKI, despite frequent binding to introns similar to FUS. CCs of $\geq 25\%$ T-to-C changes were used. Due to the smaller size of the PUM2 and QKI datasets, a minimum of 5 reads per CC instead of 10 were required. The Y-axis indicates the number of observed CCs per 4 nt segment. **(b)** Positional distribution of CCs for FET-proteins based on size-normalized datasets. Each dataset was reduced to contain the same number of sites as QKI ($n = 6822$), by random sampling of CCs, showing that the observed effect is independent of dataset size.



Supplementary Figure 5 Representation of a PAR-CLIP CC from reads corresponding to the SON gene transcript. Reads were aligned to the genome and T-to-C changes are indicated in red. Green indicates the two complementary segments of a predicted stem-loop and orange indicates the additional nucleotides proposed to constitute the FET protein family RRE.



Supplementary Figure 6 siRNA-mediated silencing of *FUS*. HEK293 T-REx Flp-In cells were transfected with three different siRNAs targeting *FUS* and two different control siRNAs (Applied Biosystems) using Lipofectamine RNAiMAX (all from Invitrogen) at 50 nM final concentration, each in duplicates (10 arrays in total). (a) Knockdown efficiency was assayed after 72 h by anti-*FUS* immunoblots. (b) Volcano plot of whole-transcriptome mRNA changes and P-values after *FUS* silencing. 200 ng of total RNA was assayed using U133 Plus 2.0 arrays (Affymetrix). Raw files were processed in R using the Bioconductor package and the RMA algorithm. Duplicates were averaged, redundant probe sets merged and genes were evaluated for differential expression using Student's t-test followed by correction for multiple testing (Storey and Tibshirani's method). No genes were significantly changed at 5% FDR, 1 gene at 10% FDR and 16 genes at 20% FDR. (c) The cumulative distribution of mRNA changes (at a zoomed-in scale compared to panel b) for transcripts with and without *FUS* binding sites (as determined by PAR-CLIP) reveals that *FUS*-bound transcripts were not notably different in their response to *FUS* silencing compared to unbound transcripts.

8.2 Supplementary Tables

Supplementary Table 1 Annotation table of mapped PAR-CLIP sequence reads.

Library	Number of mapped reads (error distance 1)	Number of uniquely mapped reads	mRNA reads (%)	miRNA reads (%)	miscRNA reads (%)	piRNA reads (%)	rRNA reads (%)	snRNA reads (%)	tRNA reads (%)	other (%)
FUS stable ¹	4,894,941	2,007,928	38.77	0.21	2.6	0.04	28.51	0.78	0.25	28.84
FUS inducible ²	5,688,260	2,064,289	40.49	0.02	2.46	0.05	6.47	0.24	0.11	50.18
EWSR1 stable	4,348,695	2,153,122	31.88	0.05	2.99	0.03	54.29	0.72	0.46	9.59
EWSR1 inducible	2,885,147	1,308,001	29.51	0.13	4.62	0.03	47.21	0.74	1.24	16.53
TAF15 stable	3,780,613	1,324,082	33.25	0.21	2.94	0.04	26.43	1.83	1.49	33.8
TAF15 inducible	4,008,522	1,332,335	46.84	0.09	2.37	0.03	10.33	1	0.78	38.55
FUS-R521G	4,271,115	1,792,475	39.68	0.24	3.78	0.04	37.65	0.89	0.91	16.8
FUS-R521H	1,826,791	850,452	34.84	0.26	4.51	0.03	47.68	1.43	1.42	9.83

¹refers to cell lines constitutively expressing the indicated protein

²refers to cell lines which express the indicated protein after induction with doxycycline

Supplementary Table 2 Normalization of target gene counts. Datasets were reduced by random sampling of reads. Read numbers were individually tuned for each dataset to equalize the number of targeted genes. Two size normalizations were performed; one where the number of target genes was dictated by EWSR1 (“Resampled to EWSR1 gene count”), and a more radical size reduction dictated by TAF15 (“Resampled to TAF15 gene count”). Size reduction decreased overall overlaps, meaning that the reduced datasets had higher fractions of uniquely targeted genes.

