Universidade de Lisboa

Faculdade de Ciências

Departamento de Biologia Vegetal

MicroRNA regulation by the DNA and RNA binding proteins EWS and FUS

Ana Miguel Guterres Coelho Fernandes

Mestrado em Biologia Molecular e Genética

2012

Universidade de Lisboa

Faculdade de Ciências

Departamento de Biologia Vegetal

MicroRNA regulation by the DNA and RNA binding proteins EWS and FUS

Dissertação orientada por

Professor Doutor Anders Lade Nielsen

(Department of Biomedicine, Aarhus University, Denmark)

Professora Doutora Margarida Telhada

(Departamento de Química e Bioquímica, Faculdade de Ciências, Universidade de Lisboa, Portugal)

Ana Miguel Guterres Coelho Fernandes

Mestrado em Biologia Molecular e Genética

2012

Table of Contents

Appendix:

Figures:

Acknowledgements

It has been an intense year and a challenging project, which would not have been finished without the help of people very dear to me.

Firstly I would like to thank Anders Lade Nielsen, my supervisor, the opportunity he has given me to work in his laboratory, all the discussions and ideas, the careful guidance, advices and motivation which allowed me to learn and grow so much during this year. A special thanks to Hanne for all the help, good advices when experiments went wrong and for all the techniques I learned from her. I would like to thank Jenny for all the discussions about which direction to go, good tips and help throughout the project. I would also like to thank Tina, Christina, Rune, Johan, Jakob and Emilie for all the lab talks, all the morning meetings and for creating an awesome atmosphere in the lab. I loved to be part of it, it was a great year.

To the department of Biomedicine, I thank you for the good work environment, collaboration, all the coffee and cake breaks and parties.

I would like to thank Professor Margarida Telhada, my coordinator at my home university, for all the guidance and important recommendations throughout the development of this paper.

To all my friends in Aarhus, Tobias, Karen, Sara, Denis, Dan and Sofie, thank you for all the dinners and parties, great moments, for introducing me to the Danish life-style and for making me feel so much at home. To Maria and Magnus, thank you for being my family in Aarhus. It was great to live with you. Maria, thank you for all the friendship, talks and support when I was freaking out about what to do with my thesis.

To all my great friends in Lisbon that were far away but will always have a place in my heart: Andreia, Bia, Bota, Joana, Beatriz, Ana Maria. Without you I wouldn’t be who I am today. Thank you for all the great moments in Lisbon.

I would like to thank my family: my grandmothers Ana e Felícia, my uncles Palmira, Luís ,Tó and Vítor and my cousins Inês, Filipe and Francisco. Thank you for all the love, for helping me grow up and for being best family one could ever wish for.

Lastly, I thank my parents Ana Paula e Domingos for being with me every step of the way, for every laugh and every tear, for the unconditional support, for teaching me what happiness and trust is. Without you I would not have fulfilled my dream. Thank you for being my parents, you are awesome!

Resumo

As proteínas EWS (*Ewing sarcoma breakpoint region 1*), FUS [*Fused in sarcoma* ou TLS (*Translocated in liposarcoma*)] e TAF15 (*TATA box binding protein-associated factor*) fazem parte da família de proteínas denominada FET (FUS, EWS e TAF15). A região amino-terminal destas proteínas é composta por um domínio de activação transcricional enquanto a região C-terminal é constituída pelo domínio de ligação ao RNA, onde existem locais de interacção com os ácidos nucleicos. Desta forma, as proteínas FET são capazes de se ligar a moléculas de RNA e DNA. Interagem também com várias proteínas e factores de transcrição, tendo sido implicadas em diversos processos celulares. As proteínas FET participam na regulação da transcrição, *splicing* geral e alternativo, reparação de DNA, transporte de RNA mensageiro (mRNA) e transitam entre o núcleo e o citoplasma (*nucleo-cytoplasmic shuttling*). Adicionalmente, foram identificadas num complexo proteico que inclui a enzima Drosha, designado "*Drosha large complex"*. A enzima Drosha é um dos principais constituintes do Complexo Microprocessador que é responsável pelo processamento de microRNA (miRNA) primários a miRNA percursores. Porém, o envolvimento das proteínas FET no processamento de miRNAs não foi ainda esclarecido. Estas proteínas foram também relacionadas com patologias humanas, tendo sido identificadas como fusões proteicas em vários tipos de cancro. Nestas fusões, o domínio de activação transcricional das proteínas FET encontrava-se fundido ao domínio de ligação ao DNA de outro factor de transcrição. Isto levava à produção de uma proteína quimérica com função celular alterada capaz de induzir tumorigénese. Estudos mais recentes implicaram as proteínas FET em doenças neurológicas. As proteínas FET foram encontradas em inclusões citoplasmáticas de doentes com Esclerose Lateral Amiotrófica, sendo FUS a proteína mais estudada. As proteínas FET foram também associadas com a doença Degeneração Lobar Frontotemporal.

Nos últimos anos, foram descobertos numerosos RNAs não codificantes denominados pequenos RNAs (*small RNAs*). Entre estes, os miRNAs são o grupo mais estudado, tendo-se revelado importantes reguladores da expressão génica a nível pós-transcricional. Estes, em conjunto com o Complexo de Indução do Silenciamento de RNA (RISC), despoletam a clivagem ou o silenciamento da tradução de mRNAs com os quais são complementares. A regulação exercida pelos miRNAs é muito ampla, abrangendo a maioria dos processos celulares. Assim, os miRNAs revelaram ter um papel importante na diferenciação celular e embriogénese, estando a sua expressão frequentemente desregulada em doença.

O principal objectivo deste trabalho foi esclarecer se as proteínas EWS e FUS participavam na regulação da expressão de miRNAs. Experiências de imunoprecipitação da cromatina realizadas anteriormente na nossa equipa demonstraram que as proteínas EWS e FUS se encontravam ligadas em várias regiões genómicas, apresentando uma tendência para se ligar em conjunto e na região 3' dos genes. Nos loci identificados estavam genes que codificavam miRNAs, entre eles o locus de *HNRNPK* (*heterogeneous nuclear ribonucleoprotein K*) onde o percursor pre-miR-7-1 é produzido para originar o miR-7. Além disso, o último exão do mRNA de hnRNPK contém um local de *splicing* alternativo, onde existem dois receptores de *splicing*, o proximal e o distal, que originam duas isoformas proteicas de hnRNPK. O presente estudo focou-se então neste locus já que apresentava a possibilidade de investigar a influência das proteínas EWS e FUS no processamento de miRNAs assim como no *splicing* geral e alternativo do mRNA de hnRNPK.

Para estudar a função das proteínas EWS e FUS realizámos uma dupla transfecção transiente de células HEK293 com pequenos RNAs de interferência (siRNA) de modo a inibir a expressão de ambas as proteínas. Foram feitos três tratamentos: (a) inibição de EWS (KOEWS), (b) inibição de FUS (KOFUS) e (c) inibição simultânea de EWS e FUS (KOEWS+FUS). As sequências dos siRNAs usados estão descritas na *Supplementary Table* 1. Para a inibição de cada proteína foi usada uma mistura de dois siRNAs. Como controlo, as células HEK293 foram transfectadas com um siRNA sem quaisquer mRNAs-alvo. Procedeu-se então à quantificação da expressão do miR-7, que se apresentou sobreexpresso nos tratamentos com KOEWS e KOEWS+FUS, em comparação com KOFUS e controlo. Quantificámos também o mRNA da proteína BCL2, um alvo *bona fide* do miR-7. De acordo com o esperado, verificou-se uma diminuição da expressão do mRNA de BCL2 nas células em que o miR-7 estava sobreexpresso. Seguidamente, desenharam-se vários conjuntos de *primers* para amplificar diferentes regiões ao longo do mRNA de hnRNPK de forma a perceber se a repressão de EWS, FUS, ou ambas afectaria o *splicing*. De facto, nos tratamentos com KOFUS e KOEWS+FUS, verificou-se uma acumulação de mRNA não processado (*unspliced* mRNA), na região 3' do mRNA de hnRNPK. No entanto, identificou-se também que as células onde a proteína EWS estava inibida (KOEWS e KOEWS+FUS) se dividiam menos e apresentavam alterações morfológicas em comparação com os restantes tratamentos. Experiências adicionais mostraram que a proteína TAF15 estava inibida nos tratamentos com KOEWS e KOEWS+FUS devido a efeitos *off-target* dos siRNAs usados na repressão de EWS. Assim, substituíram-se estes siRNAs por um novo siRNA sem complementaridade com TAF15 (*Supplementary Table* 1).

Procedemos então à repetição das experiências de dupla transfecção com um novo siRNA para a inibição de EWS, realizando os seguintes tratamentos: KOEWS, KOFUS e

KOEWS+FUS. Ao quantificar o miR-7 nas novas amostras, observou-se um aumento de expressão (~ 50%) apenas no tratamento KOEWS+FUS. Além disso, não se verificou repressão do mRNA de BCL2 no mesmo tratamento. De forma a esclarecer se EWS ou FUS estavam a afectar o processamento dos percursores do miR-7, pri-miR-7-1 e pre-miR-7-1, procedeu-se à quantificação dos mesmos. Ao contrário do que seria esperado, observámos que a expressão de pri-miR-7-1 encontrava-se diminuída nas células KOFUS, enquanto nos restantes tratamentos específicos não ocorreu qualquer alteração dos níveis de expressão do pri-miR-7-1. Além disso, a expressão do pre-miR-7-1 não sofreu qualquer variação.

Seguidamente, foi quantificada a expressão do mRNA de hnRNPK usando conjuntos de *primers* acima mencionados. Os resultados obtidos foram semelhantes aos anteriores: verificou-se uma acumulação significativa de mRNA não processado na região 3' do mRNA de hnRNPK com os tratamentos KOFUS e KOEWS+FUS, quando comparados com o controlo. Realizou-se também um ensaio para averiguar se as proteínas EWS e FUS afectavam o *splicing* alternativo no último exão do mRNA de hnRNPK. Para isso, desenhou-se um grupo de *primers* que permitiam a amplificação de duas bandas com dimensões distintas, correspondentes aos transcritos derivados do receptor de *splicing* proximal ou distal. Verificou-se que nas células controlo (tratadas com siRNA não específico, i. e., sem mRNAs-alvo) o receptor de *splicing* proximal era preferencialmente usado. No entanto, nas amostras em que as proteínas EWS, FUS ou ambas se encontravam inibidas, especialmente com KOFUS, ocorreu uma diminuição no uso do receptor de *splicing* proximal e um aumento no uso do receptor de *splicing* distal. Finalmente procedeu-se ao estudo da influência da inibição das proteínas EWS e FUS na produção da variante do mRNA de hnRNPK a partir do exão não-codificante alternativo presente na região a 5' do mRNA de hnRNPK. Observou-se que a inibição da proteína FUS provocou um aumento da variante do mRNA de hnRNPK correspondente ao exão alternativo.

Considerando os resultados obtidos, as proteínas EWS e FUS parecem regular a produção do miR-7 e o *splicing* do mRNA de hnRNPK. A proteína FUS aparenta inibir a produção de miR-7 a dois níveis: primeiramente, impedindo a clivagem do pri-miR-7-1 pela enzima Drosha no núcleo, e posteriormente reprimindo o processamento do pre-miR-7-1 pela enzima Dicer no citoplasma, com a ajuda da proteína EWS. Para além disso, as proteínas EWS e FUS influenciam o processamento do mRNA de hnRNPK na região 3' assim como a escolha do receptor de *splicing* alternativo usado. Novamente, a proteína FUS parece ter um papel predominante, sendo que a inibição da proteína EWS não evidenciou efeitos relevantes.

A região 3' do mRNA de hnRNPK contém um local de splicing alternativo e um intrão codificante para um miRNA que são processados por maquinarias celulares distintas,

nomeadamente da transcrição, do *splicing* e do processamento de miRNAs. Estudos anteriores mostraram que estas se influenciam mutuamente e estão estreitamente interligadas. Desta forma, é provável que os diversos processos a ocorrer neste local estejam a ser regulados por proteínas que os integram e coordenam. Considerando a presença das proteínas EWS e FUS na região a 3' do DNA de *HNRNPK* e a sua capacidade de se ligar a RNAs, é possível que estas funcionem como coordenadoras dos processos celulares a ocorrer na região a 3' do mRNA de hnRNPK.

Palavras chave: EWS, FUS, processamento de microRNAs, *splicing*, RNA de interferência.

Abstract

EWS and FUS proteins belong to the FET protein family, which has been implicated in cancer and neurological diseases. Their distinct DNA and RNA binding properties and their broad spectrum of protein interactions allow them to participate in various cellular processes. So far, they have been implicated in transcription, splicing, DNA repair, mRNA transport and nucleo-cytoplasmic shuttling. They have also been found in the same complex as Drosha, indicating a role in microRNA processing that has not yet been clarified. Over the past years, microRNAs have emerged as key posttranscriptional regulators, being able to influence most cellular pathways by fine-tuning and buffering internal variations. Accordingly, microRNAs have been implicated in cell differentiation and development and often appear deregulated in human pathologies.

The aim of this study was to clarify the role of EWS and FUS in microRNA function. We focused on *HNRNPK* locus, one of the production sites of miR-7, since EWS and FUS were found to bind to the DNA at the 3' end of *HNRNPK.* siRNA-mediated depletion of EWS and FUS indicates that both proteins have a role in miR-7 processing. FUS inhibits pri-miR-7-1 cleavage by Drosha, in the nucleus, and together with EWS is able to further hinder pre-miR-7-1 Dicer cleavage, in the cytoplasm. Moreover, both proteins are able to influence alternative splice acceptor site usage at the 3' end of hnRNPK mRNA as well as to regulate splicing in this region. Overall, FUS appears to be the main regulator whereas EWS has a supplementary role. The proposed hypothesis is that FUS, with the help of EWS, is possibly coordinating transcription, splicing and microRNA processing at the 3' end of hnRNPK mRNA.

Key words: EWS, FUS, microRNA processing, splicing, small interference RNA

Manuscripts

The results obtained in this study will be included in the following paper:

- Blechingberg J. **Fernandes A.M.**, Luo Y., Li S., Nielsen A. L. Identification of the DNA binding sites of the FUS and EWS proteins reveals binding to the 3' end of genes. *Manuscript in preparation*

Abbreviations

AGO	Argonaute
ALS	Amyotrophic lateral sclerosis
BCL2	B-cell CLL/lymphoma 2
bp	Base-pair
cDNA	Complementary DNA
C-terminal	Carboxyl-terminal
DGCR8	DiGeorge syndrome critical region 8
DMEM	Dulbeccoo's modified Eagle's medium
DNA	Deoxyribonucleic acid
DSA	Distal splice acceptor
EGFR	Epidermal growth factor receptor
EWS	Ewing sarcoma breakpoint region 1
FUS	Fused in sarcoma
G-rich	Glycine-rich
HEK293 cells	Human kidney embryo cells
hnRNP	Heterogeneous nuclear ribonucleoprotein
IGF1R	Insulin-like growth factor 1 receptor
LPHN2	Latrophilin 2
miRNA	MicroRNA
mRNA	Messenger RNA
PAK1	p21-activated kinase 1
PBS	Sodium Perborate
Pre-miRNA	Precursor miRNA
Pre-mRNA	Precursor mRNA
Pri-miRNA	Primary miRNA
PSA	Proximal splice acceptor
qPCR	Quantitative real-time PCR
RAF1	V-raf murine leukemia viral oncogene homolog 1
RGG-rich	Arginine-glycine-glycine-rich
RISC	RNA induced silencing complex
RNA	Ribonucleic Acid
RNA Pol II	RNA Polymerase II
RNU48	Small nucleolar RNA C/D box 48
RRM	RNA recognition motif
siRNA	Small interfering RNA
S, Y, G, Q-domain	Serine, glycine, glutamine and tyrosine-rich domain
TAF15	TATA box binding protein-associated factor 15
TBP	TATA box binding protein
TFIID	Transcription factor II D
TF	Transcription factor
UTR	Untranslated region
Zn finger	Zinc Finger

Introduction

Proteins EWS and FUS are involved in human disease as well as in numerous cellular processes, such as transcription and splicing. Recent evidence indicates that they may also be implicated in microRNA processing. It is known that transcription, splicing and microRNA processing are intimately connected and thus is likely that these processes are coordinated by common proteins. Several proteins are considered good candidates for this role, being among them EWS and FUS.

1. The FET protein family

The proteins Ewing sarcoma breakpoint region 1 (EWS), Fused in sarcoma [FUS or Translocated in liposarcoma (TLS)) and TATA box binding protein-associated factor (TAF15) are part of the FET protein family (FUS, EWS and TAF15)[1]. The FET proteins have ubiquitous expression patterns and localize mainly in the cell nucleus[2].

The first protein to be discovered was EWS, where it was fused to the transcription factor (TF) FLI-1 in Ewing's sarcoma family of tumors[3]. This was followed by the identification of FUS fused with CHOP in human myxoid liposarcomas[4; 5]. TAF15 was identified as an interaction partner of Transcription factor II D (TFIID)[1; 6]. TFIID is part of the transcription initiation complex, along with RNA polymerase II (RNA pol II) and the general transcription factors[7]. TAF15 was included in the FET protein family due to its homology with EWS and FUS[1].

1.1. Structure of FET proteins

The FET proteins are constituted by several homologous domains which confer them unique properties[8]. The amino-terminal is composed of serine, glycine, glutamine and tyrosine degenerate repeats (S, Y, G, Q- domain) which can act as a potent transcriptional activation domain when fused to other TFs in cancer associated translocations (Figure 1)[9; 10].

The other regions of the FET proteins are implicated in nucleic acid binding. The RNA-binding domain is composed of a glycine-rich (G-rich) region, an RNA recognition motif (RRM), a RanB2-type zinc finger (Zn-finger) and a carboxyl-terminal (C-terminal) arginine-glycine- glycine- rich (RGG-rich) domain (Figure 1). These domains confer the FET proteins the ability to bind both DNA and RNA. The Zn-finger domain of the FET proteins was identified as Cys_2/Cys_2 Zn-finger, a subtype usually involved in DNA and protein binding[11; 12]. Recently, it was further categorized as a RanB2-type Zn-finger which was shown to bind single stranded RNA[13; 14]. Although it was previously thought that the Zn-finger domain was responsible for the DNA binding properties of the FET proteins, studies done by Loughlin *et al* (2009) indicated that this domain may be implicated only in RNA binding, therefore

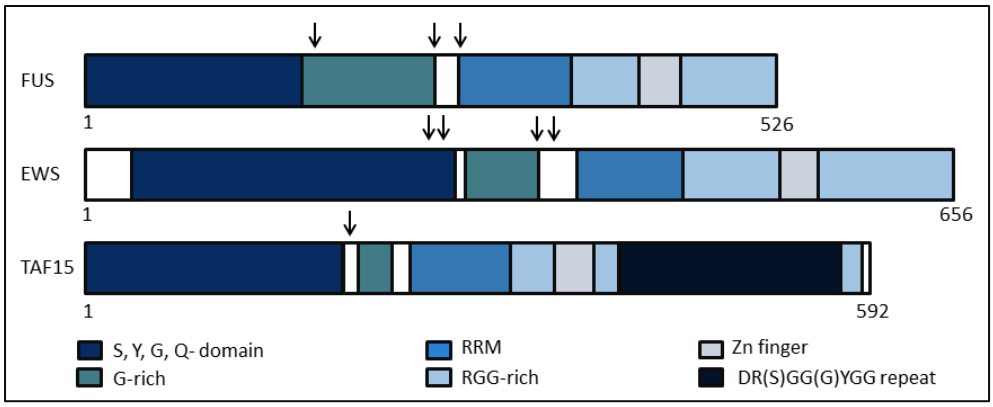


Figure 1 | Schematic representation of the FET proteins structure. Here are depicted the several domains that constitute the FET proteins. The amino-terminal is constituted by an S, Y, G, Q-rich domain. The C-terminal is composed of a G-rich domain, a RRM domain and RGG-rich repeats intercalated by a Zn-finger domain. In addition, TAF15 contains DR(S)GG(G)YGG repeats. The black arrows indicate the breakpoints of the fusion proteins in cancer associated chromosome translocations. *Adapted from Moroshi et al, 1998; Tan and Manley, 2009 and* http://www.uniprot.org/.

showing that this issue needs to be further clarified. The most conserved domain in the FET protein family is the RRM[15]. It folds into a secondary structure which is involved in sequence-specific RNA binding[11; 16-18]. The RGG-rich domains also appear to be relevant for RNA binding specificity[18; 19]. Hence, it is probably the cooperation between the RRM and the RGG-rich domains that provides optimal specificity for RNA binding. The RGG-domain of TAF15 is slightly different from the other FET proteins, consisting of several DR(S)GG(G)YGG repeats, which possibly confers different RNA-binding specificity[8]. This way, FET proteins have been reported to bind RNA as well as double-stranded and single-stranded DNA[1; 11; 17; 19]. Recently, Hoell *et al* (2011) have determined the RNA targets of FET proteins and all proteins were frequently bound to RNAs, including highly expressed cellular messenger RNAs (mRNAs). So far, EWS and FUS have been shown to interact with specific RNA sequences. Both proteins are able to bind to poly-G and -U RNA[1; 17; 20-22]. It was further determined that FUS interacts with GGUG sequences[18] and that it binds strongly to AU-rich stem-loop structures[22].

