#toutdonner

Université de Liège

Faculté des Sciences - Département des Sciences de la Vie

GIGA - Molecular Biology of Diseases

Laboratory of Gene Expression and Cancer (GEC)

"A novel role of EWSR1 protein in RNA translation regulation"

Zahrat El Oula HASSOUN

Doctoral dissertation to obtain the grade of "Doctorat en Sciences", in

"Biochimie, Biologie moléculaire et cellulaire, Bioinformatique et modélisation"

Promotor: Pr. Franck DEQUIEDT

Academic year 2022-2023

EXAMINATION COMMITTEE

Pr. Marc Thiry (Committee chair)
Université de Liège, GIGA Neurosciences, Laboratory of Cellular and Tissular Biology

Pr. Franck Dequiedt (Promotor)
Université de Liège, GIGA, Laboratory of Gene Expression and Cancer

Dr. Yvette Habraken (Secretory)
Université de Liège, GIGA, Laboratory of Gene Expression and Cancer

Dr. Denis Mottet
Université de Liège, GIGA, Laboratory of Gene Expression and Cancer

Dr. Arnaud Blomme
Université de Liège, GIGA, Cancer Signaling

Pr. Cyril Gueydan
Université Libre de Bruxelles, IBMM, Laboratoire de Biologie moléculaire du gène

Pr. Katrin Karbstein
University of Florida, UF Scripps Biomedical Research, Department of Integrative Structural and Computational Biology

ABSTRACT

The Ewing Sarcoma breakpoint region 1 protein (EWSR1) is a member of the FET (FUS, EWSR1, and TAF15) family of proteins, implicated in several aspects of cell biology and found mutated in multiple diseases. Primarily, *EWSR1* attracted broad attention as one of the most susceptible genes to breakage/translocation, originally identified because of the t(11;22)(q24;q12) chromosomal translocation, characteristic of Ewing sarcoma. Interestingly, EWSR1 contains two functional domains: a potent N-terminal transcriptional activation domain (NTD), with several conserved serine-tyrosine-glycine-glutamine (SYGQ) repeats, revealed in the context of oncogenic EWSR1-fusions, and a C-terminal domain (CTD), modulating different aspects of RNA processing. Although extensive studies on oncogenes generated in Ewing sarcoma, only limited information was available about the function of EWSR1 protein itself. It is only until recently, that various EWSR1 cellular functions have been revealed, but the exact mechanisms are not yet understood. Importantly, EWSR1 was identified as a multifunctional protein implicated in transcriptional and post-transcriptional mechanisms. Considering its implication in post-transcriptional regulation, we decided to investigate whether EWSR1 might also be involved in mRNA translation.

In this work, we highlight that EWSR1 is a repressor of translation of a specific subset of mRNA and decipher the molecular mechanisms underlying this function. We show that this new activity is mediated by the RGG2-ZnF-RGG3 region of the CTD. Interestingly, we evidence that this region is also implicated in the association of EWSR1 with the 40S ribosomal subunit. Therefore, we conclude that EWSR1-mediated repression of translation is correlated with its ability to associate with the 40S. Importantly, we show that EWSR1 activity in the repression of translation is crucial for the regulation of lipid homeostasis. This function is primordial to maintain normal endoplasmic reticulum (ER) shape and function, but also to properly activate the unfolded protein response (UPR^{ER}). In the future, these findings might lead to novel therapeutic opportunities for diseases in which EWSR1 is implicated, especially that the CTD region, which is responsible for the translational regulation, was reported to be mutated in several diseases or totally lost in oncogenic EWSR1-ETS fusions, such as Ewing sarcoma. Verily, unraveling EWSR1 at the molecular level is not only important for understanding gene expression, but is also of medical relevance, as aberrant function of the protein is the basis of many human diseases.

Keywords: EWSR1, FET family, Ewing sarcoma, multifunctional protein, mRNA translation, 40S, lipid metabolism, ER, UPR^{ER}, diseases.