Dataset Protein	Actual reads				Resampled to EWSR1 gene count				Resampled to TAF15 gene count			
	FUS	EWSR1	TAF15	Mut. FUS	FUS	EWSR1	TAF15	Mut. FUS	FUS	EWSR1	TAF15	Mut. FUS
Uniquely mapped reads	4069214	3458882	3443557	2642445	1900000	3458882	3443557	2500000	1400000	2500000	3443556	1500000
Targeted genes	6845	4488	3113	4732	4563	4488	3113	4536	3083	3159	3113	3084
% unique targets (not targeted by FUS)		14.5%	13.4%	19.4%		32.1%	28.6%	37.7%		43.3%	44.7%	47.2%

Supplementary Table 3 Gene ontology (GO) analysis of genes uniquely bound by mutant FUS (not bound by the wild-type protein), as depicted in Figure 1d.

GO Term	Unique mut. FUS targets ¹	Genome ²	Fold enrichment	P-value	Description
GO:0005515	43.4%	34.6%	1.3	7.48E-08	protein binding
GO:0005739	10.0%	5.6%	1.8	3.65E-07	mitochondrion
GO:0005829	10.1%	6.0%	1.7	2.96E-06	cytosol
GO:0005634	34.5%	27.4%	1.3	3.27E-06	nucleus
GO:0005737	29.4%	22.9%	1.3	8.39E-06	cytoplasm
GO:0005783	8.1%	4.7%	1.7	1.23E-05	endoplasmic reticulum
GO:0000502	1.0%	0.1%	6.8	1.36E-05	proteasome complex
GO:0031145	1.5%	0.4%	4.1	2.67E-05	anaphase-promoting complex-dependent proteasomal ubiquitin-dependent protein catabolic process
GO:0051436	1.5%	0.4%	4.1	2.67E-05	negative regulation of ubiquitin-protein ligase activity during mitotic cell cycle
GO:0005743	3.3%	1.4%	2.4	3.43E-05	mitochondrial inner membrane
GO:0051437	1.5%	0.4%	4	4.15E-05	positive regulation of ubiquitin-protein ligase activity during mitotic cell cycle
GO:0009313	0.4%	0.0%	23.8	7.36E-05	oligosaccharide catabolic process

¹fraction of genes uniquely bound by mutant FUS (916) annotated with a specific GO term

²background fraction (all ENSEMBL genes) for a specific GO term

8.3 Supplementary Methods

Oligonucleotides and siRNA duplexes

The following oligodeoxynucleotides were used for plasmid preparation, mutagenesis reactions and sequencing:

FLAGHA_BamHI_for,

5'GATCGACCGGTGACTACAAGGACGACGATGACAAGTACCCTTATGACGTGCCCGATTA
CGCTG;

FLAGHA_BamHI_rev,

5'GATCCAGCGTAATCGGGCACGTCATAAGGGTACTTGTTCATCGTCGTCCTTGTAGTCACC
GGTG;

FUS_pET23a_SalI_for, 5'ACGCGTCGACCCATGGCCTCAAACGATTATACC;

FUS_pET23a_NotI_rev, 5'ATAGTTTAGCGGCCGCATACGGCCTCTCCCTGCGATC;

FUS_NcoI_for, 5'CAATCCCATGGACTACAAGGACGACGATGAC;

FUS_EcoRV_rev, 5'GCAATCGATATCTCAGTGGTGGTGGTGGTGGTG;

FUS_PCR_for, 5'ACGCGTCGACATGGCCTCAAACGATTATACCCAAC;

FUS_PCR_rev, 5'ATAGTTTAGCGGCCGCTTAATACGGCCTCTCCCTGC;

EWSR1_PCR_for, 5'ACGCGTCGACATGGCGTCCACGGATTACAGTAC;

EWSR1_PCR_rev, 5'ATAGTTTAGCGGCCGCCTAGTAGGGCCGATCTCTGC;

TAF15_PCR_for, 5'ACGCGTCGACATGTTCGGATTCTGGAAG;