The FET proteins have nuclear localization[2] which implies that they need a nuclear localization signal to be imported from the cytoplasm. FUS, EWS and TAF15 have very similar nuclear localization signals located at the C-terminal and their transport to the nucleus is mediated by Transportin[23-25]. EWS nuclear import also depends on its C-terminal phosphorylation[26] and it was recently shown that FUS has an additional nuclear export signal within the RRM[27].

1.2. Cellular function of FET proteins

Since their discovery, FET proteins have been implicated in numerous cellular processes, which are described in the text bellow. One of such processes is transcription, which is mediated by a protein complex constituted by RNA Pol II associated with several TFs[7].

Among these is TFIID, which is composed of TATA box binding protein (TBP) plus TBP-associated factors. TFIID is involved in the initiation and elongation steps of transcription[7]. FET proteins were found to interact with TFIID[1; 6; 28] and were identified in a complex with RNA Pol II, where they interacted with several and different subunits (Figure 2 a)[1; 28; 29]. Similarly, FET proteins can alter gene transcription by binding to different TFs (Figure 2 b). Specifically, EWS was found to interact with Brn-3a, a TF which regulates neuronal differentiation, inhibiting its activity[30]. EWS was also found to bind to Oct-4, a TF implicated in maintenance of totipotency of embryonic stem cells, thereby enhancing its activity[31]. FUS was found to bind to STAT1[30], although its function is still unknown, and to various nuclear hormone receptors[32]. Finally, pull-down experiments have shown that FET proteins bind each other[33], possibly working as a complex.

The ability to specifically bind to DNA and RNA is the most distinct feature of the FET protein family and it greatly influences the cellular processes mediated by them. Accordingly, FUS

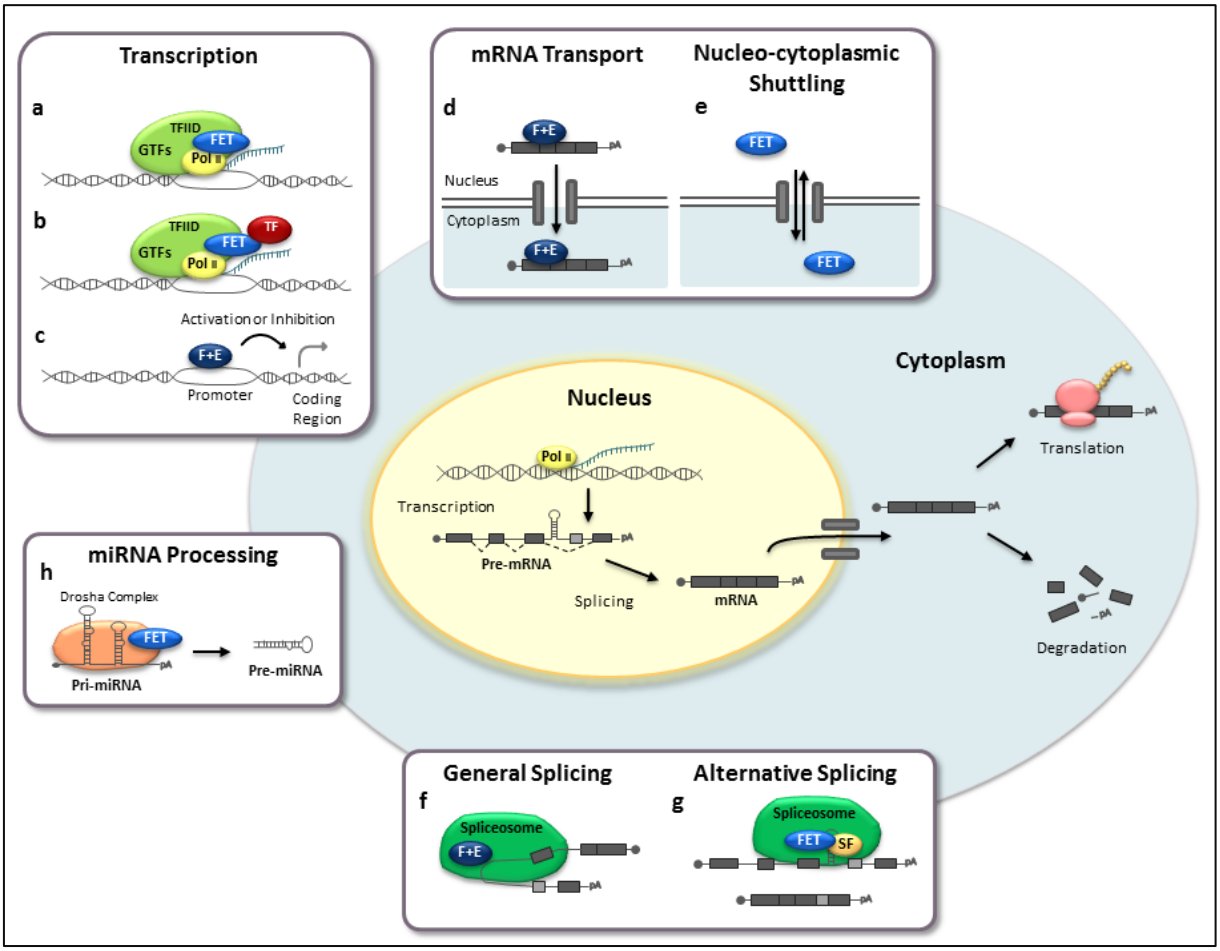


Figure 2 | Representation of the FET proteins main cellular roles. In the center is depicted the canonical pathway for transcript processing and translation. The white boxes highlight the steps where FET proteins intervene. FUS, EWS and TAF15 are represented by the light blue circle designated FET. Only FUS and EWS are implicated in some processes, being represented by the dark blue circle designated F+E. (a) FET proteins interact with general transcription machinery (RNA Pol II, GTFs and TFIID) as well as (b) with associated transcription factors (TF). (c) FUS and EWS bind to promoter elements of genes thus activating or inhibiting gene transcription. (d) FUS and EWS are involved in mRNA transport and (e) all FET proteins can shuttle between the nucleus and the cytoplasm. (f) FUS and EWS contribute to the general splicing process by being part of the spliceosome and (g) all FET proteins are able to interact with splicing factors (SF) in alternative splicing. (h) FET proteins are part of the Drosha complex, possibly influencing microRNA processing.

was implicated in mRNA transport (Figure 2 d) in dendritic spines. Upon activation of the glutamate receptor 5, FUS accumulated at excitatory postsynapses, which was concomitant with an increase in the dendritic mRNA level[34]. FET proteins are also able to shuttle between the nucleus and the cell surface or nucleus and the cytoplasm, respectively (Figure 2 e)[20; 25; 35], indicating a role in nuclear mRNA export. Moreover, FET proteins seem to participate in DNA repair and genomic stability. All FET proteins were found to mediate pairing of homologous DNA ends, a fundamental step in DNA double-strand break repair[19; 36; 37]. Additionally, knockout mice for EWS and FUS showed similar defects, such as increased sensitivity to ionizing light, reduced meiotic recombination in spermatocytes and B-cell development abnormalities[38-40]. Moreover, it was reported that in the presence of DNA damage, FUS was able to bind non-coding RNAs transcribed and tethered to the chromatin until approximately 2 Kb upstream of the cyclin D1 promoter[41]. The recruitment of FUS by non-coding RNAs allowed the inhibition of CREB-binding protein and p300 histone acetyltransferase activities, thereby preventing cyclin D1 gene activation[41]. Besides this, there have also been reports FUS and EWS proteins binding directly to gene promoters (Figure 2 c): FUS was found to bind to promoters of oxidative stress related genes[42] and EWS to the promoter of CSF-1 receptor gene[43].

Extensive studies associated the FET proteins in precursor mRNA (pre-mRNA) splicing and maturation. FUS and EWS were identified as part of the spliceosome-associated factors (Figure 2 f)[44]. FUS and TAF15 were found to interact with U1 small nuclear ribonucleoprotein[11; 45], also part of the spliceosome machinery (reviewed in Wahl *et al*[46]). In addition, FET proteins interact with numerous splicing factors and mediators (Figure 2 g). FUS was initially identified as the heterogeneous nuclear ribonucleoprotein (hnRNP) P2, which is part of a group of RNA-binding nuclear proteins involved in pre-mRNA maturation[47]. Furthermore, EWS could bind to hnRNP M and U whereas FUS could bind to hnRNP A1 and C1/C2[9; 33]. Both EWS and FUS were found interacting with the splicing factors YB-1 [48] and serine-arginine proteins, such as TASR and SC35[49; 50]. FUS was further identified binding to PTB, SRm160 and Spi-PU.1 splicing factors[51; 52] and seemed to promote the usage of distal 5' splice sites[52]. EWS and FUS, along with the above mentioned splicing factors, were able to modify the splicing pattern of a reporter gene, with the C-terminal domain responsible for the binding of both proteins to the splicing factors. In cancer-associated fusions, the C-terminal domain of the FET proteins would be absent and the splicing factors would not able to interact with the FET proteins, thus altering alternative splicing patterns. Finally, a recent study has implicated EWS in alternative splicing of DNA damage-induced genes mediated by its co-transcriptional binding to pre-mRNA[53], which may explain the results obtained in EWS knockout mice described above.

Finally, FET proteins might be implicated in the regulation of microRNAs (miRNAs), as they were identified as part of Drosha large complex (Figure 2 h)[54]. FUS was also found to interact with DGCR8[55], a fundamental protein in miRNA processing. Not much is known about the significance of this binding and so this topic will be addressed in further detail bellow.

1.3. FET proteins and disease

FET proteins have been largely implicated in human pathologies, such as cancer and neurological diseases. As it was mentioned above, FET proteins were found as fusion proteins with other TFs in several types of malignant tumors. Besides EWS-FLI1 and FUS-CHOP fusion proteins, both EWS and FUS were identified fused with several other TFs, such as ERG and ATF1[56]. TAF15 was also found as a cancer associated fusion protein with TAC in extraskeletal myxoid chondrosarcoma[57]. Interestingly, the transcription activation domain seems to be functionally interchangeable among FET proteins: the transcription activation domain of one of FET proteins can be substituted by the same domain of another FET protein and is still able to originate tumors[9; 57]. Oppositely, the TF to which the FET protein is fused seems to determine the tumor phenotype[9; 56]. Fusion proteins originate when the breaking point of a chromosomal translocation is in the middle of a gene which is then transported to a different locus and attached to a another gene, therefore producing an aberrant protein[58]. In the case of FET proteins, their amino-terminal is always fused to the DNA binding domain of another TF resulting in a chimeric TF with altered cellular function (Figure 1)[59]. This process occurs in one of the alleles and thus the tumor cell expresses both aberrant and normal proteins. It was shown that both FET protein forms were strongly expressed in tumor cells[60] and so it is thought that there is mainly a gain of function of the fusion protein, which will have different cellular functions and ultimately originate a tumor[58].

Additionally, FET proteins have been recently associated with neurological disorders. Two independent studies found genetic mutations in FUS in patients with Amyotrophic Lateral Sclerosis (ALS) [61; 62]. ALS is a neurodegenerative disease which leads to a progressive skeletal muscle atrophy, causing death after 2-5 years from symptoms onset by respiratory failure[63]. About 10% of the ALS cases have genetic causes, whereas the remaining cases are sporadic[63; 64]. In both situations, the mutated FUS was found aggregated in the cytoplasm of neuronal cells[61; 62; 65]. Furthermore, it was shown that ALS-related FUS mutants, which lack the C-terminal domain, were not imported to the nucleus and thus presented cytoplasmic mislocalization[24; 66]. Interestingly, mutant but not normal FUS co-localized with stress granule specific proteins implying that FUS cytoplasmic accumulation induced stress granule formation[66; 67]. Stress granules are non-membranous cellular foci composed of translationally stopped ribonucleoproteins which form when cells are exposed to various environmental stresses, such as heat shock or oxidative stress[68]. This may indicate that altered RNA

metabolism is an underlying cause of ALS[66]. Moreover, EWS and TAF15 were recently identified as candidate genes for ALS[69-71], being also present in cytoplasmic aggregates in neurons of ALS patients[69; 71]. However, TAF15 cytoplasmic inclusions were not present in FUS-mutated ALS patients[72]. Finally, all FET proteins were associated with Frontotemporal Lobar Degeneration where they similarly form cytoplasmic aggregates[72]. Protein aggregates have been considered a hallmark in numerous neurodegenerative disorders, such as Alzheimer or Parkinson diseases, which may point to a common pathological mechanism[65].

2. MicroRNA biogenesis and function

miRNAs are ~22 base pair (bp) non-coding RNAs which are able to posttranscriptionally regulate gene expression[73]. miRNAs are able to bind target cellular mRNAs, usually at their 3' untranslated region (UTR), by full or partial Watson-Crick base-pair complementarity of the nucleotides 2-8, the seed region. This biding induces posttranscriptional silencing, which can be accomplished by two different ways: mRNA degradation or translation repression[74]. They are the most well studied family of small RNAs, which also includes the families of endogenous small interfering RNAs and Piwi-interacting RNAs[73]. miRNAs are thought to control about 50% of all protein-coding genes in mammals, participating in the regulation of most of cellular processes[75]. The number of cellular miRNAs is equivalent to those of TFs and RNA binding proteins. miRNAs are expressed in a development-specific or tissue-specific manner, allowing them to be key players in embryogenesis and cell differentiation[75].

miRNAs are usually transcribed by RNA Pol II and originate from a capped and polyadenylated primary transcript, the pri-miRNA (Figure 3 a), which folds into a stem-loop secondary structure[76]. Then, the nuclear RNase III- type enzyme Drosha cleaves the pri-miRNA at the base of the stem-loop, producing an ~ 70 bp hairpin, the pre-miRNA (Figure 3 a)[75; 77; 78]. Drosha acts together with DGCR8, both being part of a protein complex, the Microprocessor Complex. DGCR8 helps in the definition of the pre-miRNA cleavage site[54; 79].

Pri-miRNAs can be produced from different genomic environments: intergenic or intragenic. When intergenic, the miRNA expression is coordinated with the expression of other miRNAs. These are transcribed as a miRNA cluster which is considered the canonical miRNA gene (Figure 3 a)[80]. When intragenic, miRNAs are almost exclusively expressed from introns. In this instance, pri-miRNAs appear to be processed co-transcriptionally (Figure 3 b)[81; 82]. The exons flanking the miRNA-encoding intron are thought to be tethered to the commitment complex, i.e. the early spliceosome complex, and Drosha cleaves the pri-miRNA before the intron is removed by the spliceosome[82]. Finally, some miRNAs, named mirtrons, are produced from very short introns and can completely bypass Drosha processing[83; 84] (Figure

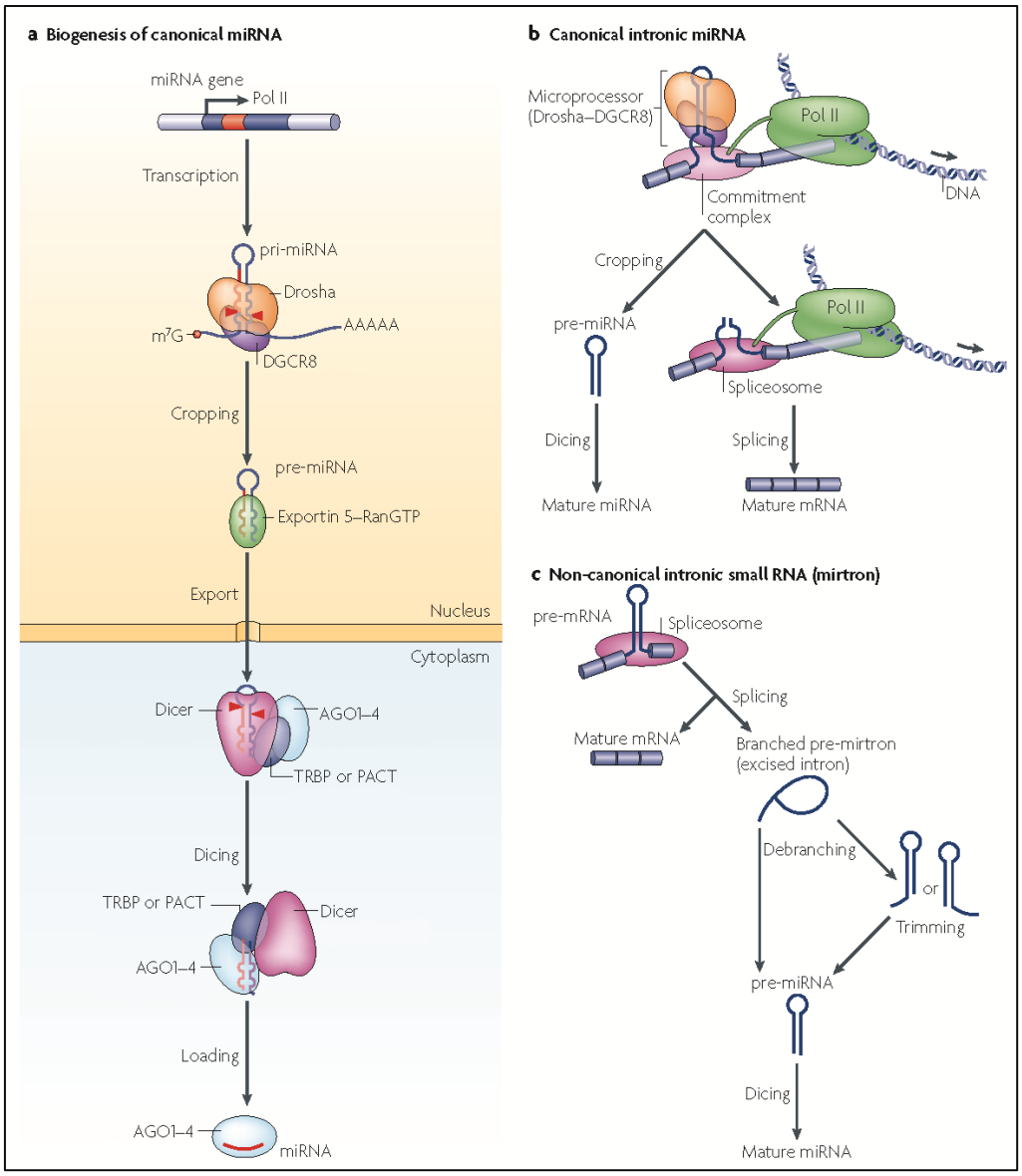


Figure 3 | Representation of miRNA biogenesis pathway. (a) Canonical miRNA genes are transcribed by RNA Pol II and originate the pri-miRNA. The pri-miRNA is cropped by Drosha and DGCR8 (Cropping), which comprise the Microprocessor Complex, generating the pre-miRNA. The pre-miRNA is exported to the cytoplasm and is further processed by Dicer (Dicing) producing the mature miRNA. Dicer is in a complex with TRBP and/or PACT and AGO (1-4), which mediate pre-miRNA processing and RNA-induced silencing complex (RISC) assembly. The mature miRNA is loaded onto one of the AGO proteins and unwound. One of the strands remains bound to RISC and mediates posttranscriptional silencing, whereas the other is degraded. (b) After being transcribed, canonical intronic miRNAs are processed co-transcriptionally. The miRNA-coding intron is thought to be tethered to the Commitment Complex (early spliceosome) and the miRNA hairpin is processed by Drosha and DGCR8 before splicing. The resulting pre-miRNA is exported to the cytoplasm and follows the canonical miRNA processing pathway while the mRNA is normally spliced and translated. (c) Non-canonical intronic small RNAs (mirtrons) originate from small introns after being spliced and debranched. This yields a pre-miRNA-like small RNA which is directly exported to the cytoplasm without being processed by Drosha. Some introns have tails at their 5' or 3' end and thus need to be trimmed before being exported. m^7G, 7-methylguanosine. *Modified from Kim* et al *(2009)*

3 c) since after intron splicing and debranching the small RNA folds into a pre-miRNA-like hairpin structure[83; 84].

Once formed, the pre-miRNA is exported by Exportin 5 from the nucleus to the cytoplasm[85]. Here, the pre-miRNA is further processed by the RNase III- type enzyme Dicer, yielding an ~ 22 bp RNA duplex, the mature miRNA (Figure 3 a)[73; 75; 86]. Dicer exists in a complex with

TRBP and/or PACT, which contribute to the RNA induced silencing complex (RISC) assembly[87; 88]. Argonaute (AGO) proteins (1-4) are also part of RISC, forming the RISC loading complex[73; 89]. The mature miRNA is then loaded onto one of the AGO proteins and unwound. One of the strands remains bound to AGO (guide strand or miRNA) originating the effector complex and guiding RISC to its targets[73-75], whereas the other strand (passenger strand or miRNA*) is discarded and degraded (Figure 3a). Strand selection depends on strand thermodynamic stability, being usually the least stable at the 5' end preferentially incorporated into RISC[90]. The effector complex then acts upon its target-mRNAs, as was described above[74; 89].