RÉSUMÉ

La protéine de la région 1 du point d'arrêt du sarcome d'Ewing (EWSR1) est un membre de la famille des protéines FET (FUS, EWSR1 et TAF15), impliquée dans plusieurs aspects de la biologie cellulaire et mutée dans de multiples maladies. Principalement, *EWSR1* a attiré une large attention comme l'un des gènes les plus susceptibles à la rupture/translocation, qui a été identifié à l'origine à cause de la translocation chromosomique t(11;22)(q24;q12), caractéristique du sarcome d'Ewing. De manière intéressante, EWSR1 contient deux domaines fonctionnels: un puissant domaine d'activation de la transcription en position amino-terminale (NTD), avec plusieurs répétitions conservées de sérine-tyrosine-glycine-glutamine (SYGQ), révélées dans le contexte de fusions oncogéniques d'EWSR1, et un autre domaine en position carboxy-terminale (CTD), qui module différents aspects du processing d'ARN. Bien qu'il y a eu des études approfondies sur les oncogènes générés dans le sarcome d'Ewing, seules des informations limitées étaient disponibles sur la fonction de la protéine EWSR1 elle-même. Ce n'est que récemment que diverses fonctions cellulaires d'EWSR1 ont été révélées, mais les mécanismes exacts ne sont pas encore compris. De manière importante, EWSR1 a été identifiée comme étant une protéine multifonctionnelle impliquée dans les mécanismes transcriptionnels et post-transcriptionnels. Compte tenu de son implication dans la régulation post-transcriptionnelle, nous avons décidé d'étudier si EWSR1 pourrait également être impliquée dans la traduction de l'ARNm.

Dans ce travail, nous mettons en évidence qu'EWSR1 est un répresseur de la traduction d'un sous-ensemble spécifique d'ARNm et déchiffrons les mécanismes moléculaires sous-jacents à cette fonction. Nous montrons que cette nouvelle activité est médiée par la région RGG2-ZnF-RGG3 du CTD. De manière intéressante, nous montrons aussi que cette région est impliquée dans l'association avec la sous-unité ribosomale, la 40S. Par conséquent, nous concluons que la répression de la traduction médiée par EWSR1 est corrélée à sa capacité de s'associer avec la 40S. De manière importante, nous montrons que l'activité d'EWSR1 dans la répression de la traduction est cruciale pour la régulation de l'homéostasie lipidique. Cette fonction est primordiale pour maintenir la forme et la fonction normales du réticulum endoplasmique (RE), mais aussi pour activer proprement la unfolded protein response (UPR^{ER}). Dans l'avenir, ces découvertes pourraient conduire à de nouvelles opportunités thérapeutiques pour les maladies dans lesquelles EWSR1 est impliquée, surtout que la région CTD, qui est en charge de la régulation de la traduction, a été rapportée comme étant mutée dans plusieurs maladies ou totalement perdue dans les maladies de fusions oncogéniques de EWSR1-ETS, comme le sarcome d'Ewing. En vérité, le fait de démêler EWSR1 au niveau moléculaire n'est pas uniquement important pour comprendre l'expression des gènes, mais ceci est également d'une pertinence médicale, car la fonction aberrante de la protéine est à la base de nombreuses maladies humaines.

Mots clés: EWSR1, famille de FET, sarcome d'Ewing, protéine multifonctionnelle, traduction d’ARNm, 40S, métabolisme lipidique, RE, UPR^{ER}, maladies.

ACKNOWLEDGEMENTS

“Nothing in life is to be feared, it is only to be understood. Now is the time to understand more, so that we may fear less.” – Marie Curie. Science is one of the greatest endeavors of humanity. It contributes to our understanding of the world and provides solutions that benefit millions across the globe. Ever since I was a young child, I have been fascinated by the scientific field. Still, to date, news about discoveries never fails to give me goosebumps. This fascination and continued pursuit of knowledge motivated me to pursue my PhD, which was indeed a great experience, and the best is yet to come!

I would like to thank my promoter, Franck, for welcoming me in his lab of Gene Expression and Cancer (GEC). Thank you so much for your time, support and for the productive and fruitful discussions during the four years of my PhD. I truly enjoyed our discussions! I appreciate your guidance and the time you have taken to assist me. I would also like to thank you for the opportunity that you gave me to perform diverse and valuable experimental techniques that will certainly help me in developing my professional skills. What I learned by your side is enormous! By your side, learning also includes having “some?” fun. No, I would rather say: I had a blast. I mean thanks to you, I met Saint-Nicola and the King of pancakes. They were awesome! Hahaha! Thank you so much for this hilarious journey!

I would like to thank my jury members for accepting to examinate my thesis manuscript. Thank you, Profs. Cyril Gueydan and Katrin Karbstein, Drs. Yvette Habraken, Denis Mottet and Arnaud Blomme, for your valuable feedback.