TAF15_PCR_rev, 5'ATAGTTTAGCGGCCGCTCAGTATGGTCGGTTGC;

FUS_C521G_for, 5'GAGCACAGACAGGATGGCAGGGAGAGG;

FUS_C521G_rev, 5'CCTCTCCCTGCCATCCTGTCTGTGCTC;

FUS_G521A_for, 5'CACAGACAGGATCACAGGGAGAGGCCG;

FUS_G521A_rev, 5'CGGCCTCTCCCTGTGATCCTGTCTGTG;

FUS_int_seq, 5'GGCAGTGGTGGCGGTTATGGC;

EWS_int_seq, 5'CGGTGGAATGGGCAGCGCTGGAGAGCGAG;

TAF15_int_seq, 5'GGCGTGGGGGATATGACAAGG;

(CDS) without the stop codon of FUS. The PCR product was SalI and NotI digested and ligated into the SalI and NotI digested pET23(a)_mod vector. Another PCR was performed (primers FUS_NcoI_for and FUS_EcoRV_rev) and the FLAGHA_FUS_His6 cDNA was amplified. A regular pENTR4 vector (Invitrogen) was NcoI and EcoRV digested. A fill-in reaction with T4 DNA polymerase was performed to create blunt ended products of both the FLAGHA_FUS_His6 cDNA and the pENTR4 vector. Then, the FLAGHA_FUS_His6 cDNA was ligated into the pENTR4 vector. The pENTR4_FLAGHA_FUS_His6 construct was recombined into pDEST8 destination vector using GATEWAY LR recombinase according to manufacturer's protocol (Invitrogen).

Mammalian cell culture and creation of stable cell lines

T-REx HEK293 Flp-In cells (Invitrogen) were grown in D-MEM high glucose (1x) with 10% fetal bovine serum, 100 U per ml of penicillin, 100 µg per ml streptomycin, 100 µg per ml zeocin and 15 µg per ml blasticidin. Cell lines constitutively or inducibly expressing FLAGHA-tagged proteins were generated by co-transfection of pFRT_TO_FLAGHA_GOI (gene of interest) or pFRT_FLAGHA_GOI constructs with pOG44 (Invitrogen) using lipofectamine 2000 (Invitrogen). Cells constitutively or inducibly expressing FUS, EWSR1, TAF15, FUS-R521G or FUS-R521H were cultivated in D-MEM high glucose (1x, Invitrogen) with 10% fetal bovine serum, 100 U per ml of penicillin, 100 µg per ml streptomycin and 100 µg per ml hygromycin (Invivogen). In the case of inducible expression 15 µg per ml blasticidin (Invivogen) was also added to this medium. Induction was achieved by adding 1 µg per ml doxycycline to the growth medium 15 to 20 h before crosslinking.

Insect cell culture, recombinant protein expression in *Spodoptera frugiperda* (Sf9) cells and protein purification

Sf9 cells were grown in Grace's Insect Medium (Invitrogen, #11605-094), supplemented with 10 % fetal bovine serum, 1 % Pluronic F-68 (Invitrogen, #24040-032), 100 U per ml of penicillin, 100 µg per ml streptomycin and maintained in room air at 26°C in spinner culture (80 rpm). The pDEST8_FLAGHA_FUS_His6 was transformed into MAX Efficiency DH10Bac competent *E. coli* (Invitrogen). Bacmid DNA was isolated using PureLink HQ Mini Plasmid Purification and transfected into Sf9 cells using Cellfectin II Reagent kit (all from Invitrogen). Three rounds of viral amplifications yielded 250 ml of cell supernatant containing 2×10^8 plaque forming units per ml virus. 25 ml of this solution were used for infection of one liter of Sf9 culture maintained at a density of 1×10^6 cells per ml. Four days after infection, Sf9 cells were washed by centrifugation (500 x g) in 1x PBS, and pellets were suspended in 5-times the pellet volume of ice-cold lysis buffer (50 mM Tris-HCl, pH 8.0, 5 mM MgCl₂, 1 M KCl, 10% glycerol, 5 mM imidazol, 0.1% triton-X-100, 1 mM beta-mercaptoethanol and Complete EDTA-free protease inhibitor cocktail (Roche)). The suspension was incubated on ice for 10 min and then additionally suspended by 20 strokes with a Dounce homogenizer. Insoluble material