Due to their small target sequence, miRNAs are able to affect multiple mRNAs and consequently coordinate various cellular processes. It is currently thought that miRNAs work mostly as fine-tuning and dampening elements of cellular pathways[91]. Hence, it is not surprising that miRNAs have been extensively implicated in human pathologies such as cancer, cardiac and neurological diseases[92-94].

3. Preliminary data on the FET proteins

Since FET proteins are able to bind DNA, it is likely that they bind specific genomic regions and target specific genes. In order to clarify this issue, our laboratory performed chromatin immunoprecipitation followed by Next-generation sequencing using antibodies for EWS, FUS and TAF15. HEK293 cells were lysed, the chromatin was fragmented by sonication, the above-mentioned proteins were precipitated and the resulting DNA fragments were sequenced and aligned to the human reference genome (hg19). EWS and FUS, but not TAF15, were found enriched in numerous cellular loci, most of them being in close proximity to (within 10 Kb) or inside genes (data not shown). EWS and FUS proteins appeared to bind more frequently together and at the 3' end region of genes (data not shown, Figure 4). Interestingly, among the enriched peaks were four loci which encoded for miRNAs, namely miR-636, miR-150, miR-17-92 cluster and pre-miR-7-1 (data not shown, Figure 4).

The fact that EWS and FUS were found binding to the DNA in the vicinity of miRNA-encoding loci indicates that both proteins may have a role in miRNA transcription and/or processing. Based on the obtained data, we chose to start our studies with the heterogeneous nuclear ribonucleoprotein K locus since EWS and FUS presented a remarkably clear enrichment peak at the 3' end of this gene (Figure 4). Furthermore, *HNRNPK* region contains unique features. It is located on chromosome 9 and originates three mRNA variants (Supplementary Figure 1). The canonical mRNA variant (variant 1) is the largest and encodes for hnRNPK protein isoform a. The second mRNA variant (variant 2) has an alternative 5' non-coding exon, thus having a different 5' UTR, and produces the same protein isoform as variant 1.

The third variant (variant 3) has the same 5' UTR as variant 1 but has a different 3' UTR that derives from an alternative splice acceptor site at the last exon. This causes a frame-shift which originates a hnRNPK protein with a different C-terminal, the isoform b. hnRNPK protein is well studied and has been implicated in numerous cellular processes such as chromatin remodeling, transcription, splicing, translation and cell proliferation. It is also able to bind RNA and is currently thought that it functions as a docking platform to integrate these cellular mechanisms[95].

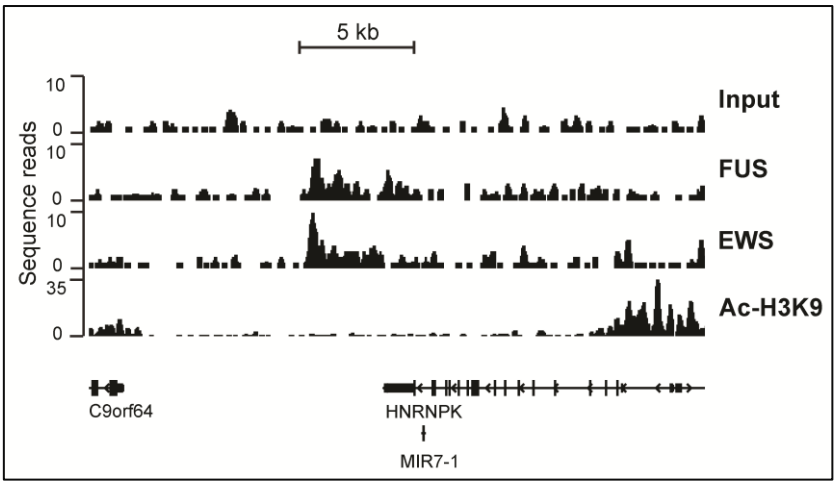


Figure 4 | FUS and EWS bind at the 3' end of *HNRNPK*. Graphic distribution of enriched peaks obtained from chromatin immunoprecipitation of HEK293 cells with antibodies for FUS and EWS. The immunoprecipitated DNA fragments were sequenced and aligned to the human genome (hg19). The graphic shows the number of sequence reads on the left side. FUS and EWS have enriched peaks at the 3' end (within 10 Kb) of *HNRNPK*. Ac-H3K9 was used as a positive control for actively transcribed genes. Input DNA in equivalent amount was used as negative control. Note the different scale of sequence reads for Ac-H3K9. At the bottom are represented the corresponding encoding loci (UCSC hg19 genome database). Note that *HNRNPK* is transcribed from the negative DNA strand and is thus depicted in reverse orientation. *Adapted from the manuscript Blechingberg et al.*

Finally, the last intron of *HNRNPK* encodes for pre-miR-7-1, one of the precursors of miR-7. This miRNA is also produced from two other pre-miRNAs: pre-miR-7-2, transcribed from an intergenic region in chromosome 15 and pre-miR-7-3, transcribed from MIR7-3 host gene, in chromosome 19. Moreover, miR-7 is reasonably well studied and has been implicated in developmental processes and disease. It appears to play a role in eye development in *Drosophila melanogaster* and vertebrates[96; 97] and is highly expressed in hypothalamus and pancreatic islets[98; 99]. Interestingly, miR-7 has been found deregulated in brain, lung and breast tumors[100-105]. Most reports have shown that miR-7 is down-regulated in human cancers having a tumor suppressor function. Correspondingly, its overexpression results in decreased cell invasiveness and viability coupled with increased apoptosis[102; 103]. Furthermore, it targets numerous genes involved in oncogenesis such as Epidermal growth factor receptor (EGFR), several components of the AKT pathway[102; 104] and p21-activated kinase 1 (PAK1)[103]. Moreover, miR-7 was associated with inhibition of epithelial-to-mesenchymal transition in metastasis of breast cancer cells[106]. Oppositely, miR-7 was found up-regulated in non-small cell lung cancer[101] which reflects the complexity of miRNA mechanisms and differences in cellular environment.

Aim of this study

It is known that transcription and splicing are greatly intertwined and thus influence each other[107]. Recently, it was further discovered that pri-miRNAs are processed co-transcriptionally and that Drosha cleavage can occur before splicing[81; 82]. An increasing amount of evidence points towards a close relationship between transcription, splicing and miRNA processing. Hence, it is not surprising that these processes might be simultaneously regulated and coordinated by various proteins. Given the FET proteins ability to bind both DNA and RNA and their role in transcription, splicing and presence in the Microprocessor Complex it is not unreasonable to think that FET proteins may provide a link between these cellular processes.

The main goal of the present study was to investigate the role of EWS and FUS in miRNA processing. This study focused on the *HNRNPK* locus and pre-miR-7-1 production in order to (a) describe the cellular role of EWS and FUS in miRNA processing and (b) clarify the influence of other molecular mechanisms in miRNA processing possibly correlated to EWS and FUS function, such as transcription and splicing.

Materials and Methods

Cell Culture

HEK293 cells were grown in Dulbeccoo's modified Eagle's medium (DMEM) (Cambrex, Verviers, Belgium) supplemented with 5% fetal bovine serum, streptomycin (2,0 g/L) and penicillin (1,2 g/L), at 37 ºC and 5 % CO_2.

Transient siRNA Double-Transfection

siRNA-mediated depletion of EWS and FUS proteins was accomplished by using different specific siRNAs (Eurofins DNA, Ebersberg, Germany) (Supplementary Table 1). In all treatments, cells were subjected twice to siRNA-mediated transfection, designated double-transfection, with siRNAs for EWS (KOEWS) or FUS (KOFUS) (Table 1). Furthermore, EWS and FUS were simultaneously depleted by mixing equal amounts of siRNA for EWS and FUS (KOEWS+FUS) (Table 1). A non-specific siRNA, which does not target any cellular mRNAs, was used as control for sample comparison (KOControl). Two double-transfection rounds with the four mentioned treatments were conducted with different siRNA mixes for EWS as well as EWS+FUS. In Transfection 1, two different siRNAs were mixed for the knockdown of each protein. In Transfection 2, a different siRNA was used for EWS depletion whereas the same mix of siRNAs as in Transfection 1 was used for FUS (Table 1).

Table 1 | siRNA mixes used in the different double-transfection experiments and corresponding amount of each added siRNA in a total of 100 nM of siRNA per well.

Transfection	KO targets	siRNAs
Transfection 1	KOEWS1	siEWSa (50 nM) + siEWSb (50 nM)
	KOFUS1	siFUSa (50 nM) + siFUSb (50 nM)
	KOEWS1+FUS1	siEWSa (25 nM) + siEWSb (25 nM) + siFUSa (25 nM) + siFUSb (25 nM)
	KOControl	siCONTROL (100 nM)
Transfection 2	KOEWS2	siEWSc (100 nM)
	KOFUS2	siFUSa (50 nM) + siFUSb (50 nM)
	KOEWS2+FUS2	siEWSc (50 nM) + siFUSa (25 nM) + siFUSb (25 nM)
	KOControl	siCONTROL (100 nM)

For double-transfection, HEK293 cells were subjected to two siRNA-mediated transfections, separated by 48h for cell adjustment. Firstly, HEK293 cells were grown until ~ 50 - 80% confluency, trypsinized and counted with NucleoCounter (Chemotec, Montreal, Canada). 2,5 x 10^5 cells were seeded in 6-well plates and transfected for the first time with 100 nM of total siRNA plus 4 µL of DharmaFECT®1 (Dharmacon, Lafayette, Colorado) in DMEM. Cells were grown for 48 hours. The same cells were subjected to a second transfection with the same amount of total siRNA plus 4 µL of *Trans*IT-siQUEST® Transfection Reagent (Mirus,

Madison, Wisconsin) in DMEM without antibiotics. Cells were harvested after 48 hours and used for RNA or protein extraction. KOEWS1, KOFUS1 and KOEWS+FUS1 mixes were done in two individual experiments. KOEWS2, KOFUS2 and KOEWS+FUS2 were done in biological triplicates.

RNA extraction and Quantitative Real-Time PCR

Total RNA was extracted with TRI®-reagent (Sigma-Aldrich, St. Louis, Missouri) according to manufacturer's protocol. Total RNA was stored at -70 ºC until further use.

Complementary DNA (cDNA) synthesis was done using iScript™ cDNA Synthesis Kit (Bio-Rad, Hercules, California) with random hexamer primers according to manufacturer's protocol. For hnRNPK mRNA quantification, whose primers could amplify genomic DNA, total RNA samples were first DNase-treated with TURBO DNA-*free*™ kit (Ambion, Austin, Texas). This was followed by cDNA synthesis as previously mentioned. EWS, FUS, TAF15, BCL2, LPHN2, IGFR1, EGFR, RAF1, hnRNPK, UBC, GAPDH and TBP mRNA expression was evaluated by quantitative real-time PCR (qPCR) with LightCycler 480 SYBR Green I Master mix (Roche, Basel, Switzerland) using a LightCycler 480 (Roche, Basel, Switzerland). Quantification was performed by 15 s at 95 ºC heating, 50 cycles of 10 s at 95 ºC plus 20 s at 58 ºC plus 15 s at 72 ºC and a final elongation of 1 min at 72 ºC. EWS, FUS, TAF15, BCL2, LPHN2, IGFR1, EGFR, RAF1 and hnRNPK transcripts were normalized to TBP reference gene expression. All above-mentioned primer sequences are described in Supplementary Tables 2 and 3.

For precursor miRNA quantification, cDNA synthesis was done using Thermoscript™ RT-PCR System (Invitrogen, Carlsbad, California) kit with gene specific reverse primers for pri-miR-7-1, pre-miR-7-1 and RNU48. For pri-miR-7-1 the same reverse primer was used for cDNA synthesis and qPCR analysis. Oppositely, for pre-miR-7-1 and RNU48 quantification, specific reverse primers closer to the 3' end of the transcript were used in order to obtain a more precise and better amplification, as these molecules have a small size and are difficult to amplify. All gene specific reverse primers were mixed prior to cDNA synthesis. A pre-incubation of the total RNA with the primer mix and dNTPs was done for 5 min at 80 ºC plus 5 min at 60 ºC to denature the small RNAs and facilitate specific reverse primer annealing[108]. cDNA synthesis was performed for 45 min at 60ºC and the reverse transcriptase was inactivated by heating the reaction for 5 min at 85 ºC. pri-miR-7-1, pre-miR-7-1 and RNU48 expression was quantified by qPCR with LightCycler 480 SYBR Green I Master (Roche, Basel, Switzerland) mix using a LightCycler 480 (Roche, Basel, Switzerland). Quantification was performed by 10 s at 95 ºC heating, 50 cycles of 15 s at 95 ºC plus 1 min at 60 ºC. pri-miR-7-1 and pre-miR-7-1 transcripts were normalized to small nucleolar RNA C/D box 48

(RNU48) reference gene expression. All above-mentioned primer sequences are described in Supplementary Table 2.

Primer specificity and quality for all above mentioned primers used in qPCR was evaluated by melting curves. These were obtained by a denaturation protocol after the qPCR amplification. Final qPCR products were run on a 1,5 % agarose TAE gel in order to verify if there was a single amplification product and if the amplification product had the correct size.

Finally, for mature miR-7 analysis, cDNA synthesis was performed using TaqMan® MicroRNA Reverse Transcription Kit, according to manufacturer's protocol. Quantification of miR-7 was done by TaqMan® microRNA Assays (Applied Biosystems, Foster City, California) according to manufacturer's protocol. miR-7 and RNU48 expression was determined using miR-7 and RNU48 primers and probes (Assay ID 000268 and 001006, respectively). qPCR was performed with TaqMan® Universal Master Mix II, No UNG (Applied Biosystems, Foster City, California) using a LightCycler 480 (Roche, Basel, Switzerland). miR-7 expression was normalized to RNU48 reference gene.

For all qPCR assays, relative mRNA levels and amplification efficiency were determined using the relative standard curve method. This was accomplished by obtaining a standard curve for each gene from a dilution series of cDNA. The standard curve was then used to calculate the relative amount of specific RNA present in each sample. Relative gene expression levels of each gene were quantified using the X_0-method[108]. All qPCR assays were performed at least twice and data presented as mean + SEM (standard error of the mean).

Western Blot

Protein was extracted from siRNA-depleted HEK293 cells using the 1x Lane Marker Reducing Sample Buffer (Thermo Scientific, Waltham, Massachusetts) and heated for 5 min at 95 ºC. Samples were cooled on ice and loaded onto a Mini- Protean® TGX™ Gel (Bio-Rad, Hercules, California) along with the ProSieve® Color protein marker 9-75 kDa (Lonza, Basel, Switzerland). Gel was run for ~ 40 min at 40 mA in Running Buffer (0.25 M Tris, 1.92 M Glycine, 1 % SDS). Proteins were then transferred to an Amersham Hybond™-P membranes (GE Healthcare, Waukesha, Wisconsin) previously activated for 20 s in ethanol 99.9 %. Protein transfer was done in cold Transfer Buffer (0.25 M Tris, 1.92 M Glycine, pH 8.3) for 30 min at 60 V and 4 ºC. Membrane were blocked in 10% skimmed milk powder (Difco, Franklin Lakes, New Jersey) mixed with sodium perborate (PBS) and 1% Tween® 20 (Sigma-Aldrich, St. Louis, Missouri) overnight, at 4 ºC, in Blocking Buffer (PBS with 0.04 % skimmed milk powder and 0.1 % Tween 20). Membranes were then incubated for 1.5 h at 4 ºC with the primary antibodies for EWS (sc-48404, Santa Cruz Biotechnology, Santa Cruz, California),

FUS (sc-47711, Santa Cruz Biotechnology, Santa Cruz, California), TAF15 (8TA2B10, provided by Laszlo Tora, Strasbourg, France) and 4F4-hnRNP C1+C2 (ab10294, Abcam®, Cambridge, United Kingdom). EWS, FUS and TAF15 primary antibodies were diluted 1:2000 in Washing Buffer (PBS with 0.5 % skimmed milk powder and 0.1 % Tween 20) whereas 4F4 was used as a loading control with 1: 10000 dilution. Membranes were washed three times with Washing Buffer and incubated with the secondary polyclonal goat anti-mouse (P 0447, DakoCytomation, Glostrup, Denmark) antibody with 1:10000 dilution in Washing Buffer for 1 h, at room temperature. After incubation, membranes were washed with Washing Buffer five times. Antibody detection was done using BM Chemoluminescence Blotting substrate (Roche, Basel, Switzerland), exposed to A plus film (Konica Minolta, Ramsey, New Jersey) and then developed with AGFA Curix 60.

Alternative Splicing Assay

The same cDNA used for hnRNPK mRNA quantification was used for this assay. Briefly, EWS, FUS, EWS+FUS and control siRNA-depleted HEK293 cells were tested for variations of the alternative splicing pattern at the last exon of hnRNPK mRNA. For this, a special primer set was designed (E15F_E16R-S, Supplementary Table 3 and Supplementary Figure 2) which allowed the amplification of two different bands: the largest (~ 491 bp) corresponded to the proximal splice acceptor (PSA) site usage and the smallest (~ 431 bp) corresponded to distal splice acceptor (DSA) site usage. PCR amplification was performed with a 2720 Thermal Cycler (Applied Biosystems). Amplification was done between 28 and 30 cycles to ensure that the PCR reaction was within the linear amplification range and excessive product accumulation did not occur. PCR products were run on a 2% agarose TAE gel with a 100 bp DNA ladder (Bio Labs, Ipswich, Massachusetts) and relative band intensity was quantified by *GelQuant.NET software* provided by http://biochemlabsolutions.com. Ratio between DSA and PSA was determined and plotted onto a graph. Each experiment was done at least in duplicates and results presented as mean + SEM.

Statistical Analysis

P-values were calculated using unpaired Student's T-test. In all statistical analysis, *p-values* bellow 0.05 were considered statistically significant.

Results

Effect of EWS and FUS depletion (Transfection 1)

With the purpose of evaluating the cellular role of EWS and FUS, transient siRNA-mediated depletion of both proteins was performed. HEK293 cells were subjected to three different siRNA treatments using a mix of two siRNAs for each protein: (a) KOEWS1, with siEWSa and siEWSb, (b) KOFUS1, with siFUSa and siFUSb and (c) KOEWS1+FUS1 simultaneously, with a mix of the four mentioned siRNAs (Table 1 and Supplementary Table 1). HEK293 cells were also transfected with a control siRNA (KOControl), without cellular targets, for data comparison (Table 1 and Supplementary Table 1). Cells were subjected to a double-transfection to obtain an efficient knockdown both proteins. All samples used for FUS and EWS mRNA quantification were normalized to TBP reference gene since it was the most stably expressed among the different siRNA-mediated knockdowns (Supplementary Figure 3). EWS, FUS and EWS together with FUS were successfully depleted at the mRNA level in all specific treatments (81 %, 82 % and 75 % and 78%, respectively), when compared to control. Knockdowns were equally efficient at the protein level (Figure 5 a).

After obtaining successful siRNA-mediated depletion of EWS and FUS mRNAs and proteins, the experiments were focused on the *HNRNPK* locus. In order to investigate if EWS and

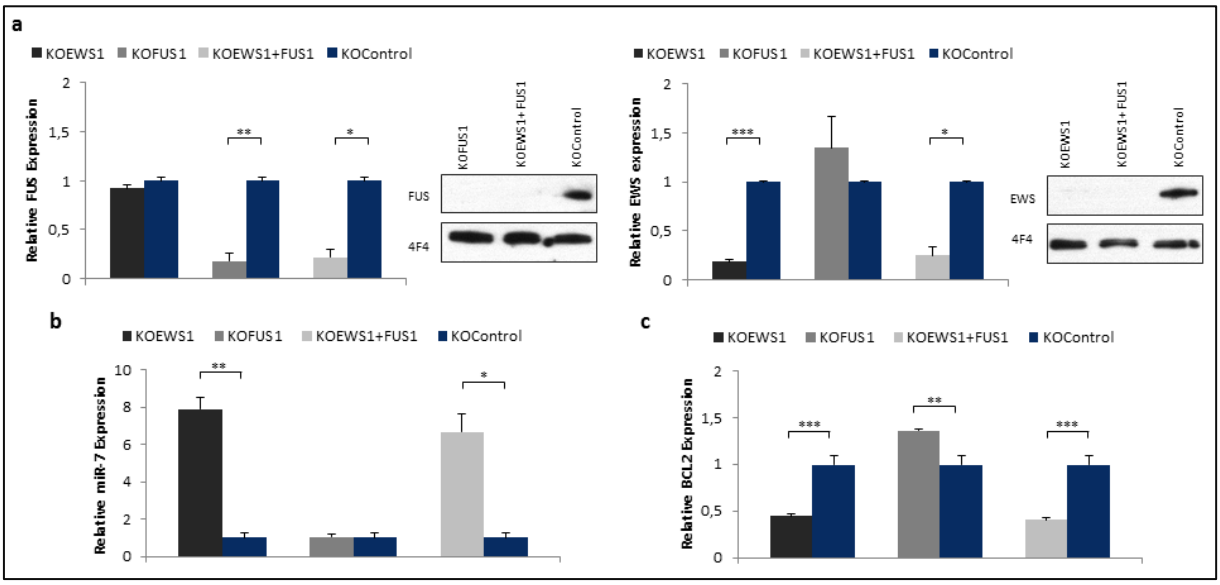


Figure 5 | Transient siRNA-mediated depletion of EWS and FUS and miR-7 and BCL2 mRNA quantification (Transfection 1). HEK293 cells were double-transfected with specific siRNAs for EWS (KOEWS1), FUS (KOFUS1) and EWS together with FUS (KOEWS+FUS1) as well as with a control siRNA (KOControl) with no mRNA cellular targets. (a) Relative mRNA quantification of FUS and EWS by qPCR in siRNA-depleted cells. FUS and EWS were normalized to reference gene TBP. On the right of the graphs is protein quantification of FUS and EWS by western blot of the corresponding siRNA-depleted cells. 4F4-HNRNP C1+C2 (4F4) was used as a loading control. (b) KOEWS1, KOFUS1, KOEWS+FUS1 and KOControl siRNA-depleted cells were used for relative quantification of miR-7 and BCL2 mRNA by qPCR. Obtained data was normalized to reference genes RNU48 and TBP, respectively. *P-values* for the comparisons are indicated by brackets as follows: *P < 0.05; **P < 0.01, ***P < 0.001. Two individual experiments with the same conditions were performed. Data presented as mean + SEM.