Particularly, I thank Pr. Marc Thiry for keeping me posted, so I can manage the procedure for my thesis defense on time. Thank you for your kindness and readiness to answer my questions!

I would like to thank Yvette and Jean-Claude for their help during my PhD. Thank you for always being ready to answer my questions. I truly appreciate your generosity for sharing ideas and lab materials! Thank you, Yvette, for the most delicious “carotte cake”! Thank you, Jean-Claude, for the amazing discussions and advice!

Florence, what a shining person with a warm heart! I know that Allah loves me so much for sending me such an amazing “bench/life friend”. Your actions during these four years, were full of forgiveness, kindness and much of love towards others. I know that sometimes (only sometimes… Hahaha!), I can be very messy. However, I never heard you (my dear neighbor) complaining. I would never forget those times when you put my things (that were already on your bench, so sorry for that!) in their place, without even telling me about it. Keep it up like that girl, humanity needs more wonderful and kind people like you! And, as I was always saying: you are such an intelligent scientific. Overall, you are just perfect, and I am not complaining!

Margaux, or should I say the very elegant Margaux Claes/Classe! Besides being a very lovely, beautiful, kind and very smart girl; Margaux, you are a classy lady, who chooses very wisely her words. For instance, all humanity says OK; you say “D’accord!” We say that’s funny; you say “C’est hilarant!” We say that’s very good; you say “Excellent!” Maybe someday we can go and meet the Queen of Belgium and ask her for a title? Just give me a call, whenever ready... Hahaha! I wish you all the best and happiness in your life, you deserve it!

Bartimée, you have been such a wonderful friend! Thank you for being so kind, so polite and so respectful. I truly enjoyed working with you during these years. Among other qualities, I do appreciate your willingness to help without even taking the second to think about it. You explain and provide the information with patience and kindness. You are a modest and humble person. Thank you for being there whenever I needed you, and a big “thank you” for all the rides! Hope this friendship will last forever (maybe even for the afterlife? Hahaha!). Indeed, indeed, I was laughing (finally) at your jokes! Wish you all the best and happiness in your new journey!

Djonhathan, or should I say President J.B.? Hahaha! Thank you for being such a nice and kind gentleman. Thank you for your hospitality, for supporting my stupid jokes, and for even laughing at them! Hahaha! But you agree that they are awesome and even the best jokes in our galaxy, right? I am very glad that I met your family and your amazing dog. Your kids are just stunning. Oh, these beautiful angels, I am melting!! Do not forget to tell them that they **must** come to visit me in Lebanon: “it only takes four hours from Belgium”! I wish you all the happiness in the world!

Loic Ongena, a special person with a special knowledge. Honestly, I admired and still admire your ability to acquire so much information in various disciplines. Your willingness to learn is unique. But the most precious information (among all) that you should keep in mind is that “COLD CAUSES DISEASES”! You can really (like really) get sick!! PMID: 01.01.1994. Hahaha! Apart from being an excellent scientific, you are an amazing, kind, and wonderful friend. Once you love, you give from all your heart, without even asking for the “merci”.
Thank you for your valuable help for my project, for your time and your readiness to correct and to discuss ideas! Thank you for everything! You will stay in my top 5 WhatsApp list!

Michelle, I would like to thank for your kindness and your patience! Thank you for the experiments that you performed when I was no more in Belgium. I wish you all the best in your life!

Emeline, thank you for your nice and sweet spirit. You are a very genuine soul! I appreciate your help during my thesis. You were always there to answer my questions with no hesitation. I am very happy that I met someone like you, someone that appreciate friendship. All the happiness in your life!

Our beautiful Italian Benedetta, our little amazing Eva, Mr. Greg, and our new Laurence, although we didn't spend a lot of time together, I really enjoyed every second with you, guys!

"Les anciens" of the GEC laboratory, thank you for the memories!

Judita Sanchez Gil, I am very honored that I already met the future Nobel prize holder of Virology! I am glad that I had someone during my PhD, with who I could share my big ideas in science. Someone with the same passion and love for research, someone who truly loves viruses, Jajajaja! Besides science, I also enjoyed our discussions about other type of big ideas like "The Kardashians" and "The Real Housewives of Beverly Hills"; like of course, Duh! Do not forget: one day, we should meet Kim and Kanye (or Ye!) and ask them to fund our research, OK? See you soon, inchallah!