was removed by centrifugation for 20 min at 20,000 x g and the supernatant was further cleared by passing through a 5µm Supor membrane syringe filter (Pall Acrodisc). Sf9-expressed FUS was purified using the AektaExplorer (Amersham Bioscience). 10 ml of Co²⁺ TALON beads were washed 3 times with de-ionized water and packed into a XK 16 column (Amersham Bioscience). The column was equilibrated with 4 column volumes (CV) lysis buffer supplemented with 5 mM imidazole and 1 mM DTT. The cell lysate was loaded onto the equilibrated Co²⁺ TALON column (GE Healthcare Life Sciences XK 16/20 column with an AK16 adapter) using a flow rate of 1 ml per min. The column was washed with 2 CV lysis buffer supplemented with 5 mM imidazole and 2 CV supplemented with 13 mM imidazole. The protein was eluted from the column in 4 CV elution buffer (50 mM Tris-HCl, pH 8.0, 5 mM MgCl₂, 1 M KCl, 10% glycerol, 400 mM imidazole, 0.1% triton-X-100, 1 mM beta-mercaptoethanol) running a gradient with a final concentration of 400 mM imidazole. During elution, 1 ml fractions were collected and analyzed on SDS-gels. Fractions containing FUS were pooled and dialyzed overnight in 2 times 1 l of dialysis buffer (20 mM Tris-HCl, pH 8.0, 300 mM KCl, 5 mM MgCl₂, 0.1% v per v Triton X-100, 50% v per v glycerol, 1 mM DTT) using dialysis bags (Spectrum, SpectraPor, 08-667E) with a molecular cutoff of 12 to 14 kDa. Protein concentrations were estimated by comparing Coomassie stain intensity against a BSA standard on a 10% SDS-gels.

Electrophoretic mobility shift assay (EMSA)

10 pmol oligonucleotide were labeled with 5 pmol [γ -³²P]-ATP in a 10 µl reaction containing 70 mM Tris-HCl (pH 7.6 @ 25), 10 mM MgCl₂, 5 mM DTT, and 5 U T4 PNK. The reaction was denatured (95°C, 30 sec), and placed on ice. After 1 min, 5 U T4 PNK was added and the reaction was incubated at 37°C for 15 min. After 15 min, regular ATP was added to a final concentration of 1 mM and the reaction was incubated for an additional 5 min at 37°C. The reaction volume was increased to 50 µL and denatured at 95°C for 30 sec. The unincorporated [γ -³²P]-ATP was removed by passing the reaction mixture through a G25 column (GE Life Science). The eluate volume was increased to 100 µl and the 100 nM oligonucleotide was stored at -20°C. In a 20 µl reaction, 1 nM labeled RNA was incubated with protein concentrations varying from 0-1 µM in buffer containing 20 mM Tris-HCl (pH 7.65 @ 25°C), 300 mM KCl, 5 mM MgCl₂, 1 mM DTT, 0.1U per µl RNasin (Promega), 100 ng yeast tRNA, and 0.1 mg per ml acetylated BSA at 30°C for 1 hour in 1.5 ml passivated (50 µL 1 mg per ml acetylated BSA, 27°C, 1 hour), siliconized eppendorf tubes. After 1 hour, 5 µl of buffer with 50% glycerol and bromophenol blue was added. Following a 30 min pre-run, the reaction was separated by native page (25 mM Tris, 0.2 M Glycine, 6% 49:1 acrylamide:bisacrylamide, ammoniumpersulfat, TEMED), at 4°C at 300V in 25 mM Tris 0.2 M Glycine containing buffer. The reaction was loaded onto running gels, and the species were separated for 2 hours at 300 V at 4°C. The [γ -³²P] radioactive signal was detected using phosphorimager screens and the signal was quantified using ImageGauge software. Curves and binding constants were calculated using Kaleidagraph software.