FUS proteins had a role in miR-7 production, quantification of mature miR-7 in siRNA-depleted cells was performed. In KOEWS1 and KOEWS1+FUS1 knockdown treatments, miR-7 was significantly over-expressed (~ 8 and 7 fold, respectively) when compared to control cells (Figure 5 b). On the contrary, KOFUS1 did not show any variation in miR-7 expression levels. To verify if the up-regulation of miR-7 would have any consequences on its cellular targets the literature was investigated in order to find previously validated miR-7 target mRNAs. Several mRNAs were found to be targeted by miR-7 and five were chosen for further tests: RAF1, IGFR1, EGFR, LPHN2 and BCL2. From these, only BCL2 mRNA proved to be specifically affected by miR-7 up-regulation, since it was down-regulated in the siRNA-depleted cells where miR-7 was correspondingly overexpressed (Figure 5 c). In KOEWS1 and KOEWS1+FUS1 treatments, BCL2 mRNA expression was reduced approximately 60 %, when compared to control cells. In KOFUS1 treatment was observed an up-regulation of 40 % in BCL2 mRNA levels. Oppositely, the other target mRNAs were not affected by miR-7 over-expression (Supplementary Figure 4). Based on the obtained data, it appears that the knockdown of EWS as well as both EWS and FUS caused an overexpression of miR-7 which consequently down-regulated its target mRNA BCL2.

Considering the roles of EWS and FUS in transcription and splicing it is possible that the knockdown of EWS and FUS influences these processes. To investigate this, several primer combinations were designed along hnRNPK mRNA. Three kinds of primer sets spanning different mRNA regions were used: (a) exon-exon spanning primers for spliced mRNA quantification, (b) exon-intron spanning primers for unspliced mRNA quantification and (c) intron-intron spanning primers for unspliced mRNA and intron quantification (Figure 6 a). hnRNPK mRNA quantification showed an accumulation of unspliced mRNA towards the end of the hnRNPK mRNA. In KOEWS1 there was an increase of the unspliced mRNA in the region of exons 14 and 15 (30 % and 32 % in I13F_E14R and I14F_E15, respectively) (Figure 6 b). In KOFUS1 the same was observed, with unspliced hnRNPK mRNA accumulated in the region spanning exons 14 and 15 (51%, 46% and 42 % in I13F_E14R, I14F_E15 and E15F_I15R, respectively) (Figure 6 c). Finally, in KOEWS1+FUS1 the unspliced hnRNPK mRNA was again increased in the region of exons 14 and 15 (26 % and 73 % in I13F_E14R and E15F_I15R, respectively) (Figure 6 d). Oppositely, no significant changes in all siRNA-depleted cells were observed in the fully spliced mRNA (E9F_E10R and E15F_E16R) or unspliced mRNA in the beginning and middle of the hnRNPK mRNA (E3F_I3R, E10F_I10R). Furthermore, no significant changes were detected at the intron levels (I10F_I10R, I15F_I15R). Obtained results indicated that depletion of EWS, FUS or both caused an accumulation of unspliced transcripts between exons 14 and 15, and thus affected splicing at the 3’ end of hnRNPK mRNA.

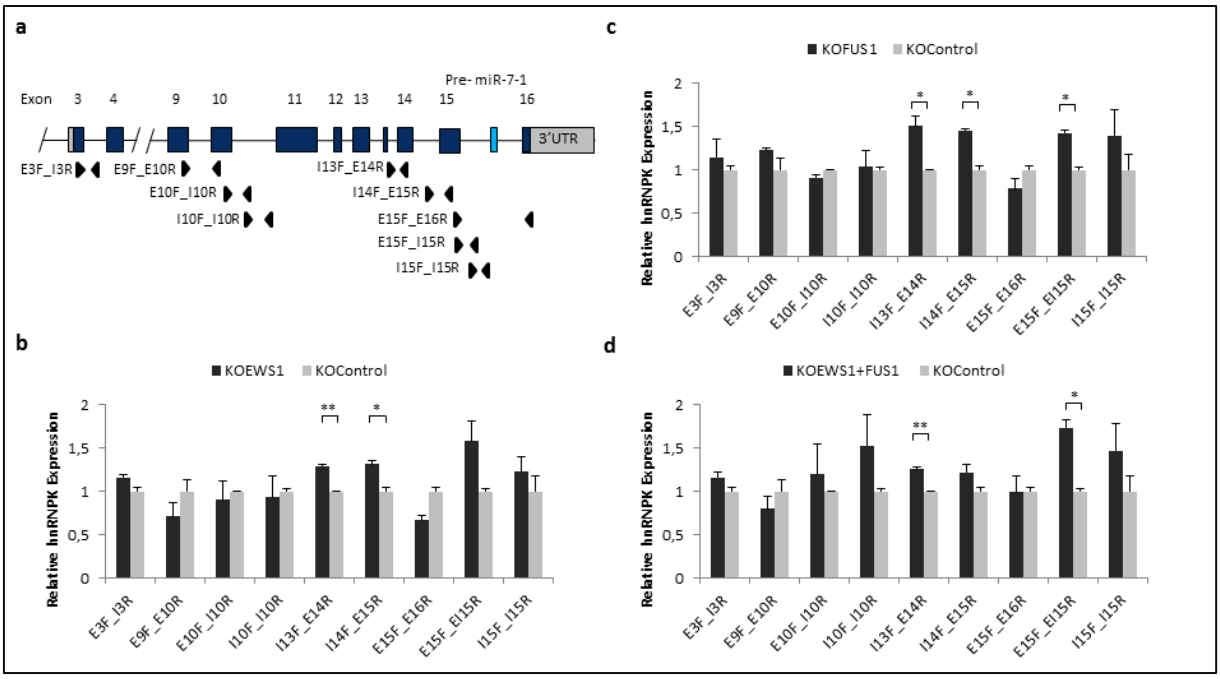


Figure 6 | Quantification of spliced and unspliced hnRNPK mRNA (Transfection 1). (a) Schematic representation of parts of *HNRNPK* locus with the location of the primer sets used for mRNA quantification. The dark blue boxes represent the exons, identified with the corresponding number above, and grey boxes represent 5' and 3'UTR sequences. The light blue box indicates pre-miR-7-1 encoding region. The paired arrows represent the forward and reverse primers used to amplify the corresponding mRNA region. (b, c, d) KOEWS1, KOFUS1, KOEWS1+FUS1 and KOControl siRNA-depleted HEK293 cells were used for relative hnRNPK mRNA quantification by qPCR using all primer combinations indicated in (a). All samples were normalized to reference gene TBP. *P-values* for the comparisons are indicated by brackets as follows: *P < 0.05; **P < 0.01. Two individual experiments with the same conditions were performed. Data presented as mean + SEM.

Off-targeting effects on TAF15 mRNA

An unexpected effect was detected in KOEWS1 and KOEWS1+FUS1 siRNA-depleted cells: in both treatments cells divided less and presented a round shape when compared to the elongated shape of KOFUS1 and KOControl treatments (Supplementary Figure 5). Since this effect had not been mentioned in any of the previous literature, this issue was investigated further. The siRNAs used in the siRNA-mediated knockdowns (siEWSa, siEWSb, siFUSa and siFUSb) were blasted against the human RefSeq mRNA database using the NCBI Nucleotide BLAST. Results showed numerous possible target mRNAs for all siRNAs. Among siEWSb targets was TAF15, with a base-pair complementarity of 13 out of 19 nucleotides (Figure 7 a). Furthermore, since TAF15 is homologous to EWS and FUS and is thought to have similar cellular functions, it became relevant to test if TAF15 was affected by siRNA-mediated depletion of EWS. Hence, quantification of TAF15 expression in all siRNA-depleted samples was done. Results showed that TAF15 mRNA was down-regulated 87 % and 54 % in KOEWS1 and KOEWS1+FUS1 treatments, respectively, when compared to control (Figure 7 b). Oppositely, in KOFUS1 treatment, TAF15 mRNA was more than 2 fold overexpressed thus showing that the siRNAs used in KOFUS1 treatment were not targeting TAF15 mRNA. These results indicated that either siEWSb was targeting TAF15 mRNA or that both siEWSa and siEWSb were having unknown indirect effects in the cell that caused

the down-regulation of TAF15 mRNA. To test this hypothesis, a test-run experiment was conducted where HEK293 cells were subjected to a transfection with siEWSa and siEWSb separately, to analyze which would be the effects of the individual siRNAs, when compared to control. Cells were also transfected with two siRNAs specific for TAF15 (KOTAF15) in order to compare the phenotype obtained in EWS-depleted cells with TAF15-depleted cells. Regarding cell morphology, both siEWSa and siEWSb treated cells displayed a round shape whereas TAF15-depleted cells had no significant morphological changes when compared to control (data not shown). On the contrary to expected, TAF15 mRNA was 53 % down-regulated in siEWSa treatment, while it was 62 % up-regulated in siEWSb treatment, when compared to control (Supplementary Figure 6). TAF15 mRNA was down-regulated 81% in TAF15 specific knockdown (KOTAF15) and EWS mRNA was down-regulated in both siEWSa and siEWSb treatments (42% and 71%, respectively). Finally, EWS mRNA was also slightly down-regulated (28%) in KOTAF15 treated cells. These results indicated that siEWSb was not targeting TAF15 mRNA. Instead, it showed that it was siEWSa affecting TAF15 mRNA expression. Regardless, TAF15 mRNA was affected in KOEWS1 and KOEWS1+FUS1 siRNA-mediated knockdowns, which could lead to difficulties in data interpretation. Also, both siEWSa- and siEWSb- depleted cells presented altered morphology which could point to cell toxicity derived from the siRNA used and thus could significantly alter obtained data. Therefore, we decided to use a new siRNA for EWS mRNA depletion, siEWSc (Supplementary Table 1). Henceforth, all siRNA-mediated knockdowns of EWS were done using siEWSc siRNA.

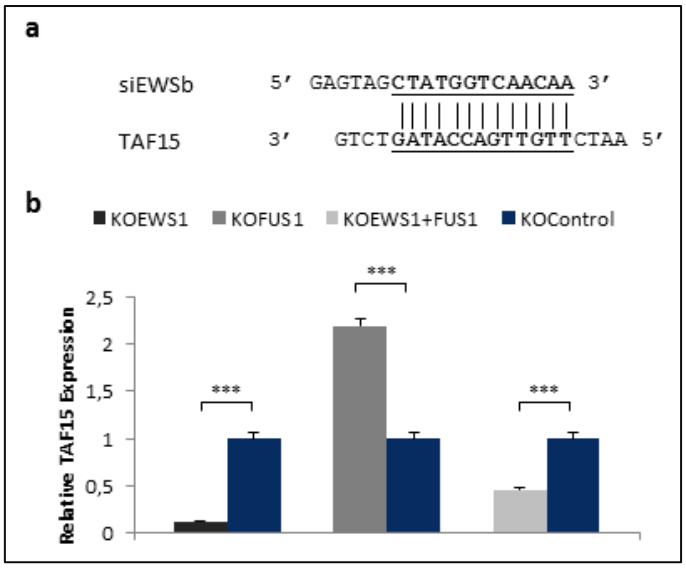


Figure 7 | TAF15 mRNA is down-regulated in KOEWS1 and KOEWS1+FUS1 treatments. (a) Schematic representation of siEWSb and TAF15 mRNA base-pair complementarity (13 out of 19 bases). The figure was based on NCBI Nucleotide BLAST of siEWSb against RefSeq mRNA database. (b) TAF15 mRNA expression was quantified by qPCR in KOEWS1, KOFUS1, KOEWS1+FUS1 and KOControl siRNA-depleted HEK293 cells. All samples were normalized to reference gene TBP. *P-values* for the comparisons are indicated by brackets as follows: ***P < 0.001. Two individual experiments with the same conditions were performed. Data presented as mean + SEM.

EWS and FUS role in splicing and miRNA processing (Transfection 2)

Given the above-mentioned hurdles with siRNA-mediated depletion of EWS, new double-transfection experiments were conducted. HEK293 cells were subjected to siRNA-mediated double-transfection with the same siRNA mix for FUS, with siFUSa and siFUSb (KOFUS2)

and a single siRNA for EWS, siEWSc (KOEWS2 and KOEWS2+FUS2), as shown in Table 1. The same control siRNA was used as in Transfection 1 experiments.

EWS, FUS and EWS together with FUS were effectively depleted at the mRNA level in all specific siRNA treatments, with knockdown efficiencies of 86 %, 81 % and 86% and 73 %, respectively, when compared to control (Figure 8 a, top panel). The knockdown treatments were as efficient at the protein level, as is depicted in Figure 8 a, bottom panel. TAF15 mRNA and protein levels were overexpressed in all siRNA-depleted cells (Figure 8 a), which showed that the siRNA used for EWS depletion was not targeting TAF15 mRNA. FUS mRNA and protein levels were also slightly up-regulated in KOEWS2-depleted cells (Figure 8 a). Remarkably, no morphological changes were detected in any of the siRNA-depleted cells when compared to control (data not shown), possibly indicating less cellular toxicity.

Having accomplished successful depletion of EWS and FUS with minimal nonspecific

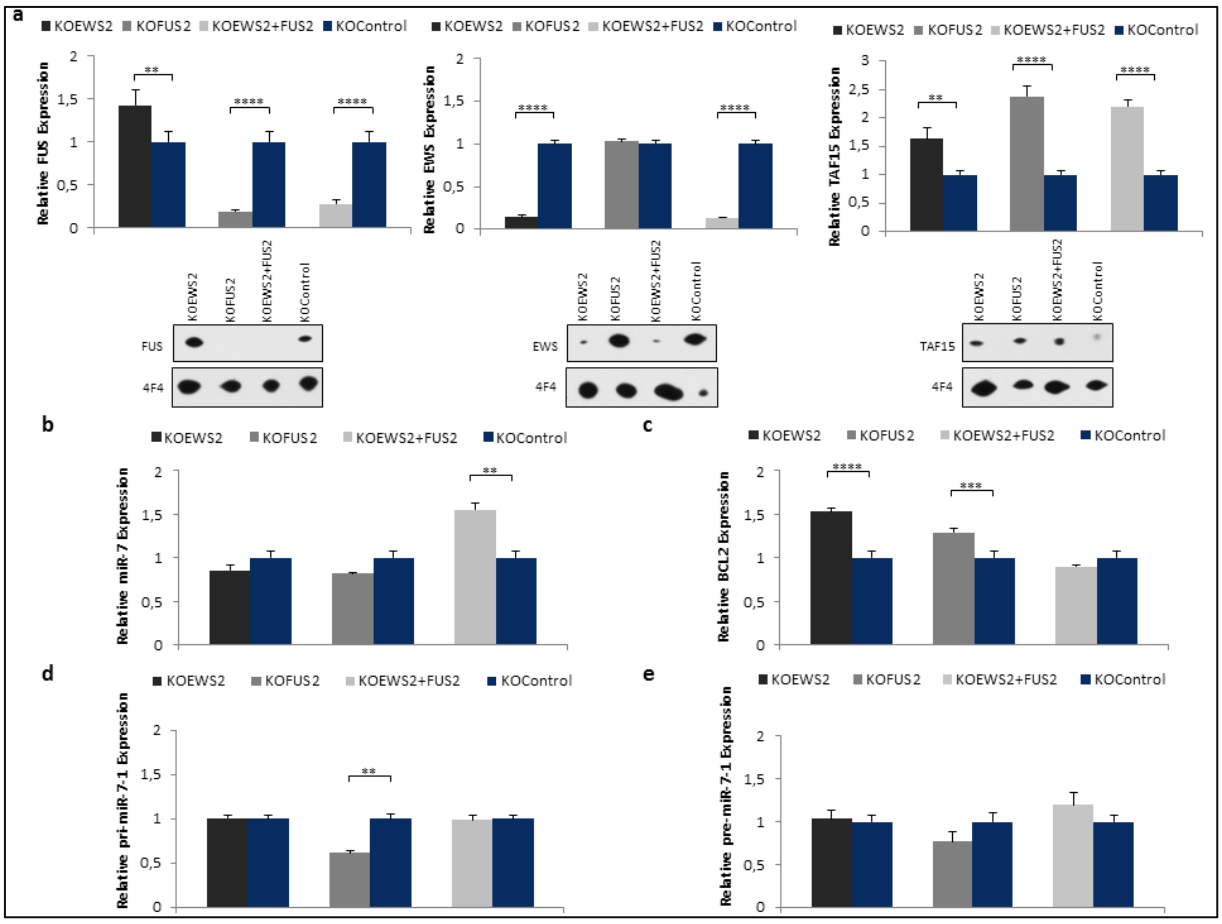


Figure 8 | Transient siRNA-mediated depletion of EWS and FUS and miR-7, BCL2 mRNA and miR-7 precursors quantification (Transfection 2). HEK293 cells were double-transfected with specific siRNAs for EWS (KOEWS2), FUS (KOFUS2) and EWS together with FUS (KOEWS2+FUS2) as well as with a control siRNA (KOControl). (a) In the top panel is the relative mRNA quantification of FUS, EWS and TAF15 by qPCR of siRNA-depleted cells. FUS, EWS and TAF15 were normalized to reference gene TBP. In the bottom panel is protein quantification of FUS, EWS and TAF15 by Western blot of the corresponding siRNA-depleted cells. 4F4- HNRNP C1+C2 (4F4) was used as a loading control. (b,c) The above-mentioned siRNA-depleted HEK293 cells were used for relative quantification of miR-7 and BCL2 mRNA by qPCR. Obtained values were normalized to reference gene RNU48 and TBP, respectively. (d, e) The above-mentioned siRNA-treated cells were used for relative quantification of pri-miR-7-1 and pre-miR-7-1 by qPCR. pri-miR-7-1 and pre-miR-7-1 values were normalized to reference gene RNU48. *P-values* for the comparisons are indicated by brackets as follows: **P < 0.01, ***P < 0.001, ****P < 0.0001. Experiments were performed in biological triplicates. Data presented as mean + SEM

effects, the previous assays were repeated. Firstly, quantification of miR-7 in all siRNA-depleted cells was done. miR-7 quantification results were quite different to expected, as it was only up-regulated about 55% in the double-knockdown treatment (KOEWS2+FUS2) when compared to control (Figure 8 b). Depletion of EWS and FUS alone (KOEWS2 and KOFUS2) did not cause any significant variation of miR-7 levels. Furthermore, we quantified the mRNA levels of BCL2 in the siRNA-depleted cells. There was no down-regulation of BCL2 mRNA levels in any of the specific knockdown treatments (Figure 8 c). In fact, BCL2 mRNA was overexpressed in KOEWS2 and KOFUS2 treatments (53% and 29%, respectively).

Given the fact that the results obtained for miR-7 were not according to expected, we decided to quantify the precursors for miR-7 produced from *HNRNPK* locus. Hence, primers were designed which specifically amplified pri-miR-7-1 and pre-miR-7-1. Precursor miRNAs cannot be normalized to a regular mRNA transcript, as their expression has to be compared to the stable expression of a cellular small nuclear RNA. Therefore, primers for a reference small nuclear RNA were also designed. To facilitate data comparison, the same reference gene was used as in miR-7 quantification, RNU48. Also, since precursor miRNAs were to be quantified by SYBR Green dye and mature miRNA by TaqMan Assays, a test run was conducted to verify if both probes behaved equally. As is shown in Supplementary Figure 7, despite the difference in the detected cycle threshold values for SYBR green dye and TaqMan assays (~ 15 and 25, respectively), both probes varied similarly. Then, quantification of miR-7 precursors, pri-miR-7-1 and pre-miR-7-1, in siRNA-depleted HEK293 cells was done. There was no variation of the pri-miR-7-1 in KOEWS2+FUS2 as well as in KOEWS2 siRNA-mediated depletions (Figure 8 d). Instead, pri-miR-7-1 was 23 % down-regulated in KOFUS1 depleted cells, when compared to control. Moreover, no significant variation of pre-miR-7-1 levels was found in all siRNA specific knockdowns (Figure 8 e). These results indicate that FUS and EWS possibly play a role in miR-7 processing.