Chloé, the most beautiful flower!! Thank you so much for your kindness! I am truly happy that I knew someone like you. Someone who cares about others. I am impressed by your ability to organize many events but also to check if everything was great. I mean what would we do without you?! I am also very impressed by your ability to remember so much information, especially at the administrative level, which was very helpful to me! I wish you all the best in your life! So much love!!

Mademoiselle Megane Jassin: what a beautiful, wonderful, and amazing soul. I am just speechless! Thank you for being there during this time. I truly enjoyed your friendship and liked your open spirit. Your respect towards others is unique. I adored your gifts: the painting of the Lebanese flag, my cup of milk with Nesquik and the honey. My father is very grateful for the Belgian honey. He sends his greetings! May Allah protect you, your cats, your dog, your rabbit, and your bees! Take care of them. I hope we will meet again very soon! And do not forget all the amazing Arabic words that I taught you. A quizzzzz will be waiting for you.

Sibusisso, Sibu, My friend! What an amazing friend! Thank you for your actions that reflect the noble person behind. As you said once: "That's what friends are for". I really enjoyed the time we spent together. Thank you for your advice, you are a very good listener. Thank you for being always ready to help, to share, but also to act. All the best and success in your life!

Jeremy, although I always wanted to hit you, I'm kidding, Hahaha! Listen, it is only because Loic told me to do so (kidding again, Hahaha!!). I discovered how nice and kind you are! Thank you for the memories and all the best for your doc!

Yasmine and Stef, thank you for your kindness! I wish you all the best in your life and career!

I would like to thank the virology and immunology lab members: Cédric, Marielle, Julien, Maxime, Margaud, Amandine, Catherine and Sylvie for all the beautiful moments (team building, Christmas dinners...) that we shared together! Thank you for your kindness!

Big thanks for Denis's lab: Sanaa, Laurent, Alixia, Clara, Céline, Amandine, Guiliua. Thank you, for being so nice, so cheerful and full of happiness!

I would like to thank the cancer signaling lab members, especially: Kataryna, Arnaud, Xinyi, Sebastian, Christian, Najla for their kindness and for being ready to share knowledge and materials.

Thank you, Arnaud Blomme for your kindness! You are a blessing for science. As I already told you: I truly believe that the scientific field needs more people like you! Wishing you all the best for the future. May you achieve everything that you want!

Sarah Hanach, my dear Lebanese mate! Being away from our sweet home could be hard and challenging. However, you were there whenever I needed you. Thank you for your honesty, kindness, and your hospitality! I mean how can someone be that kindness?! All the best for your new chapter in life! You deserve the best!

Despina, my precious Greek mate! Isn't just amazing that my geographical neighbor is also my very dear friend! You are such a blessing! Thank you for your calm and kind spirit! Thank you for all the memories that will last forever. I wish you all the best in your life!

Houda, my sister! What can I add more?! You are a blessing. I couldn't imagine that one day I will meet someone with all your qualities. Indeed, I am honored that I knew you. You forget yourself for the sake of others. People like you are keeping the world turning! Houda, you are very blessed for having Jalil by your side. He is an amazing and incredible person! May Allah protect you and your family. We will meet soon inchallah! Bisous bisous à mon petit Dani!

Sihem, thank you for all the delicious food! We will talk very soon!

Thank you Fabienne for your kindness!

My beloved unique family: Mother, Father, Mohamad, Hamza, Fatima, Ali; my "World…my backbone"! No words can describe what you are and what you have done for me. You were there since the beginning. I am very grateful for having such a loyal, supportive, and caring family! A family with unconditional love. I am just honored to be part of this fabulous, amazing family! None of these achievements would have seen the light without you. Thank you for the endless encouragements. Thank you for remaining alongside me in thick and flimsy. We will

continue the adventure together; we will be grateful for the good and learn lessons from struggles. All the love!

Thank you, Allah! You are who I look up to. Thank you for gracing my life with uncountable opportunities that I know they are not of my hands or any other human hands. You showed me that it is a scientific fact that gratitude reciprocates. You taught me that everyone has his own battle, so it is primordial to stay humble, and that each chapter of our life is a lesson. Thank you for giving me the opportunity to learn but also to develop! Thank you!