Preparation of whole cell extracts and Western blotting

For whole cell mammalian lysates, cells were washed with ice-cold PBS and lysed in 3 pellet volumes 10 mM HEPES-KOH, pH 7.5, 150 mM KCl, 2 mM EDTA, 1 mM NaF, 0.5% NP-40, 0.5 mM DTT and complete EDTA-free protease inhibitor cocktail (Roche). The lysate was incubated for 10 min on ice and cleared for 10 min at 20,000 x g and 4°C. Total protein concentration was measured by Bradford protein assay (Biorad). Whole cell lysates were analyzed on a 10% SDS-gel. Protein samples were then transferred to nitrocellulose membrane (BioRAD, Trans-Blot; 1.5 mAmp per cm² membrane for 1.5 hrs) and probed with the indicated antibodies. Signals were developed using the ECL kit (Amersham) under standard conditions. The luminescence signal was recorded with a Fujifilm Image Reader LAS-3000.

Preparation of nuclear extracts

Cells were harvested by trypsinization and centrifuged for 5 min at 2,000 x g. All following steps were performed at 4°C. The cell pellet was resuspended in 1x PBS and centrifuged for 5 min at 2,000 x g. Next, the packed cell volume (PCV) was recorded and the cells were resuspended in 5x PCV Hypotonic Lysis Buffer (10 mM HEPES pH 7.9 (KOH), 10 mM KCl, 1.5 mM MgCl₂, 0.5 mM DTT, Complete EDTA-free protease inhibitor cocktail (Roche)) and incubated on ice for 10 min. After a further centrifugation step (5 min at 2,000 x g) the pellet was resuspended in 2x PCV Hypotonic Lysis Buffer. The suspension was homogenized with 5 strokes in a Dounce glass homogenizer (type B pestle) and cell lysis was ensured microscopically. The lysate was centrifuged for 10 min at 2,000 x g, the supernatant was saved as the cytoplasmic extract (further centrifuged at 13,000 x g for 30 min) and the exact volume was recorded. To wash the nuclei, they were resuspended in 2x PCV Hypotonic Lysis Buffer and further centrifuged for 10 min at 2,000 x g. After discarding the supernatant, the nuclei were resuspended in 1x SDS sample buffer so that the final volume equaled that of the cytoplasmic extract, sonicated for 15 sec and boiled for 3 min. Purity of the fractions was tested by probing with anti-lamin and anti-tubulin antibodies.

Antibodies

Monoclonal anti-HA.11 (clone 16B12, Covance), polyclonal anti-FUS (Abcam, AB23439), beta-tubulin (Sigma, T4026), beta-actin (Sigma, SAB3500350) and lamin C (Abcam, AB16048) were used as primary antibodies at a 1:1000 dilution. Anti-Flag M2 (Sigma, F3165) was used for PAR-CLIP. HRP-conjugated anti-rabbit Ig and anti-mouse Ig (both from DAKO) were used as secondary antibodies for Western blot analysis.