Subsequently, hnRNPK mRNA quantification was repeated, as different results could also be obtained regarding its transcription and processing. The same primer combinations were used as described above (Figure 9 a). In KOEWS2-depleted cells there was a decrease of 26 % in the unspliced mRNA at the 5’ end of hnRNPK (E3F_I3R) while the remaining mRNA levels were unchanged (Figure 9 b). In KOFUS2 treatment, there was a decrease in the unspliced mRNA (15 % in E3F_I3R) concomitant with an increase of the spliced mRNA (20 % in E9F_E10R), when compared to control (Figure 9 c). Oppositely, at the 3’ end of the mRNA there was significant accumulation of unspliced transcripts (31 % in I13F_E14R, 32% in I14F_E15R, 24 % E15F_I15R and 18% I15F_I15R). A similar effect was detected on KOEWS2+FUS2 depleted cells, with increased levels of unspliced mRNA at the 3’ end of

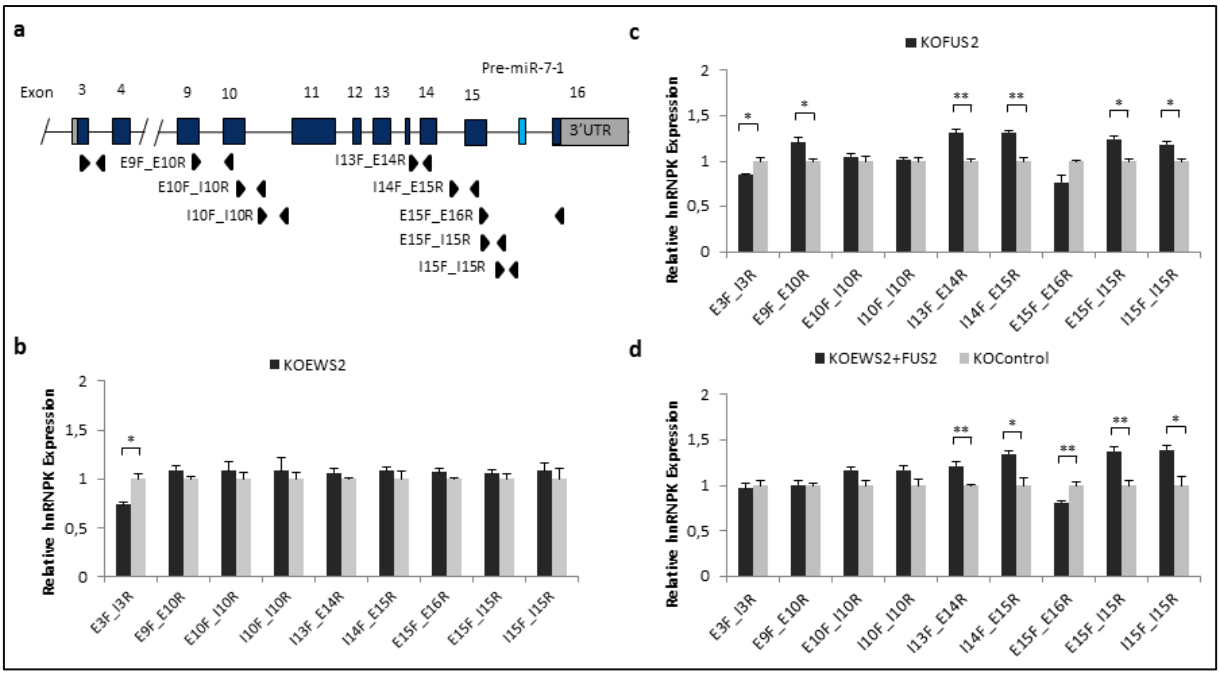


Figure 9 | Quantification of spliced and unspliced hnRNPK mRNA (Transfection 2). (a) Schematic representation of parts of *HNRNPK* locus with the location of the primer sets used for mRNA quantification. The dark blue boxes represent the exons, identified with the corresponding number above, and grey boxes represent 5' and 3' UTR sequences. The light blue box indicates pre-miR-7-1 encoding region. The paired arrows represent the forward and reverse primers used to amplify the corresponding mRNA region. (b, c, d) HEK293 cells treated with KOEWS2, KOFUS2, KOEWS2+FUS2 and KOControl were tested for relative hnRNPK mRNA quantification by qPCR with all primer combinations indicated in (a). All values were normalized to reference gene TBP. *P-values* for the comparisons are indicated by brackets as follows: *P < 0.05; **P < 0.01. All experiments were performed in biological triplicates. Data presented as mean + SEM

hnRNPK mRNA (21 % in I13F_E14R, 35% in I14F_E15R, 38 % in E15F_I15R and 40 % I15F_I15R) (Figure 9 d). Finally, we observed a concomitant decrease of the spliced mRNA levels at the 3' end of hnRNPK mRNA (19 % in E15F_E16R). Overall, similar results were obtained in FUS and EWS plus FUS siRNA-mediated knockdowns regarding hnRNPK mRNA processing in Transfections 1 and 2, indicating that FUS and EWS depletion affects splicing at the 3' end of hnRNPK mRNA.

Considering that hnRNPK mRNA has an alternative splicing site between the penultimate and last exons an assay was created to evaluate whether there were any changes in the alternative splicing pattern. It was not possible to design primers for qPCR quantification which would only amplify one of the isoforms, since the alternative splicing sequences were short and very similar to each other (Supplementary Figure 2). Therefore, a different reverse primer (E15F_E16R-S) was designed which allowed the amplification of the two isoforms by regular PCR. The forward primer would be placed in exon 15 (the same forward primer used in E15F_E16R) and the reverse primer would be placed downstream of both alternative splice acceptor sites, in exon 16 (Figure 10 a). PCR amplification would generate two amplicons with different sizes: a larger one, with ~ 491 bp, corresponding to the proximal splice acceptor (PSA) site and a smaller one, with ~ 431 bp, corresponding to the distal splice acceptor (DSA) site. It would be possible to compare both bands intensities in each sample and verify if the ratio of distal/proximal splice site usage would suffer any variation in the siRNA-depleted cells

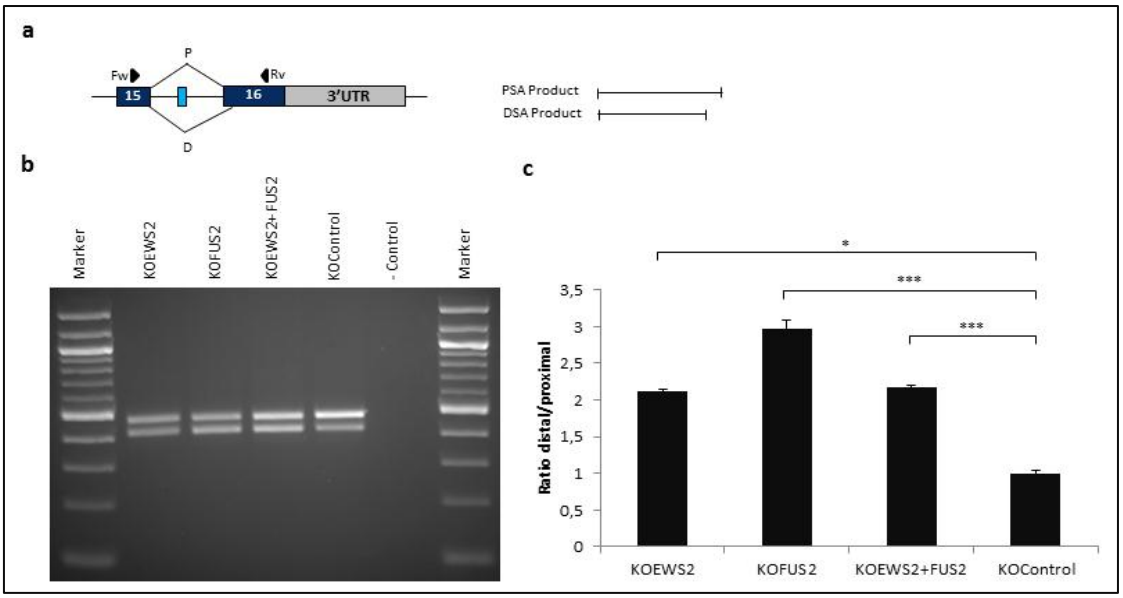


Figure 10 | Influence of EWS and FUS in alternative splicing of hnRNPK mRNA. (a) Schematic representation of the alternative splicing site at the 3' end of hnRNPK mRNA. The dark blue boxes represent the exons, identified with the corresponding number in white, and the grey box represents the 3'UTR. The light blue box indicates pre-miR-7-1 encoding region. P and D indicate the proximal acceptor splice (PSA) site and the distal acceptor splice (DSA) site, respectively. The arrows represent forward and reverse primers used to amplify the corresponding mRNA region (E15F_E16R-S). On the right are depicted the two amplicons generated by PCR amplification, being the largest the PSA product (~491 bp) and the smallest the DSA product (~431 bp). (b) cDNA from KOEWS2, KOFUS2, KOEWS2+FUS2 and KOControl depleted HEK293 cells was used for PCR amplification with E15F_E16R-S primer set. PCR products were run on a 2% agarose TAE gel and pictures were taken. The image is representative of 3 independent experiments. Marker: 100 bp DNA ladder. – Control: negative control of PCR mix without template cDNA. (c) Ratio of distal to proximal splice acceptor site usage in KOEWS2, KOFUS1, KOEWS+FUS2 and KOControl depleted cells. The ratio was calculated from relative band intensity values provided by *GelQuant.NET software*. The ratio for KOControl was assigned to the value 1 and the ratios for siRNA-depleted samples were calculated accordingly. *P-values* for the comparisons are indicated by brackets as follows: *P < 0.05; ***P < 0.001. Experiments were performed in biological triplicates. Data presented as mean + SEM.

when compared to control. The PCR reaction was carried on within the linear amplification range to avoid product saturation which would prevent the detection of variations in the amount of both bands. Apart from the negative control, it was possible to detect two bands in all treatments: the top one corresponded to the PSA site and the bottom one corresponded to the DSA site (Figure 10 b). KOcontrol cells showed that proximal splice site in hnRNPK mRNA was preferentially used. However, there was an increase in the DSA site usage in all siRNA-mediated knockdowns, when comparing to control. Additionally, the DSA/PSA ratio increased significantly, ranging from 2 to 3-fold, in all siRNA-depleted cells (Figure 10 c). This indicates that EWS and FUS depletion affected splice acceptor site choice in hnRNPK mRNA.

Finally, hnRNPK mRNA has two alternative 5' non-coding exons, hence originating mRNA variants with different 5' UTRs (Figure 11 a). Variants 1 and 3 both start at exon 1a, whereas variant 2 starts at exon 1b. hnRNPK mRNA quantification showed that variant 1 is the most expressed, followed by variant 3 and variant 2, which was very low expressed (Supplementary Figure 8). Next, it was tested whether the depletion of EWS and FUS would affect the production of either variant 1 and 3 or 2. For this, a primer set was designed where the forward primer annealed either in exon 1a (Var 1+3) or 1b (Var 2) and the reverse primer annealed in exon 2. Results showed that expression of variants 1 and 3 did not suffer any

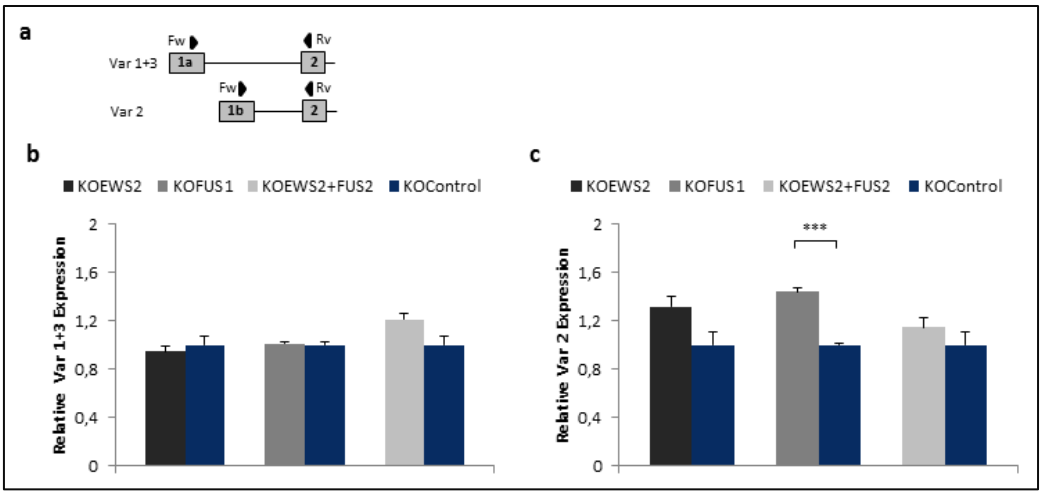


Figure 11 | Influence of EWS and FUS in the alternative 5' non-coding exon of hnRNPK mRNA. (a) Schematic representation of the alternative 5' non-coding exons of hnRNPK mRNA. The grey boxes represent the non-coding exons, identified with the corresponding number in white. The arrows represent forward and reverse primers used to amplify the corresponding hnRNPK mRNA (Var1+3 and Var 2). (b, c) HEK293 cells treated with KOEWS2, KOFUS1, KOEWS2+FUS2 and KOControl siRNAs were tested for relative hnRNPK mRNA variant expression by qPCR with Var 1+3 and Var 2 primer sets. All values were normalized to reference gene TBP. *P-values* for the comparisons are indicated by brackets as follows: ***P < 0.001. All experiments were performed in biological triplicates. Data presented as mean + SEM.

change in the siRNA-depleted cells, when compared to control (Figure 11 b). Oppositely, variant 2 appeared significantly up-regulated in KOFUS2 knockdown, when compared to control, whereas no significant changes were observed in KOEWS2 or KOEWS2+FUS2 knockdowns (Figure 11 c). These results indicate that FUS depletion favors usage of the alternative promoter for exon 1b and the production of hnRNPK mRNA variant 2.

Overall, data obtained from Transfection 2 experiments indicated that proteins FUS and EWS have a role in miR-7 processing as well as in hnRNPK general and alternative splicing.

Discussion

The main goal of this study was to elucidate the role of EWS and FUS in hnRNPK and miR production and processing. To accomplish this, EWS, FUS and EWS together with FUS were transiently depleted from HEK293 cells.

In the first round of transfection experiments (Transfection 1), a mix of two siRNAs for EWS and FUS was used to achieve protein depletion, both at mRNA and protein levels. When analyzing miR-7 expression in siRNA-depleted cells, it was up-regulated from 7 to 8-fold in both EWS and double knockdown cells (KOEWS1 and KOEWS1+FUS1), whereas no variation was detected in FUS-depleted cells (KOFUS1). This was concomitant with a down-regulation of the *bona fide* miR-7 mRNA target BCL2. However, both EWS and double knockdown cells divided less and presented a round shape when compared to FUS-depleted and control cells. These morphological changes could be related to miR-7 up-regulation, since it has been previously reported that miR-7 decreases cell proliferation and favors cell apoptosis[102; 103]. As nothing similar has been previously described for EWS-depleted HEK293 cells, this issue was investigated further. After blasting all siRNA sequences against the human reference genome it was discovered that one of the siRNAs used for EWS depletion, siEWSb, could be having off-targeting effects on TAF15 mRNA since it was complementary to TAF15 mRNA. In fact, TAF15 mRNA was down-regulated in both EWS and double knockdowns, when compared to control, while it was up-regulated in FUS knockdown. This was a problem for future data analysis, as TAF15 is a protein homolog of EWS and FUS and could thus be interfering with or contributing to the observed effects. A test-run transfection was conducted with the separate siRNAs used in EWS depletion (siEWSa and siEWSb) as well as with a mix of two specific siRNAs for TAF15 (siTAF15a and siTAF15b). We would expect a down-regulation of TAF15 mRNA in siEWSb-depleted cells. On the contrary, TAF15 mRNA was down-regulated in siEWSa-depleted cells instead of in siEWSb. We also quantified miR-7 expression in these samples and found it up-regulated in siEWSa-depleted cells as well as slightly up-regulated in siTAF15-depleted cells (data not shown). This could indicate that TAF15 and EWS coupled knockdown was responsible for the effects seen in miR-7 expression. However, in a following test-run transfection where TAF15 and EWS were depleted together using a new siRNA for EWS depletion (siEWSc), no significant variation of miR-7 was observed (data not shown). These results showed that miR-7 up-regulation and TAF15 down-regulation were being caused by the deregulation of unknown mechanisms and were therefore a side-effect induced by the siRNAs used in EWS depletion. In order to avoid these problems in the future, the siRNAs designed for protein depletion should be blasted

against the human genome before use. It would also be easier to use company pre-designed siRNAs, which have minimal side-effects. Moreover, it is important to be careful very about morphological changes of the cells subjected to the siRNA-mediated depletion, which could either specifically derive from the treatment or just indicate cellular stress and toxicity.

A second transfection round was done (Transfection 2) using the same siRNAs for the depletion of FUS and a new siRNA for EWS, siEWSc. The new siRNA proved to be as effective as the previously used at knocking down EWS both at mRNA and protein levels. Furthermore, EWS and double knockdown cells (KOEWS2 and KOEWS2+FUS2) did not present morphological alterations when compared to FUS-depleted (KOFUS2) and control cells, thus indicating that less cellular toxicity was induced by the new siRNA. Correspondingly, TAF15 was found up-regulated in all siRNA-depleted cells when compared to control, showing that it was no longer repressed by the siRNA treatment.

Surprisingly, in Transfection 2 siRNA-depleted cells, miR-7 expression was enhanced only about 50% in double knockdown cells (KOEWS2+FUS2). When evaluating BCL2 mRNA expression, it was up-regulated in EWS and FUS knockdown cells (KOEWS2 and KOFUS2) and unchanged in double knockdown cells. This may be explained by the fact that the variation of miR-7 in double knockdown cells was not sufficient to affect BCL2 mRNA expression. Despite this, we cannot exclude that miR-7 could have an effect on BCL2 protein levels since we did not quantify BCL2 protein levels. To clarify if EWS and FUS were influencing miR-7 precursor processing, pri-miR-7-1 and pre-miR-7-1, their expression levels were evaluated. pri-miR-7-1 was down-regulated in FUS-depleted cells (KOFUS1), with no variation in the other knockdowns, when compared to control. In addition, no significant changes were detected at pre-miR-7-1 level. Only the simultaneous knockdown of EWS and FUS caused overexpression of miR-7, which may indicate that EWS and FUS work together to inhibit mature miR-7 production. Since the levels of precursor miRNA levels remained stable, this inhibition would take place at the Dicer cleavage step, in the cytoplasm. It is possible that FUS and EWS bind to pre-miR-7-1 terminal loop and prevent Dicer and its associated proteins from accessing and recognizing the cleavage site, thus inhibiting its processing into mature miRNA. The shuttling ability of EWS and FUS further supports this hypothesis[20; 35]. Also, when knocking down FUS there is a decrease in pri-miR-7-1 but no variation of pre-miR-7-1 or miR-7. This could be explained by either lower pri-miRNA transcription or faster pri-miRNA processing. The first option seems less likely, since there was an accumulation of unspliced exon and intron 15, where pri-miR-7-1 is originated from. Therefore, when FUS is absent, there could faster recognition and cleavage of pri-miR-7-1 by Drosha and DGCR8. Oppositely, when FUS is present, it could be preventing pri-miR-7-1 processing also at the Drosha cleavage step. Unfortunately, a corresponding increase in pre-

miR-7-1 and miR-7 expression was not observed. pre-miR-7-1 could still be processed to miR-7, despite the fact that EWS was present and possibly repressing its conversion at the Dicer cleavage step, though to a lesser extent than if FUS was also there. Furthermore, since miR-7 is produced from two other cellular loci, it can contribute to buffering the variations caused by FUS depletion. Hence, FUS appears to inhibit miR-7 processing at two different levels, Drosha and Dicer cleavage, in the nucleus and cytoplasm. Numerous proteins have been shown to regulate miRNA expression at Drosha and/or Dicer steps[109]. For instance, the NF90-NF45 protein complex is able to bind endogenous pri-miRNAs and prevent its access to the Microprocessor Complex, thus inhibiting pri-miRNA processing[110]. Another protein, KSRP, is able to bind to the terminal loop of target miRNA precursors and enhance miRNA production both at Drosha and Dicer processing steps[111]. Despite this, we cannot rule out the fact that pre-miR-7-1 has been reported to be produced as part of hnRNPK transcriptional unit as well as having its own promoter, which appears to vary according to the study[101; 103; 112; 113].