PAR-CLIP

PAR-CLIP was performed as described before³. Briefly, the growth medium of HEK293 T-REx Flp-In cells expressing FLAGHA-tagged FET proteins was supplemented with 100 μ M 4SU for 12 hours prior to crosslinking. After decanting the growth medium, cells were irradiated uncovered with 0.15 J per cm² total energy of 365 nm UV light in a Stratalinker 2400. Cells were harvested at 500 x g and lysed in 3 cell pellet volumes of NP40 lysis buffer (50 mM HEPES, pH 7.5, 150 mM KCl, 2 mM EDTA, 1 mM NaF, 0.5% (v per v) NP40, 0.5 mM DTT, complete EDTA-free protease inhibitor cocktail (Roche)). After centrifugation at 13,000 x g the cleared cell lysate was treated with RNase T1. FLAGHA-tagged FET proteins were immunoprecipitated with an anti-FLAG antibody conjugated to Protein G Dynabeads (Invitrogen). After a second RNase T1 digestion, beads were washed in high-salt wash buffer (50 mM HEPES-KOH, pH 7.5, 500 mM KCl, 0.05% (v per v) NP40, 0.5 mM DTT, complete EDTA-free protease inhibitor cocktail (Roche)) and resuspended in dephosphorylation buffer (50 mM Tris-HCl, pH 7.9, 100 mM NaCl, 10 mM MgCl₂, 1 mM DTT). Following incubation with calf intestinal alkaline phosphatase beads were washed again. Next, the crosslinked RNA was radiolabelled using T4 polynucleotide kinase. After a final wash step, RNA–FET complexes were released from the beads by incubating at 90°C and subsequently separated on SDS-gels. The excision of the bands corresponding to the expected masses of the proteins was followed by the electroelution of the RNA–FET complexes. Following proteinase K digestion, RNA was recovered from the eluate by acidic phenol-chloroform extraction and ethanol precipitation. The 5'-³²P-phosphorylated RNA was then carried through a standard cDNA library preparation protocol⁴. Both the 3' and the 5' Solexa adapters were ligated to the RNA followed by its reverse transcription. The resulting cDNA was amplified by PCR and the final PCR product was Solexa sequenced.

Processing of PAR-CLIP reads

The raw sequencing data was processed as previously described³. Briefly, after the removal of the Solexa adapters sequences that were either too short (less than 20 nucleotides) or too repetitive (for exact scoring parameters please refer to⁵) were discarded. The remaining sequences were mapped against the human genome (NCBI36 hg18 assembly) while allowing at most one error. Non-uniquely mapping reads were discarded and clusters were built from overlapping reads, requiring at least ten reads per cluster and at least one T to C change (“crosslinked clusters” or ”CCs”).

Annotation

Human genomic coordinates of RefSeq transcripts, including exon-intron boundaries and CDS start or stop positions, were obtained from the UCSC browser⁶. Transcripts in unfinished genomic segments or segments with unknown location (‘_random’ chromosomes) were disregarded, as were transcripts that

were not uniquely mapped so a single locus. Center positions of CCs were matched to transcript coordinates for the purpose of gene annotation, and to determine whether binding took place in the 5' UTR, CDS, 3' UTR or intronic regions. Due to multiple overlapping transcript isoforms, a minor fraction could be assigned a single such region and was thus classified as 'ambiguous'. Likewise, CCs that could not be assigned to a single specific gene were annotated as 'ambiguous' and disregarded in subsequent gene-centric analyses, as were clusters that matched to ribosomal RNA ('miscRNA' track in the UCSC browser). The annotation procedure was initially performed on individual reads, for the purpose of the clustering analysis (see below), and later on the final CCs.

Hierarchical clustering

Unfiltered binding profiles for each dataset were calculated by summarizing the total number of reads in each gene. These profiles were normalized based on the mean intensity in each dataset. Correlations between binding profiles were calculated based on the 5000 most variable genes as defined by the standard deviation across all datasets and using Spearman correlation as distance metric. Hierarchical clustering was performed using average linkage. After establishing reproducibility of replicates and uniqueness of the different proteins, replicate reads were pooled and clusters built and annotated as described above.

Site-level overlap analysis

We considered the top 1000 CCs in each datasets, as defined by the number of uniquely mapped reads and with the usual requirement of 25% T-to-C-containing reads. This approach simplifies downstream interpretation, as datasets that were originally of different sizes are thereby reduced to smaller and identically sized sets of high-ranking CCs. The methodology is similar to a previously described approach⁷. We screened for pairs of such CCs where each CC was from a different datasets and where center positions were within 10 nt of each other. These were considered to represent the same binding site. A small number of CCs with very high read counts (>10.000, maximum 3 cases in each pair wise comparison) were disregarded in this analysis as they contributed positively to correlation values while potentially being the products of erroneous short read mapping. The analysis was also repeated, with a similar outcome, while only considering exonic sites (not shown).

Cluster density plots

To determine the frequency of binding near intron-exon and exon-intron junctions, all individual exons were extracted from RefSeq transcripts. For each unique junction, all CCs binding with ± 200 bp were identified and the frequency of binding for each 4 bp segment in this 400 bp region was determined.

Gene ontology analysis

Gene Ontology (GO) annotations were obtained from ENSEMBL using the BioMart tool⁸, and enrichment of GO terms was evaluated statistically using Fisher's exact test. Genes that could not be mapped to the ENSEMBL GO annotation were excluded from the analysis. To account for multiple testing (30164 GO terms were evaluated), the observed P-values were compared to a simulated null-distribution (repeated scrambling of gene identities). A P-value of 1e-4 was found to be useful as a conservative threshold, as no false positives were observed at this level.

Motif discovery

To elucidate the RRE of FET proteins, we analyzed CCs as well as CCs with ≥ 2 T-to-C positional changes, as these could potentially show a stronger motif signal. The CCs were enriched for A and T as compared to random intronic regions, but use of standard bioinformatic tools^{9,10} as well as screening for short (4 nt) overrepresented sequence patterns (relative to mono- or dinucleotide shuffled sequences) did not return a significant motif. Stem-loop structures were mapped using the Matlab Bioinformatics Toolbox, and were required to have perfectly complementary stems of at least 3 bp and loops of 3 bp or longer.

Supplemental References

1. Kwiatkowski, T.J., Jr. et al. Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. *Science* **323**, 1205-8 (2009).
2. Vance, C. et al. Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. *Science* **323**, 1208-11 (2009).
3. Hafner, M. et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129-41 (2010).
4. Hafner, M. et al. Identification of microRNAs and other small regulatory RNAs using cDNA library sequencing. *Methods* **44**, 3-12 (2008).
5. Berninger, P., Gaidatzis, D., van Nimwegen, E. & Zavolan, M. Computational analysis of small RNA cloning data. *Methods* **44**, 13-21 (2008).
6. Kent, W.J. et al. The human genome browser at UCSC. *Genome Res* **12**, 996-1006 (2002).
7. Kishore, S. et al. A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. *Nat Methods*.
8. Smedley, D. et al. BioMart--biological queries made easy. *BMC Genomics* **10**, 22 (2009).
9. Bailey, T.L. & Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* **2**, 28-36 (1994).
10. Larsson, E., Lindahl, P. & Mostad, P. HeliCis: a DNA motif discovery tool for colocalized motif pairs with periodic spacing. *BMC Bioinformatics* **8**, 418 (2007).

9. Acknowledgements

“Success is a journey, not a destination. The doing is often more important than the outcome”- Arthur Ashe.

I thank everyone who was part of this journey and helped me to achieve my goal.

I would like to first of all thank Dr. Jessica Hoell for all the guidance and mentoring during my PhD. She was always there to lend an ear, be it for scientific discussion or a friendly chat.

A special thanks to Prof. Dr. Arndt Borkhardt for giving me the opportunity to work in his lab and for excellent supervision during my project. I would also like to thank Prof. Dr. Dieter Willbold for continuous evaluation of my thesis as second referee.

Furthermore I would like to thank my best buddy and colleague Kebria Hezaveh who was with me during my ups and downs and helped me in all my trivial times.

Thanks to my other group members Steffi, Kunal, Mareike for their kind support and helpful suggestions. I also owe a big thanks to Andreas kloetgen and Phillipp Muench for all the help with NGS data.

Thanks to all my lab members Sanil, Cyrill, Schafiq, Deborah, Micheal, Kati, Katharina, Teresa, Svenja, Nadja, Franzi, Julia, Sven, Yvonne, Andrea, Ute, Daniel 1, Daniel 2, Silke, Bianca, Frau Oellers, Claudia, Birgit, Özer, Rene, Nadine, Heidi and Frau Wieczorek for the great support in the lab and Frau Lesch for all the official help.

Thanks to my sisters Rajitha and Vijitha who were the two shoulders I could lean onto everytime. Also, a very special thanks to Cecilia Braun-Müller and Hubert Müller for all the love and support. Thanks to my in laws Vijaya and Kondiah for their wonderful support and understanding.

Finally, I would like to thank my husband and my best friend Sravan for his unending love and encouragement during all these years.