Additionally, EWS and FUS proteins seem to also have a role in hnRNPK mRNA splicing. Quantification of spliced and unspliced hnRNPK mRNA indicated that in FUS or double knockdowns (KOFUS2 and KOEWS+FUS2) there was accumulation of unspliced exons and introns at the 3' end of the transcript. Furthermore, there was a decrease of the unspliced transcripts at the 5' end of hnRNPK mRNA in EWS and FUS knockdowns (KOFUS2 and KOEWS2). There also seemed to be a decrease of spliced mRNA at the 3' end, though to a lesser extent, since the variation was significant only in the double knockdown (KOEWS+FUS2). This may be explained by the total amount of molecules quantified by the primer sets used. It is expected that total amount of spliced mRNA is larger than the unspliced. Therefore, small changes in the unspliced hnRNPK mRNA pool would be significant whereas the same variation would not be significant in the spliced hnRNPK pool. Significant hnRNPK variations from qPCR experiments were very small. However, since these results were obtained from three biological replicates, they should be considered biologically significant. In addition, FUS, EWS and double knockdowns clearly demonstrated a shift in alternative splice acceptor site usage: in control cells the proximal acceptor splice site was mostly used, whereas in siRNA-depleted cells the distal acceptor splice site was at least as frequently used, thus increasing hnRNPK variant 3 and consequently isoform b production. Thus, EWS and FUS favor proximal splice acceptor site usage at the 3'end of hnRNPK and inhibit the production of variant 3. Interestingly, FUS also seems to prevent production of hnRNPK variant 2 which is produced by an alternative start codon, thus indicating that FUS may also contribute to which promoter is preferentially used for the transcription of hnRNPK.

Given the accumulation of unspliced transcripts at the 3'end of hnRNPK mRNA and the changes in alternative splicing site selection, evidence points toward FUS and EWS regulating hnRNPK splicing. However, this does not happen in a homogenous way since splicing was mainly affected at the end of the mRNA. This could be correlated with EWS and FUS binding the DNA downstream of *HNRNPK*. It is possible that EWS and FUS are coordinating transcription and splicing at the 3' end of hnRNPK mRNA. With FUS or both FUS and EWS depletion, there would be an uncoupling between RNA Pol II stalling at the 3' end of hnRNPK mRNA and recruitment of transcription termination and spliceosome factors, which would not be done as effectively. This would cause a less efficient splicing of the last exons, consequently accumulating at this region. Despite this, transcription and splicing would still occur, at a lower rate, since spliced mRNA levels did not vary much. The exons and introns at the 5' end of the gene would not be affected as they would be exposed longer to the splicing factors available in the nucleoplasm, thus having more probability of being recognized and spliced. FUS, and EWS to a certain extent, would therefore help RNA Pol II recruiting pre-mRNA processing factors and would coordinate the action of both transcription and splicing machineries at the 3' end of hnRNPK mRNA. Their absence would result in an impaired recruitment of mRNA processing factors to the 3'end of hnRNPK mRNA and thus to a less efficient pre-mRNA processing in this region. It is known that transcription and splicing are intimately connected and both are able to influence each other[114-116]. RNA Pol II has been shown to pause at the 5' and 3' end of genes promoting recruitment of pre-mRNA processing factors, which facilitates co-transcriptional pre-mRNA processing[115]. It was also reported that transcription termination is coupled to spliceosome assembly since preventing spliceosome recruitment caused unspliced mRNA to remain associated with stalled RNA Pol II and tethered to the DNA template downstream of the poly(A) site[116]. Additionally, EWS has been implicated in co-transcriptional exon skipping, where after inducing genotoxic stress in cells, the communication between EWS and YB-1 splicing factor was disrupted and several exons were skipped in *MDM2*. Thus EWS was pointed out as a bridge between RNA Pol II and the spliceosome[117]. EWS and FUS were found to frequently bind mRNA in introns, upstream of 3' end splice sites and to directly bind hnRNPK mRNA[22]. This may indicate that EWS and FUS simultaneously use DNA and RNA binding sites in order to coordinate transcription and splicing at the *HNRNPK* locus.

Considering obtained results, EWS seems to have a secondary role in this study, since unspliced mRNA accumulates in FUS or double knockdown cells but not in EWS-depleted cells. We cannot discard the fact that other mechanisms are at play here, such as mRNA degradation and transport. Besides, we did not address the possibility that RNA Pol II elongation speed along hnRNPK mRNA could alter pre-mRNA processing. It is known that RNA Pol II elongation rate is able to alter mRNA splicing: slower elongation rates would

delay synthesis of downstream competitor splice site sequences and thus favor the inclusion of weak upstream exons whereas faster elongation rates would lead to exon skipping[118]. For instance, EWS-FLI1 fusion protein has been shown to decrease the elongation rate of RNA Pol II, thus enhancing the production of a cyclin D1 isoform frequently overexpressed in tumors[119]. Therefore, EWS and FUS may also be modulating RNA Pol II elongation rate in order to control pre-mRNA processing at hnRNPK 3' end.

All the above-mentioned evidence suggests both FUS and EWS have a function in miRNA processing and splicing. FUS seems to be the main player whereas EWS completes and supports FUS in its function. Moreover, FUS and EWS role in splicing and miRNA processing seem to be distinct: while splicing is less efficient and distal splice acceptor site usage is favored in the absence of FUS and EWS, miRNA processing is enhanced at nuclear as well as cytoplasmic levels. A study done by Kim and Kim (2007)[81] mentioned that partially spliced transcript accumulation was concomitant with decreased Drosha pri-miRNA cleavage and no variation of spliced mRNA. This goes against the data obtained in this study, where unspliced mRNA levels were increased while miRNA processing was enhanced. However, a direct comparison may not be possible since the primer sets used in Kim and Kim (2007)[81] study amplified partially spliced mRNAs, whereas the ones used in this study amplified unspliced mRNA. Furthermore, the splicing factor SF2/ASF has been found to regulate splicing of hnRNPK mRNA as well as the expression of pre-miR-7-1[120]. SF2/ASF is able to directly bind pri-miR-7-1 and enhance its cleavage by Drosha. At the same time, SF2/ASF promotes proximal splice site acceptor usage in the pre-miR-7-1 containing intron. This would be accomplished by different domains of SF2/ASF protein.

A new concept has emerged that mRNA splicing and pri-miRNA processing take place co-transcriptionally and are consequently coupled[109]. Pri-miRNAs that are retained longer at transcription sites are more efficiently processed than the pri-miRNAs readily capped, polyadenylated and released from the transcriptional machinery[121]. Drosha is also found to co-localize with RNA Pol II and unspliced mRNAs are main Drosha substrates[82]. Furthermore, intronic pri-miRNAs are processed from unspliced introns before splicing occurs. These introns are processed more slowly when compared to pri-miRNA non-containing introns, indicating that Drosha cleavage does not prevent splicing but possibly delays it[81]. Therefore, strong evidence suggests that transcription, splicing and pri-miRNA processing are interconnected and are likely coordinated by common proteins. Spliceosome and miRNA processing machineries are probably competing for access to the same location in hnRNPK mRNA and also for available auxiliary factors, so it is possible that FUS and EWS are managing these two processes by stimulating one and preventing the other.

Conclusion and Future Perspectives

The aim of this study was to elucidate the role of EWS and FUS in miRNA processing. By transient depletion of both proteins it was discovered that they both participate in the processing of pre-miR-7-1, produced from *HNRNPK*. Results indicate that this regulation is done at two different steps: Drosha and Dicer processing. At Drosha level, FUS would possibly be binding to pri-miRNA and preventing Drosha cleavage. At Dicer level, both FUS and EWS would be hindering Dicer access to pre-miR-7-1, thus impeding its cleavage. Furthermore, we found that FUS and EWS regulate the splicing of hnRNPK mRNA and modulate the production of all three mRNA variants. Evidence points towards FUS and EWS having a simultaneous role in miRNA processing and splicing. Hence, they are most likely coordinating and integrating both miRNA and splicing machineries at the hnRNPK mRNA through the use of their DNA and RNA binding properties.

For future experiments, it would be interesting to overexpress EWS and FUS, transiently or stably, and verify if hnRNPK and miR-7 regulation could be compared to the results obtained in the current study. Hoell *et al*[22] results were based on the overexpression of the FET proteins and it is mentioned in that study that the correspondent knockdown yielded weak results which did not correlate with the overexpression data. In order to validate the proposed models, we would need to verify if EWS and/or FUS bind directly to miR-7 and its precursors by RNA immunoprecipitation. It would also be interesting to design a minigene containing the last two exons and intron of hnRNPK, similar to the one used in Wu *et al*[120]. This would allow a much more precise study of what is happening in hnRNPK, pre-miR-7-1 and miR-7 production since the other loci from which miR-7 is produced could be disregarded. Furthermore, it would be interesting to clarify if EWS and FUS are able to modify RNA Pol II elongation speed in hnRNPK transcription and whether this would have an effect on splicing at the 3' end. Finally, it would be important to verify if EWS and FUS have a general role in miRNA processing or if their regulation is confined to either intronic miRNAs or just to hnRNPK locus. This could be done with a miRNA microarray using RNA from cells subjected to EWS and FUS depletion or overexpression.

References

1. Bertolotti, A., Lutz, Y., Heard, D.J., Chambon, P., and Tora, L. (1996). 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 15, 5022-5031.
2. Andersson, M.K., Stahlberg, A., Arvidsson, Y., Olofsson, A., Semb, H., Stenman, G., Nilsson, O., and Aman, P. (2008). The multifunctional FUS, EWS and TAF15 proto-oncoproteins show cell type-specific expression patterns and involvement in cell spreading and stress response. BMC cell biology 9, 37.
3. Delattre, O., Zucman, J., Plougastel, B., Desmaze, C., Melot, T., Peter, M., Kovar, H., Joubert, I., de Jong, P., Rouleau, G., et al. (1992). Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours. Nature 359, 162-165.
4. Crozat, A., Aman, P., Mandahl, N., and Ron, D. (1993). Fusion of CHOP to a novel RNA-binding protein in human myxoid liposarcoma. Nature 363, 640-644.
5. Rabbitts, T.H., Forster, A., Larson, R., and Nathan, P. (1993). Fusion of the dominant negative transcription regulator CHOP with a novel gene FUS by translocation t(12;16) in malignant liposarcoma. Nature genetics 4, 175-180.
6. Hoffmann, A., and Roeder, R.G. (1996). Cloning and characterization of human TAF20/15. Multiple interactions suggest a central role in TFIID complex formation. The Journal of biological chemistry 271, 18194-18202.
7. Baumann, M., Pontiller, J., and Ernst, W. (2010). Structure and basal transcription complex of RNA polymerase II core promoters in the mammalian genome: an overview. Molecular biotechnology 45, 241-247.
8. Morohoshi, F., Ootsuka, Y., Arai, K., Ichikawa, H., Mitani, S., Munakata, N., and Ohki, M. (1998). Genomic structure of the human RBP56/hTAFII68 and FUS/TLS genes. Gene 221, 191-198.
9. Zinszner, H., Albalat, R., and Ron, D. (1994). A novel effector domain from the RNA-binding protein TLS or EWS is required for oncogenic transformation by CHOP. Genes & development 8, 2513-2526.
10. Bertolotti, A., Bell, B., and Tora, L. (1999). The N-terminal domain of human TAFII68 displays transactivation and oncogenic properties. Oncogene 18, 8000-8010.
11. Hackl, W., and Luhrmann, R. (1996). 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 264, 843-851.
12. Matthews, J.M., and Sunde, M. (2002). Zinc fingers--folds for many occasions. IUBMB life 54, 351-355.
13. Loughlin, F.E., Mansfield, R.E., Vaz, P.M., McGrath, A.P., Setiyaputra, S., Gamsjaeger, R., Chen, E.S., Morris, B.J., Guss, J.M., and Mackay, J.P. (2009). The zinc fingers of the SR-like protein ZRANB2 are single-stranded RNA-binding domains that recognize 5' splice site-like sequences. Proceedings of the National Academy of Sciences of the United States of America 106, 5581-5586.
14. Nguyen, C.D., Mansfield, R.E., Leung, W., Vaz, P.M., Loughlin, F.E., Grant, R.P., and Mackay, J.P. (2011). Characterization of a family of RanBP2-type zinc fingers that can recognize single-stranded RNA. Journal of molecular biology 407, 273-283.
15. Tan, A.Y., and Manley, J.L. (2009). The TET family of proteins: functions and roles in disease. Journal of molecular cell biology 1, 82-92.
16. Burd, C.G., and Dreyfuss, G. (1994). Conserved structures and diversity of functions of RNA-binding proteins. Science 265, 615-621.
17. Ohno, T., Ouchida, M., Lee, L., Gatalica, Z., Rao, V.N., and Reddy, E.S. (1994). The EWS gene, involved in Ewing family of tumors, malignant melanoma of soft parts and desmoplastic small round cell tumors, codes for an RNA binding protein with novel regulatory domains. Oncogene 9, 3087-3097.
18. Lerga, A., Hallier, M., Delva, L., Orvain, C., Gallais, I., Marie, J., and Moreau-Gachelin, F. (2001). Identification of an RNA binding specificity for the potential splicing factor TLS. The Journal of biological chemistry 276, 6807-6816.

19. Guipaud, O., Guillonneau, F., Labas, V., Praseuth, D., Rossier, J., Lopez, B., and Bertrand, P. (2006). An in vitro enzymatic assay coupled to proteomics analysis reveals a new DNA processing activity for Ewing sarcoma and TAF(II)68 proteins. Proteomics 6, 5962-5972.
20. Zinszner, H., Sok, J., Immanuel, D., Yin, Y., and Ron, D. (1997). TLS (FUS) binds RNA in vivo and engages in nucleo-cytoplasmic shuttling. Journal of cell science 110 (Pt 15), 1741-1750.
21. Takahama, K., Kino, K., Arai, S., Kurokawa, R., and Oyoshi, T. (2008). Identification of RNA binding specificity for the TET-family proteins. Nucleic Acids Symp Ser (Oxf), 213-214.
22. Hoell, J.I., Larsson, E., Runge, S., Nusbaum, J.D., Duggimpudi, S., Farazi, T.A., Hafner, M., Borkhardt, A., Sander, C., and Tuschl, T. (2011). RNA targets of wild-type and mutant FET family proteins. Nature structural & molecular biology 18, 1428-1431.
23. Zakaryan, R.P., and Gehring, H. (2006). Identification and characterization of the nuclear localization/retention signal in the EWS proto-oncoprotein. Journal of molecular biology 363, 27-38.
24. Dormann, D., Rodde, R., Edbauer, D., Bentmann, E., Fischer, I., Hruscha, A., Than, M.E., Mackenzie, I.R., Capell, A., Schmid, B., et al. (2010). ALS-associated fused in sarcoma (FUS) mutations disrupt Transportin-mediated nuclear import. The EMBO journal 29, 2841-2857.
25. Marko, M., Vlassis, A., Guialis, A., and Leichter, M. (2012). Domains involved in TAF15 subcellular localisation: Dependence on cell type and ongoing transcription. Gene 506, 331-338.
26. Leemann-Zakaryan, R.P., Pahlich, S., Grossenbacher, D., and Gehring, H. (2011). Tyrosine Phosphorylation in the C-Terminal Nuclear Localization and Retention Signal (C-NLS) of the EWS Protein. Sarcoma 2011, 218483.
27. Lanson, N.A., Jr., Maltare, A., King, H., Smith, R., Kim, J.H., Taylor, J.P., Lloyd, T.E., and Pandey, U.B. (2011). A Drosophila model of FUS-related neurodegeneration reveals genetic interaction between FUS and TDP-43. Human molecular genetics 20, 2510-2523.
28. Bertolotti, A., Melot, T., Acker, J., Vigneron, M., Delattre, O., and Tora, L. (1998). EWS, but not EWS-FLI-1, is associated with both TFIID and RNA polymerase II: interactions between two members of the TET family, EWS and hTAFII68, and subunits of TFIID and RNA polymerase II complexes. Molecular and cellular biology 18, 1489-1497.
29. Zhou, H., and Lee, K.A. (2001). An hsRPB4/7-dependent yeast assay for trans-activation by the EWS oncogene. Oncogene 20, 1519-1524.
30. Gascoyne, D.M., Thomas, G.R., and Latchman, D.S. (2004). The effects of Brn-3a on neuronal differentiation and apoptosis are differentially modulated by EWS and its oncogenic derivative EWS/Fli-1. Oncogene 23, 3830-3840.
31. Lee, J., Rhee, B.K., Bae, G.Y., Han, Y.M., and Kim, J. (2005). Stimulation of Oct-4 activity by Ewing's sarcoma protein. Stem Cells 23, 738-751.
32. Powers, C.A., Mathur, M., Raaka, B.M., Ron, D., and Samuels, H.H. (1998). TLS (translocated-in-liposarcoma) is a high-affinity interactor for steroid, thyroid hormone, and retinoid receptors. Mol Endocrinol 12, 4-18.
33. Pahlich, S., Quero, L., Roschitzki, B., Leemann-Zakaryan, R.P., and Gehring, H. (2009). 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 8, 4455-4465.
34. Fujii, R., Okabe, S., Urushido, T., Inoue, K., Yoshimura, A., Tachibana, T., Nishikawa, T., Hicks, G.G., and Takumi, T. (2005). The RNA binding protein TLS is translocated to dendritic spines by mGluR5 activation and regulates spine morphology. Current biology : CB 15, 587-593.
35. Belyanskaya, L.L., Gehrig, P.M., and Gehring, H. (2001). Exposure on cell surface and extensive arginine methylation of ewing sarcoma (EWS) protein. The Journal of biological chemistry 276, 18681-18687.
36. Bertrand, P., Akhmedov, A.T., Delacote, F., Durrbach, A., and Lopez, B.S. (1999). Human POMp75 is identified as the pro-oncoprotein TLS/FUS: both POMp75 and POMp100 DNA homologous pairing activities are associated to cell proliferation. Oncogene 18, 4515-4521.
37. Baechtold, H., Kuroda, M., Sok, J., Ron, D., Lopez, B.S., and Akhmedov, A.T. (1999). Human 75-kDa DNA-pairing protein is identical to the pro-oncoprotein TLS/FUS and is able to promote D-loop formation. The Journal of biological chemistry 274, 34337-34342.
38. Hicks, G.G., Singh, N., Nashabi, A., Mai, S., Bozek, G., Klewes, L., Arapovic, D., White, E.K., Koury, M.J., Oltz, E.M., et al. (2000). Fus deficiency in mice results in defective B-lymphocyte development and activation, high levels of chromosomal instability and perinatal death. Nature genetics 24, 175-179.
39. Kuroda, M., Sok, J., Webb, L., Baechtold, H., Urano, F., Yin, Y., Chung, P., de Rooij, D.G., Akhmedov, A., Ashley, T., et al. (2000). Male sterility and enhanced radiation sensitivity in TLS(-/-) mice. The EMBO journal 19, 453-462.