Last but not least, I would like to thank my parents Annamary and Sudhakar for being the wind under my wings and helping me to believe that dreams can be achieved.

I finally thank God for his abundant blessings.

Curriculum vitae

Personal details – Sujitha Smruthy Duggimpudi

Date of Birth: 09.07.1986 in Velgonda, India
E-Mail: duggimpudi@med.uni-duesseldorf.de
Telephone: +49 0211 81 19097
Address: Klinik für Kinder-Onkologie, -Hämatologie und Klin. Immunologie
Moorenstraße 5, 40225 Düsseldorf, Germany

Akademic profile

Since 02/2011 PhD at the Department of Paediatric Oncology, Haematology and clinical Immunology at *University Hospital, Duesseldorf, Germany*
➤ Research on posttranscriptional gene regulation of the RNA-binding proteins EWSR1, MSI1 and MSI2
09/2008 – 08/2010 Masters in Molecular Biosciences, *Ruprecht Karls Universität, Heidelberg, Germany*
➤ Final grade: 2,4
04/2005 – 05/2008 Bachelors in Biotechnology, Genetics and Chemistry, *Loyola Academy, India*
➤ Final grade: 81/100
06/2002 – 04/2004 Pre University, Secunderabad, *India*,
➤ Final grade: 77/100
06/1991 – 04/2002 Secondary school education, Secunderabad, *India*
➤ Final grade: 80/100

Languages known and IT

Languages: Telugu (Mother tongue), English (Fluent), German (Intermediate)
BioInformatics: Office, Photoshop, Illustrator, Vector-N'TI, Endnote, Graphpad prism

Poster & Talks

Duggimpudi S (2012): "*Characterisation of RNA-protein interaction networks using PAR-CLIP*"; BMFZ Symposium, Düsseldorf, Germany. (Talk)

Duggimpudi S, Hoell JI, Larsson E, Tuschl T, Sander C, Borkhardt A (2012): "*Role of the RNA-binding protein EWSR1 in Ewing Sarcoma*"; Keystone Meeting on RNA binding proteins, Sante Fe, USA. (Poster)

Duggimpudi S, Larsson E, Hoell JI, Borkhardt A (2013). "RNA binding protein EWSR1 regulates CCDC6"; RNA society meetings, Davos, Switzerland. (Poster)

Publications

Duggimpudi S., Larsson E., Nabhani S., Borkhardt A., Hoell JI. "*The cell cycle regulator CCDC6 is a key target of RNA-binding protein EWS*" PLOS One, 2015,10-1371

Hoell JI., Larsson E., Runge S., Nusbaum J., **Duggimpudi S.**, Farazi T., Hafner M., Borkhardt A., Sander C., Tuschl T. "*RNA targets of wild-type and mutant FET family proteins.*" Nat Struct Mol Biol, 2011, 18:1428-31

Gigova A., **Duggimpudi S.**, Pollex T., Schaefer M., Koš M. "*A cluster of methylations in the domain IV of 25S rRNA is required for ribosome stability.*" RNA, 2014, 10:1632-44

Doose G, Wojciech F., Bernhart S., Haake A., **Duggimpudi S.**, Lopez C., Hoell JI., Borkhardt A., Hoffmann S., Siebert R, Iaccarino I. "*MINCR1 is a novel MYC-induced lncRNA able to modulate MYC's transcriptional network in Burkitt lymphoma cells*". (under review at PNAS)

Duggimpudi S., Muench P., Kloetgen A., Hezaveh K., Shaikalishahi H., Borkhardt A., Hoell JI. "*Transcriptome-wide mRNA targets of the RNA-Binding protein MSI2 involved in myeloid malignancies*" (Manuscript in preparation)

Affirmation

Hereby, I declare on oath that I composed this dissertation independently and self consistently. I used only the references and resources indicated in this thesis. This PhD thesis was never submitted in presented or similar form to any other institution or examination board and I have not ventured a doctoral examination without success so far.

Düsseldorf, 27/4/2015

Sujitha Smruthy Duggimpudi