40. Li, H., Watford, W., Li, C., Parmelee, A., Bryant, M.A., Deng, C., O'Shea, J., and Lee, S.B. (2007). Ewing sarcoma gene EWS is essential for meiosis and B lymphocyte development. The Journal of clinical investigation 117, 1314-1323.
41. Wang, X., Arai, S., Song, X., Reichart, D., Du, K., Pascual, G., Tempst, P., Rosenfeld, M.G., Glass, C.K., and Kurokawa, R. (2008). Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. Nature 454, 126-130.
42. Sanchez-Ramos, C., Tierrez, A., Fabregat-Andres, O., Wild, B., Sanchez-Cabo, F., Arduini, A., Dopazo, A., and Monsalve, M. (2011). PGC-1alpha regulates translocated in liposarcoma activity: role in oxidative stress gene expression. Antioxidants & redox signaling 15, 325-337.
43. Hume, D.A., Sasmono, T., Himes, S.R., Sharma, S.M., Bronisz, A., Constantin, M., Ostrowski, M.C., and Ross, I.L. (2008). The Ewing sarcoma protein (EWS) binds directly to the proximal elements of the macrophage-specific promoter of the CSF-1 receptor (csf1r) gene. J Immunol 180, 6733-6742.
44. Rappsilber, J., Ryder, U., Lamond, A.I., and Mann, M. (2002). Large-scale proteomic analysis of the human spliceosome. Genome research 12, 1231-1245.
45. Leichter, M., Marko, M., Ganou, V., Patrinou-Georgoula, M., Tora, L., and Guialis, A. (2011). A fraction of the transcription factor TAF15 participates in interactions with a subset of the spliceosomal U1 snRNP complex. Biochimica et biophysica acta 1814, 1812-1824.
46. Wahl, M.C., Will, C.L., and Luhrmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. Cell 136, 701-718.
47. Calvio, C., Neubauer, G., Mann, M., and Lamond, A.I. (1995). Identification of hnRNP P2 as TLS/FUS using electrospray mass spectrometry. RNA 1, 724-733.
48. Chansky, H.A., Hu, M., Hickstein, D.D., and Yang, L. (2001). Oncogenic TLS/ERG and EWS/Fli-1 fusion proteins inhibit RNA splicing mediated by YB-1 protein. Cancer research 61, 3586-3590.
49. Yang, L., Embree, L.J., Tsai, S., and Hickstein, D.D. (1998). Oncoprotein TLS interacts with serine-arginine proteins involved in RNA splicing. The Journal of biological chemistry 273, 27761-27764.
50. Yang, L., Chansky, H.A., and Hickstein, D.D. (2000). EWS.Fli-1 fusion protein interacts with hyperphosphorylated RNA polymerase II and interferes with serine-arginine protein-mediated RNA splicing. The Journal of biological chemistry 275, 37612-37618.
51. Meissner, M., Lopato, S., Gotzmann, J., Sauermann, G., and Barta, A. (2003). Proto-oncoprotein TLS/FUS is associated to the nuclear matrix and complexed with splicing factors PTB, SRm160, and SR proteins. Experimental cell research 283, 184-195.
52. Hallier, M., Lerga, A., Barnache, S., Tavitian, A., and Moreau-Gachelin, F. (1998). The transcription factor Spi-1/PU.1 interacts with the potential splicing factor TLS. The Journal of biological chemistry 273, 4838-4842.
53. Paronetto, M.P., Minana, B., and Valcarcel, J. (2011). The Ewing sarcoma protein regulates DNA damage-induced alternative splicing. Molecular cell 43, 353-368.
54. Gregory, R.I., Yan, K.P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N., and Shiekhattar, R. (2004). The Microprocessor complex mediates the genesis of microRNAs. Nature 432, 235-240.
55. Shiohama, A., Sasaki, T., Noda, S., Minoshima, S., and Shimizu, N. (2007). Nucleolar localization of DGCR8 and identification of eleven DGCR8-associated proteins. Experimental cell research 313, 4196-4207.
56. Kovar, H. (2011). Dr. Jekyll and Mr. Hyde: The Two Faces of the FUS/EWS/TAF15 Protein Family. Sarcoma 2011, 837474.
57. Sjogren, H., Meis-Kindblom, J., Kindblom, L.G., Aman, P., and Stenman, G. (1999). Fusion of the EWS-related gene TAF2N to TEC in extraskeletal myxoid chondrosarcoma. Cancer research 59, 5064-5067.
58. Aman, P. (1999). Fusion genes in solid tumors. Seminars in cancer biology 9, 303-318.
59. Riggi, N., Cironi, L., Provero, P., Suva, M.L., Stehle, J.C., Baumer, K., Guillou, L., and Stamenkovic, I. (2006). Expression of the FUS-CHOP fusion protein in primary mesenchymal progenitor cells gives rise to a model of myxoid liposarcoma. Cancer research 66, 7016-7023.
60. Spitzer, J.I., Ugras, S., Runge, S., Decarolis, P., Antonescu, C., Tuschl, T., and Singer, S. (2011). mRNA and protein levels of FUS, EWSR1, and TAF15 are upregulated in liposarcoma. Genes, chromosomes & cancer 50, 338-347.
61. Kwiatkowski, T.J., Jr., Bosco, D.A., Leclerc, A.L., Tamrazian, E., Vanderburg, C.R., Russ, C., Davis, A., Gilchrist, J., Kasarskis, E.J., Munsat, T., et al. (2009). Mutations in the FUS/TLS

gene on chromosome 16 cause familial amyotrophic lateral sclerosis. Science 323, 1205-1208.
62. Vance, C., Rogelj, B., Hortobagyi, T., De Vos, K.J., Nishimura, A.L., Sreedharan, J., Hu, X., Smith, B., Ruddy, D., Wright, P., et al. (2009). Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. Science 323, 1208-1211.
63. Colombrita, C., Onesto, E., Tiloca, C., Ticozzi, N., Silani, V., and Ratti, A. (2011). RNA-binding proteins and RNA metabolism: a new scenario in the pathogenesis of Amyotrophic lateral sclerosis. Archives italiennes de biologie 149, 83-99.
64. Lagier-Tourenne, C., Polymenidou, M., and Cleveland, D.W. (2010). TDP-43 and FUS/TLS: emerging roles in RNA processing and neurodegeneration. Human molecular genetics 19, R46-64.
65. Deng, H.X., Zhai, H., Bigio, E.H., Yan, J., Fecto, F., Ajroud, K., Mishra, M., Ajroud-Driss, S., Heller, S., Sufit, R., et al. (2010). FUS-immunoreactive inclusions are a common feature in sporadic and non-SOD1 familial amyotrophic lateral sclerosis. Annals of neurology 67, 739-748.
66. Gal, J., Zhang, J., Kwinter, D.M., Zhai, J., Jia, H., Jia, J., and Zhu, H. (2011). Nuclear localization sequence of FUS and induction of stress granules by ALS mutants. Neurobiology of aging 32, 2323 e2327-2340.
67. Bosco, D.A., Lemay, N., Ko, H.K., Zhou, H., Burke, C., Kwiatkowski, T.J., Jr., Sapp, P., McKenna-Yasek, D., Brown, R.H., Jr., and Hayward, L.J. (2010). Mutant FUS proteins that cause amyotrophic lateral sclerosis incorporate into stress granules. Human molecular genetics 19, 4160-4175.
68. Anderson, P., and Kedersha, N. (2009). Stress granules. Current biology : CB 19, R397-398.
69. Couthouis, J., Hart, M.P., Shorter, J., DeJesus-Hernandez, M., Erion, R., Oristano, R., Liu, A.X., Ramos, D., Jethava, N., Hosangadi, D., et al. (2011). A yeast functional screen predicts new candidate ALS disease genes. Proceedings of the National Academy of Sciences of the United States of America 108, 20881-20890.
70. Ticozzi, N., Vance, C., Leclerc, A.L., Keagle, P., Glass, J.D., McKenna-Yasek, D., Sapp, P.C., Silani, V., Bosco, D.A., Shaw, C.E., et al. (2011). Mutational analysis reveals the FUS homolog TAF15 as a candidate gene for familial amyotrophic lateral sclerosis. American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics 156B, 285-290.
71. Couthouis, J., Hart, M.P., Erion, R., King, O.D., Diaz, Z., Nakaya, T., Ibrahim, F., Kim, H.J., Mojsilovic-Petrovic, J., Panossian, S., et al. (2012). Evaluating the role of the FUS/TLS-related gene EWSR1 in amyotrophic lateral sclerosis. Human molecular genetics 21, 2899-2911.
72. Neumann, M., Bentmann, E., Dormann, D., Jawaid, A., DeJesus-Hernandez, M., Ansorge, O., Roeber, S., Kretzschmar, H.A., Munoz, D.G., Kusaka, H., et al. (2011). FET proteins TAF15 and EWS are selective markers that distinguish FTLD with FUS pathology from amyotrophic lateral sclerosis with FUS mutations. Brain : a journal of neurology 134, 2595-2609.
73. Kim, V.N., Han, J., and Siomi, M.C. (2009). Biogenesis of small RNAs in animals. Nature reviews Molecular cell biology 10, 126-139.
74. Eulalio, A., Huntzinger, E., and Izaurralde, E. (2008). Getting to the root of miRNA-mediated gene silencing. Cell 132, 9-14.
75. Krol, J., Loedige, I., and Filipowicz, W. (2010). The widespread regulation of microRNA biogenesis, function and decay. Nature reviews Genetics 11, 597-610.
76. Lee, Y., Kim, M., Han, J., Yeom, K.H., Lee, S., Baek, S.H., and Kim, V.N. (2004). MicroRNA genes are transcribed by RNA polymerase II. The EMBO journal 23, 4051-4060.
77. Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S., et al. (2003). The nuclear RNase III Drosha initiates microRNA processing. Nature 425, 415-419.
78. Lee, Y., Jeon, K., Lee, J.T., Kim, S., and Kim, V.N. (2002). MicroRNA maturation: stepwise processing and subcellular localization. The EMBO journal 21, 4663-4670.
79. Han, J., Lee, Y., Yeom, K.H., Nam, J.W., Heo, I., Rhee, J.K., Sohn, S.Y., Cho, Y., Zhang, B.T., and Kim, V.N. (2006). Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. Cell 125, 887-901.
80. Altuvia, Y., Landgraf, P., Lithwick, G., Elefant, N., Pfeffer, S., Aravin, A., Brownstein, M.J., Tuschl, T., and Margalit, H. (2005). Clustering and conservation patterns of human microRNAs. Nucleic acids research 33, 2697-2706.
81. Kim, Y.K., and Kim, V.N. (2007). Processing of intronic microRNAs. The EMBO journal 26, 775-783.

82. Morlando, M., Ballarino, M., Gromak, N., Pagano, F., Bozzoni, I., and Proudfoot, N.J. (2008). Primary microRNA transcripts are processed co-transcriptionally. Nature structural & molecular biology 15, 902-909.
83. Berezikov, E., Chung, W.J., Willis, J., Cuppen, E., and Lai, E.C. (2007). Mammalian mirtron genes. Molecular cell 28, 328-336.
84. Ruby, J.G., Jan, C.H., and Bartel, D.P. (2007). Intronic microRNA precursors that bypass Drosha processing. Nature 448, 83-86.
85. Lund, E., Guttinger, S., Calado, A., Dahlberg, J.E., and Kutay, U. (2004). Nuclear export of microRNA precursors. Science 303, 95-98.
86. Bernstein, E., Caudy, A.A., Hammond, S.M., and Hannon, G.J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature 409, 363-366.
87. Gregory, R.I., Chendrimada, T.P., Cooch, N., and Shiekhattar, R. (2005). Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. Cell 123, 631-640.
88. Maniataki, E., and Mourelatos, Z. (2005). A human, ATP-independent, RISC assembly machine fueled by pre-miRNA. Genes & development 19, 2979-2990.
89. Carthew, R.W., and Sontheimer, E.J. (2009). Origins and Mechanisms of miRNAs and siRNAs. Cell 136, 642-655.
90. Schwarz, D.S., Hutvagner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P.D. (2003). Asymmetry in the assembly of the RNAi enzyme complex. Cell 115, 199-208.
91. Bartel, D.P. (2009). MicroRNAs: target recognition and regulatory functions. Cell 136, 215-233.
92. Croce, C.M. (2009). Causes and consequences of microRNA dysregulation in cancer. Nature reviews Genetics 10, 704-714.
93. Mendell, J.T., and Olson, E.N. (2012). MicroRNAs in stress signaling and human disease. Cell 148, 1172-1187.
94. Salta, E., and De Strooper, B. (2012). Non-coding RNAs with essential roles in neurodegenerative disorders. Lancet neurology 11, 189-200.
95. Bomsztyk, K., Denisenko, O., and Ostrowski, J. (2004). hnRNP K: one protein multiple processes. BioEssays : news and reviews in molecular, cellular and developmental biology 26, 629-638.
96. Frederikse, P.H., Donnelly, R., and Partyka, L.M. (2006). miRNA and Dicer in the mammalian lens: expression of brain-specific miRNAs in the lens. Histochemistry and cell biology 126, 1-8.
97. Li, X., and Carthew, R.W. (2005). A microRNA mediates EGF receptor signaling and promotes photoreceptor differentiation in the Drosophila eye. Cell 123, 1267-1277.
98. Lee, H.J., Palkovits, M., and Young, W.S., 3rd. (2006). miR-7b, a microRNA up-regulated in the hypothalamus after chronic hyperosmolar stimulation, inhibits Fos translation. Proceedings of the National Academy of Sciences of the United States of America 103, 15669-15674.
99. Bravo-Egana, V., Rosero, S., Molano, R.D., Pileggi, A., Ricordi, C., Dominguez-Bendala, J., and Pastori, R.L. (2008). Quantitative differential expression analysis reveals miR-7 as major islet microRNA. Biochemical and biophysical research communications 366, 922-926.
100. Chen, H., Shalom-Feuerstein, R., Riley, J., Zhang, S.D., Tucci, P., Agostini, M., Aberdam, D., Knight, R.A., Genchi, G., Nicotera, P., et al. (2010). miR-7 and miR-214 are specifically expressed during neuroblastoma differentiation, cortical development and embryonic stem cells differentiation, and control neurite outgrowth in vitro. Biochemical and biophysical research communications 394, 921-927.
101. Chou, Y.T., Lin, H.H., Lien, Y.C., Wang, Y.H., Hong, C.F., Kao, Y.R., Lin, S.C., Chang, Y.C., Lin, S.Y., Chen, S.J., et al. (2010). EGFR promotes lung tumorigenesis by activating miR-7 through a Ras/ERK/Myc pathway that targets the Ets2 transcriptional repressor ERF. Cancer research 70, 8822-8831.
102. Kefas, B., Godlewski, J., Comeau, L., Li, Y., Abounader, R., Hawkinson, M., Lee, J., Fine, H., Chiocca, E.A., Lawler, S., et al. (2008). microRNA-7 inhibits the epidermal growth factor receptor and the Akt pathway and is down-regulated in glioblastoma. Cancer research 68, 3566-3572.
103. Reddy, S.D., Ohshiro, K., Rayala, S.K., and Kumar, R. (2008). MicroRNA-7, a homeobox D10 target, inhibits p21-activated kinase 1 and regulates its functions. Cancer research 68, 8195-8200.
104. Webster, R.J., Giles, K.M., Price, K.J., Zhang, P.M., Mattick, J.S., and Leedman, P.J. (2009). Regulation of epidermal growth factor receptor signaling in human cancer cells by microRNA-7. The Journal of biological chemistry 284, 5731-5741.
105. Xiong, S., Zheng, Y., Jiang, P., Liu, R., Liu, X., and Chu, Y. (2011). MicroRNA-7 inhibits the growth of human non-small cell lung cancer A549 cells through targeting BCL-2. International journal of biological sciences 7, 805-814.

106. Kong, X., Li, G., Yuan, Y., He, Y., Wu, X., Zhang, W., Wu, Z., Chen, T., Wu, W., Lobie, P.E., et al. (2012). MicroRNA-7 Inhibits Epithelial-to-Mesenchymal Transition and Metastasis of Breast Cancer Cells via Targeting FAK Expression. PloS one 7, e41523.
107. Bieberstein, N.I., Oesterreich, F.C., Straube, K., and Neugebauer, K.M. (2012). First exon length controls active chromatin signatures and transcription. Cell reports 2, 62-68.
108. Thomsen, R., Solvsten, C.A., Linnet, T.E., Blechingberg, J., and Nielsen, A.L. (2010). Analysis of qPCR data by converting exponentially related Ct values into linearly related X0 values. Journal of bioinformatics and computational biology 8, 885-900.
109. Pawlicki, J.M., and Steitz, J.A. (2010). Nuclear networking fashions pre-messenger RNA and primary microRNA transcripts for function. Trends in cell biology 20, 52-61.
110. Sakamoto, S., Aoki, K., Higuchi, T., Todaka, H., Morisawa, K., Tamaki, N., Hatano, E., Fukushima, A., Taniguchi, T., and Agata, Y. (2009). The NF90-NF45 complex functions as a negative regulator in the microRNA processing pathway. Molecular and cellular biology 29, 3754-3769.
111. Trabucchi, M., Briata, P., Garcia-Mayoral, M., Haase, A.D., Filipowicz, W., Ramos, A., Gherzi, R., and Rosenfeld, M.G. (2009). The RNA-binding protein KSRP promotes the biogenesis of a subset of microRNAs. Nature 459, 1010-1014.
112. Corcoran, D.L., Pandit, K.V., Gordon, B., Bhattacharjee, A., Kaminski, N., and Benos, P.V. (2009). Features of mammalian microRNA promoters emerge from polymerase II chromatin immunoprecipitation data. PloS one 4, e5279.
113. Ozsolak, F., Poling, L.L., Wang, Z., Liu, H., Liu, X.S., Roeder, R.G., Zhang, X., Song, J.S., and Fisher, D.E. (2008). Chromatin structure analyses identify miRNA promoters. Genes & development 22, 3172-3183.
114. Dye, M.J., Gromak, N., and Proudfoot, N.J. (2006). Exon tethering in transcription by RNA polymerase II. Molecular cell 21, 849-859.
115. Glover-Cutter, K., Kim, S., Espinosa, J., and Bentley, D.L. (2008). RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. Nature structural & molecular biology 15, 71-78.
116. Martins, S.B., Rino, J., Carvalho, T., Carvalho, C., Yoshida, M., Klose, J.M., de Almeida, S.F., and Carmo-Fonseca, M. (2011). Spliceosome assembly is coupled to RNA polymerase II dynamics at the 3' end of human genes. Nature structural & molecular biology 18, 1115-1123.
117. Dutertre, M., Sanchez, G., De Cian, M.C., Barbier, J., Dardenne, E., Gratadou, L., Dujardin, G., Le Jossic-Corcos, C., Corcos, L., and Auboeuf, D. (2010). Cotranscriptional exon skipping in the genotoxic stress response. Nature structural & molecular biology 17, 1358-1366.
118. Carrillo Oesterreich, F., Bieberstein, N., and Neugebauer, K.M. (2011). Pause locally, splice globally. Trends in cell biology 21, 328-335.
119. Sanchez, G., Bittencourt, D., Laud, K., Barbier, J., Delattre, O., Auboeuf, D., and Dutertre, M. (2008). 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 105, 6004-6009.
120. Wu, H., Sun, S., Tu, K., Gao, Y., Xie, B., Krainer, A.R., and Zhu, J. (2010). A splicing-independent function of SF2/ASF in microRNA processing. Molecular cell 38, 67-77.
121. Pawlicki, J.M., and Steitz, J.A. (2008). Primary microRNA transcript retention at sites of transcription leads to enhanced microRNA production. The Journal of cell biology 182, 61-76.
122. Huang, L., Nakai, Y., Kuwahara, I., and Matsumoto, K. (2012). PRAS40 is a functionally critical target for EWS repression in Ewing sarcoma. Cancer research 72, 1260-1269.
123. Fox-Walsh, K., Davis-Turak, J., Zhou, Y., Li, H., and Fu, X.D. (2011). A multiplex RNA-seq strategy to profile poly(A+) RNA: application to analysis of transcription response and 3' end formation. Genomics 98, 266-271.
124. Jobert, L., Argentini, M., and Tora, L. (2009). PRMT1 mediated methylation of TAF15 is required for its positive gene regulatory function. Experimental cell research 315, 1273-1286.

Appendix

Appendix I. Oligonucleotide sequences

Supplementary Table 1 | Sequences used in double transfection knockdown.

Name	Sequence (5'-3')	Size (bp)	Start Position (bp)	Source
siEWSa	CGAGGAGGAAGGAGAGAAA	19	310	siDesign*
siEWSb	GAGUAGCUAUGGUCAACAA	19	1015	siDesign*
siEWSc	GACUCUGACAACAGUGCAAUU	21	1411	Huang *et al*[122]
siFUSa	ACAGCCCAUGAUUAAUUUGUA	21	1037	Fox-Walsh *et al*[123]
siFUSb	GGGAGAAGGCCAAAUGAUA	19	1957	siDesign*
siCONTROL	CUGAUGCAGGUAAUCGCGU(dTdT)	19	-	Nielsen's Lab**
siTAF15a	UGAUCAGCGCAACCGACCATT	21	1870	siDesign*
siTAF15b	UGAUCAGCGCAACCGACCA	19	1870	Jobert *et al*[124]

*siRNA obtained from siDesign Center, Dharmacon. ** siRNA designed by our laboratory.

Supplementary Table 2 | Primers sequences used except for hnRNPK mRNA.

Gene	Forward Primer (5'-3')	Reverse Primer (5'-3')
EWS	CTACAGCCAAGCTCCAAGTC	CCCATAAACACCCATGCTAC
FUS	CCCTACGGACAGCAGAGTTA	CTGGGGAGTTGACTGAGTTC
TAF15	GGGAGCACAGTGTCATTT	ATTTCCGCATGACGGATTAG
BCL2*	ATGTGTGTGGAGAGCGTCAACC	TGAGCAGAGTCTTCAGAGACAGCC
RAF1	TGCATGACTGCCTTATGAAA	TCAAAGACGCAGCATCAGTA
IGFR1	AGGGGAATTTCATCCCAAAT	CTGATAGTCGTTGCGGATGT
EGFR	ACTGCTGCCACAACCAGT	GGGGTTGTAGAGCATGAGTG
LPHN2	AGAAGCCGATGTGTGAGAAG	GTGGCAGCACTCATTGTTCT
TBP	TGCACAGGAGCCAAGAGTCAA	CACATCACAGCTCCCCACCA
UBC	CCTGGTGCTCCGTCTTAGAG	TTTCCCAGCAAAGATCAACC
GAPDH	ATGGGGAAGGTGAAGGTCGGAG	GATGACAAGCTTCCCGTTCTCAGC
Pri-miR7-1	GCCATGGTGTCTCAACCTTT	GGTTAAGGCTTCTTCCAGCA
Pre-miR-7-1	GTTGGCCTAGTTCTGTGTGG	CAGACTGTGATTTGTTGTCGAT
RNU48	GTGATGATGACCCCAGGTAAC	GTGATGGCATCAGCGACAC
Pre-miR-7-1 rev.**	-	CCTGTGCCATATGGCAGACT
RNU48 rev.**	-	GGTCAGAGCGCTGCGGTGAT

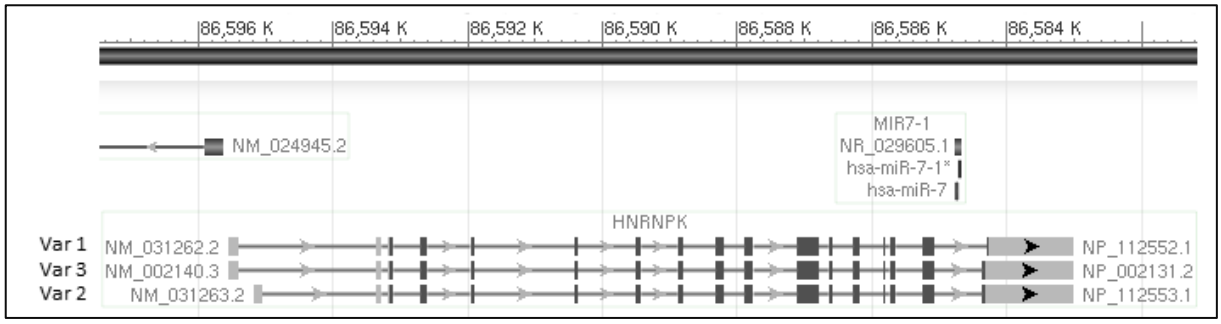
* Primer sequence obtained from Xiong *et al*[105]. ** Gene specific reverse primers used only for cDNA synthesis.

Supplementary Table 3 | Primer sequences for hnRNPK mRNA quantification.

Name	Forward Primer (5'-3')	Reverse Primer (5'-3')	Location
E3F_I3R	GCCAGAAGAAACCTTCCCTA	CCACACACTCCTAAGGCAAT	Exon 3- Intron 3
E9F_E10R	ACTTTGACTGCGAGTTGAGG	AAGCTTGATGGTGGTTTGAG	Exon 9–Exon 10
E10F_I10R	ATTGGAGGAAAACCCGATAG	GCTCTGAAGCTACTTTTGCAG	Exon 10- Intron 10
I10F_I10R	CCTTTGAGCCTTTGAATGAA	CTGAGGCACCTGAGATAGGA	Intron 10- Intron 10
I13F_E14R	GGCTCCGGATATGGTAAGTT	TAGGTCCACCAAGATCACCA	Intron 13- Exon 14
I14F_E15R	TGACAGGTTTAGGGAGCGTTA	CCTTTGCCAATAATAGATCCAG	Intron 14- Exon 15
E15F_E16R*	GCCTTTAGAAGGATCCGAAG	CAACATCTGCATACTGCTTCAC	Exon 15- Exon 16
E15F_I15R	GGACCAGATACAGAATGCACA	CCAAAAGGTTGAGACACCAT	Exon 15- Intron 15
I15F_I15R	CAACCTTTTGGGGACCTAAC	GCTTGGTTAAGGCTTCTTCC	Intron 15- Intron 15
Var 1+3**	CAGACGCCATTATCCTCTGT	TATTAAACGGGCACACCAAT	Exon 1- Exon 2
Var 2***	CTGCGCTCGTTTTCTGTCTA	TATTAAACGGGCACACCAAT	Exon 1- Exon 2
E15F_E16R-S	GCCTTTAGAAGGATCCGAAG	CCCCAAATGTTACAGTGACC	Exon 15- Exon 16

* Amplifies hnRNPK variant 1 and 2. ** Amplifies hnRNPK variant 1 and 3. *** Amplifies hnRNPK Variant 2

Appendix II. *HNRNPK* locus and primers.



Supplementary Figure 1 | *HNRNPK* and pre-miR-7-1 locus. Representation of *HNRNPK* and pre-miR-7-1 locus on chromosome 9, modified from NCBI database. hnRNPK has three mRNA variants, Var 1,2 and 3, which encode for two different protein isoforms. In the last intron there is the production of pre-miR-7-1. On the top is the scale bar where each interval represents 2 Kb.

TTGGCTGGATCTATTATTGGCAAAGGTGGTCAGCGGATTAAACAAATCCGTCATGAGTCGGGAGCTTCG

ATCAAAATTGATGAGCCTTTAGAAGGATCCGAAGATCGGATCATTACCATTACAGGAACACAGGACCAG

ATACAGAATGCACAGTATTTGCTGCAGAACAGgtcagtttaagtttagctttgtgttagcttatacata

ctaaaacctttaaaaagcttttcttctcaattgatttttttcttttagaagccatggtgtctcaaccttt

ttggggacctaacttctaaacattctaatagtttgccttaatttttcttctgctttcttactaaaaatg

aagacattcaatactaatcttgctggaagaagccttaaccaagcaaacttctcatttctctggtgaaaa

ctgctgccaaaaccacttgttaaaaattgtacagagcctgtagaaaatatagaagattcattggatgtt

ggcctagttctgtgtggaagactagtgattttgttgtttttagataactaaatcgacaacaaatcacag

tctgccatatggcacaggccatgcctctacaggacaaatgattggtgctgtaaaatgcagcatttcaca

ccttactagcattctttgtcttttctaccaaatattaacaactttcaattccgttttcttaattctgtt

ctactaatgtccgatttactactcatcatttttcttgacacttaacattgctttaatttgtaattgcta

atggtttttgaactcttctaattgtaatggacgtgtttatcattttaatttagcattgaaattgtcttg

PSA
atgttgattaagatgttacgagcaaaaattctaattcgcacttttttttctttttctttttatagTGTGAA

DSA
GCAGTATGCAGATGTTGAAGGATTCTAATGCAAGATATTTTTTCTTTTTTATAGTGTGAAGCAGTATTC

TGGAAAGTTTTTCTAAGACTAGTGAAGAACTGAAGGAGTCCTGCATCTTTTTTTTTTTATCTGCTTCTG

TTTAAAAAGCCAACATTCCTCTGCTTCATAGGTGTTCTGCATTTGAGGTGTAGTGAAATCTTTGCTGTT

CACCAGATGTAATGTTTTAGTTCCTTACAAACAGGGTTGGGGGGGGGAAGGGCGTGCAAAAACTAACAT

TGAAATTTTGAAACAGCAGCAGAGTGAGTGGATTTTATTTTTCGTTATTGTTGGTGGTTTAAAAAATTC

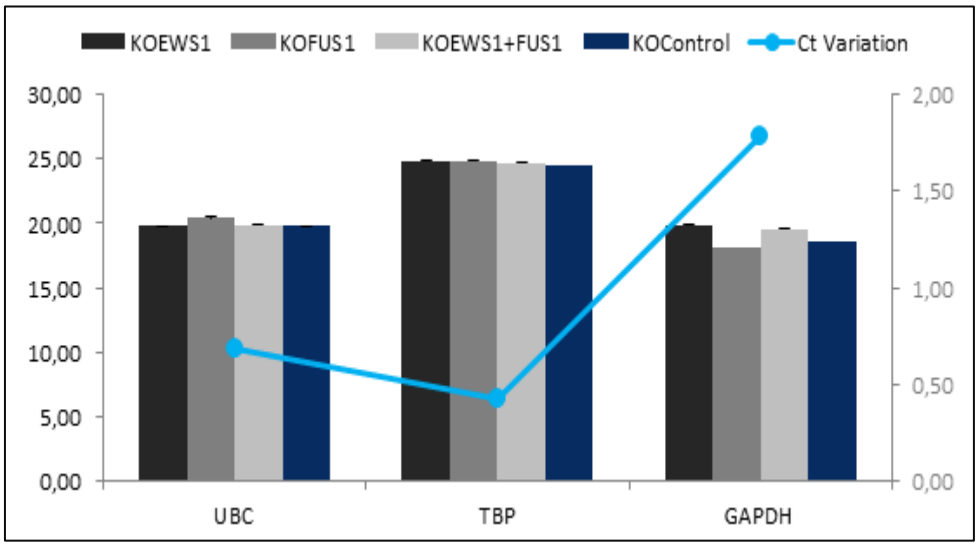
CCCCCATGTAATTATTGTGAACACCTTGCTTTGTGGTCACTGTAACATTTGGGGGGGTGGGACAGGGAGG

AAAAGTAACAATAGTCCACATGTCCCTGGCATCTGTTCAGAGCAGTGTGCAGAATGTAATGCTCTTTTG

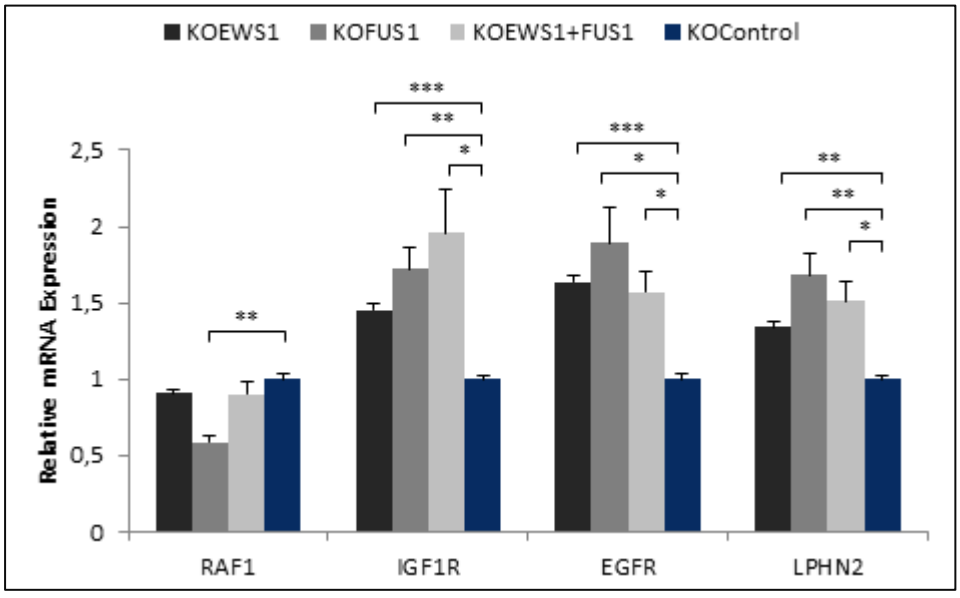
TAAGAAACGTTTTATGATTTTTAAAATAAATTTAGTGAACCTATTTTTGGTGGTCATTTTTTTTTTAAG

Supplementary Figure 2 | *HNRNPK* 3' end sequence. This sequence contains the penultimate exon, last intron and part of the last exon of *HNRNPK*. Exons and introns are represented in uppercase and lowercase, respectively. The underlined blue sequence is present in hnRNPK variant 1 and 2 but not in variant 3 and derives from an alternative proximal splice acceptor (PSA) site. The last exon of variant 3 starts immediately after the blue underline sequence and originates from an alternative distal splice acceptor (DSA) site. The underlined red sequences are the locations where the forward and reverse primers used in the alternative splicing assay anneal. The genomic sequence was obtained from UCSC Genome Browser (http://genome.ucsc.edu/, hg19).

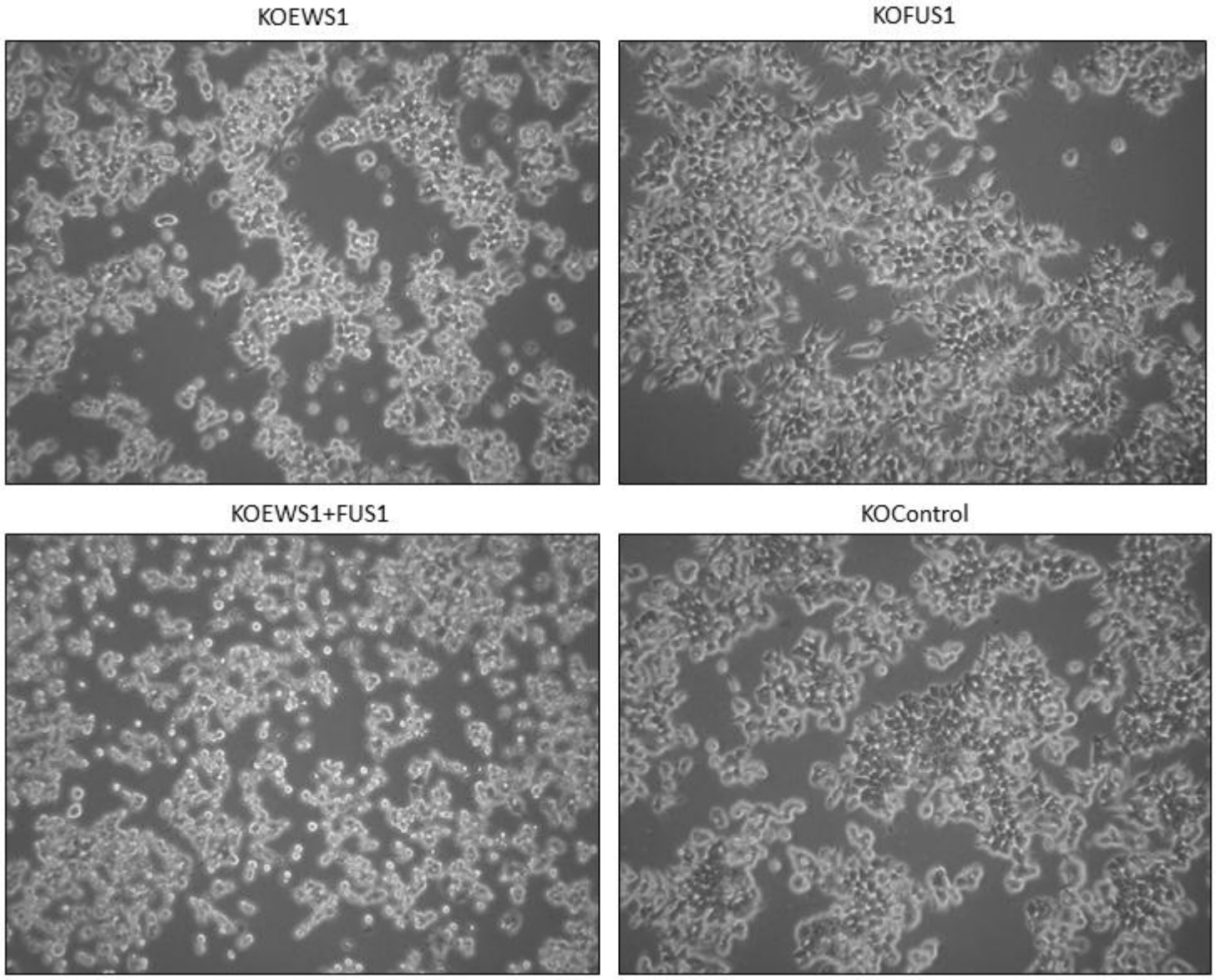
Appendix III. Supplementary Data



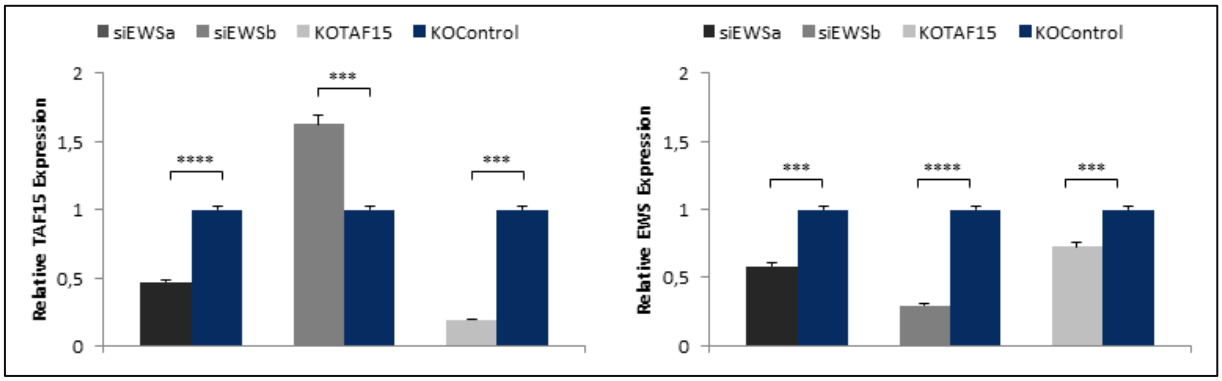
Supplementary Figure 3 | TBP is the most stable reference gene tested in siRNA-depleted EWS and FUS cells. UBC, TBP and GAPDH reference genes were quantified by qPCR in KOEWS1,KOFUS1 and KOEWS1+FUS1 depleted HEK293 cells. In the left axis are depicted the cycle threshold (Ct) values obtained for each knockdown treatment and control. On the right axis is depicted the Ct variation between knockdown treatments for each gene. This experiment was performed once. Data presented as mean of technical triplicates + SEM.



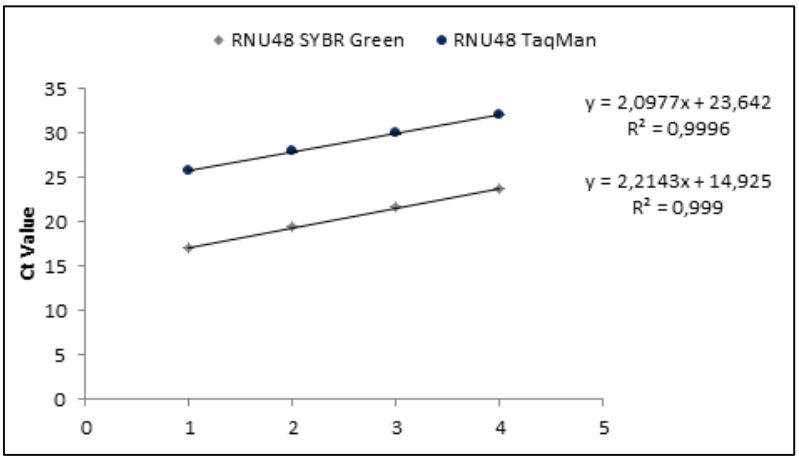
Supplementary Figure 4 | miR-7 target gene quantification. RAF1, IGFR, EGFR, LPHN2 genes were quantified by qPCR in KOEWS1, KOFUS1, KOEWS1+FUS1 and KOControl siRNA-depleted HEK293 cells. All genes were normalized to the reference gene TBP. *P-values* for the comparisons are indicated by brackets as follows: *P < 0.05; **P < 0.01, ***P < 0.001 Two individual experiments with the same conditions were performed. Data presented as mean + SEM.



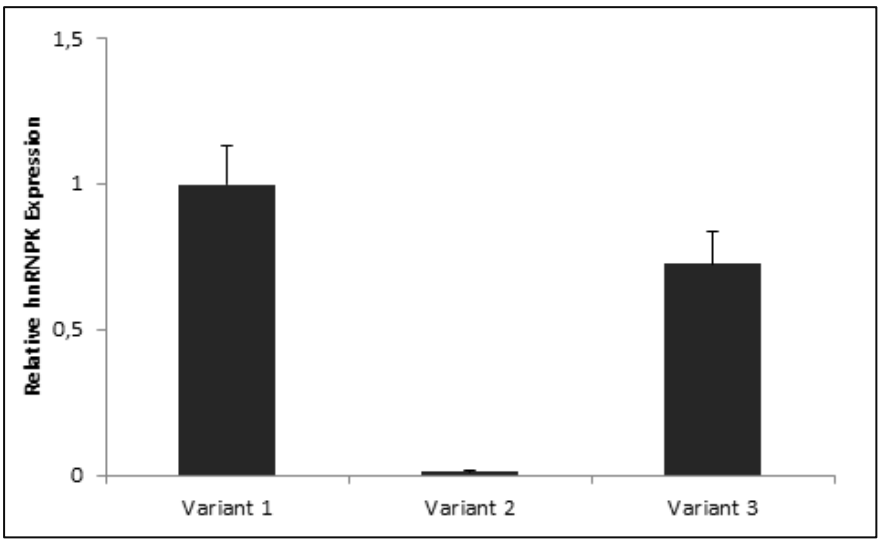
Supplementary Figure 5 | KOEWS1 and KOEWS1+FUS1 show altered cell morphology. Pictures of HEK293 cells subjected to siRNA-mediated depletion of EWS (KOEWS1), FUS (KOFUS1), EWS and FUS (KOEWS1+FUS1) and control (KOControl) were obtained by phase contrast microscopy. KOEWS1 and KOEWS1+FUS1 present a round shape when compared to KOFUS1 and KOControl, which have a more elongated shape.



Supplementary Figure 6 | TAF15 is down-regulated in siEWSa siRNA-depleted cells. HEK293 cells were subjected to double-transfection with siEWSa, siEWSb, KOTAF15 (mix of two siRNAs: siTAF15a and siTAF15b) and KOControl siRNAs. siRNA-depleted cells were used for TAF15 and EWS quantification by qPCR. All genes were normalized to the reference gene TBP. *P-values* for the comparisons are indicated by brackets as follows: *P < 0.05; **P < 0.01, ***P < 0.001, ****P < 0.0001. Experiment was performed once. Data presented as mean of technical triplicates + SEM.



Supplementary Figure 7 | RNU48 probe designed for SYBR Green dye varies equally to RNU48 probe used for TaqMan Assays. cDNA from HEK293 cells specific for SYBR Green or TaqMan master mixes was diluted in a 1:4 dilution series with four points and then quantified by qPCR with the correspondent master mixes. The mean of the obtained Ct values for each probe was plotted onto a graph and compared. On the right are the equations corresponding to both trendlines. Both probes vary in a similar way since the slope of both amplifications is comparable. This experiment was done only once. Data presented as mean of technical triplicates + SEM.



Supplementary Figure 8 | Relative quantification of hnRNPK mRNA variants. HEK293 cells were tested for hnRNPK mRNA variants by qPCR. This was done indirectly, since it was not possible to design primers which amplified the variants individually. Therefore, primers E15F_E16R (amplifies variants 1 and 2), Var 1+3 (amplifies variants 1 and 3) and Var 2 (amplifies variant 2) were used. The final quantification of mRNA variants was accomplished by subtraction of the different primer sets relative amounts. This experiment was done in biological triplicates. Data presented as mean + SEM.