TABLE OF CONTENTS

1 INTRODUCTION 1

1.1 The EWSR1 protein 1
1.1.1 FET proteins 1
1.1.2 Domain structure and function of EWSR1 5
1.1.3 EWSR1 functions in DNA and RNA metabolism 8
1.1.4 Physiological roles of EWSR1 12
1.1.5 FET proteins and neurodegenerative disorders 13
1.1.6 EWSR1 and sarcomas 16

1.2 mRNA Translation 22
1.2.1 Gene expression: a highly regulated process 22
1.2.2 Cap-dependent translation in eukaryotes 23
1.2.3 Regulation of translation steps 29
1.2.4 Alternative translation initiation mechanisms 42
1.2.5 Ribosomes 47
1.2.6 Methods of translatome profiling 59

1.3 The endoplasmic reticulum (ER) 64
1.3.1 ER functions 64
1.3.2 ER structure 67
1.3.3 ER shaping and remodeling 70
1.3.4 Membrane contact sites between ER and organelles 71
1.3.5 Protein quality control systems of the ER 74
1.3.6 ER chaperones 75
1.3.7 The unfolded protein response (UPR^{ER}) 76
1.3.8 ERAD 91
1.3.9 ER-Phagy 91
1.3.10 ER stress and diseases 92
1.3.11 UPR^{ER} modulating compounds 93

2 AIM 94

3 RESULTS 95

3.1 EWSR1 is a repressor of mRNA translation 95

3.2 EWSR1 translational target mRNAs exhibit specific features 106

3.3 EWSR1 binds its translational targets and represses the translation of a tethered reporter 110
3.4 The RGG2-ZnF-RGG3 region is responsible for the translational function of EWSR1 115
3.5 EWSR1 associates with the 40S ribosomal subunit 121
3.6 EWSR1 represses translation downstream of ATG scanning 129
3.7 The translational function of EWSR1 is linked to lipid homeostasis in cancer cells . 133
3.8 Deletion of EWSR1 impairs the activation of the unfolded protein response of the ER (UPR^{ER}) 141
3.9 EWSR1 KD leads to impairment activation of UPR^{ER} under ER stress 146

4 DISCUSSION AND PERSPECTIVES 150
4.1 EWSR1 is an inhibitor of mRNA translation 150
4.2 RIBO-seq: a powerful tool to determine the translational regulation by EWSR1 152
4.3 FET proteins and the repression of translation 153
4.4 Specific features of EWSR1 translational target mRNAs 159
4.5 EWSR1 binds to ribosomes via its CTD 160
4.6 Deregulation of EWSR1 translational activity might be implicated in several diseases 161
4.7 Possible EWSR1 mechanism in the repression of translation 167
4.8 EWSR1 localization and activity under stress conditions 169
4.9 EWSR1 and PTMs 170
4.10 EWSR1 and cellular homeostasis 174

5 MATERIALS AND METHODS 181
5.1 Provenience of cell lines and cell culture conditions 181
5.2 Cell treatments 181
5.3 Plasmids and cloning 181
5.4 Plasmid DNA and siRNA transfection 182
5.5 RNA isolation and quantitative PCR 182
5.6 Western blotting 183
5.7 MS2-tethering assay 183

5.8 SUnSET assay 183
5.9 Proliferation assay 184
5.10 Annexin V apoptosis assay 184
5.11 FACS labeling and analysis 184
5.12 Co-immunoprecipitation (co-IP) 184
5.13 Oligo(dT) pulldowns 185
5.14 RNA-immunoprecipitation-qPCR analysis 185
5.15 Polysome profiling 186
5.16 Nuclear and cytoplasmic fractionation 186
5.17 Cap-association assay using m^7GTP-Agarose 186
5.18 Subcellular fractionation 187
5.19 GST-fusion proteins purification and pulldowns 187
5.20 Ribosome profiling (RIBO-seq) 188
5.21 Lipidomic analysis 189
5.22 Transmission electron microscopy (TEM) 190
5.23 Statistics 191
5.24 Acknowledgements 191

6 APPENDIX 201
6.1 Plasmids map 201

7 REFERENCES 208

8 PUBLICATIONS 251

LIST OF FIGURES

Figure 1: Domain organization of FET proteins...... 3
Figure 2: Model for RNA-nucleated assembly of FUS proteins and subsequent recruitment of RNA polymerase II (Pol II)...... 5
Figure 3: Schematic representation of the EWSR1-FLI1 fusion gene.. 20
Figure 4: Overview of gene expression pathways.. 23
Figure 5: Model of the canonical pathway of eukaryotic translation initiation, elongation, termination, and ribosome recycling...... 25
Figure 6: Model of eukaryotic translation elongation.. 25
Figure 7: Translation termination in eukaryotes.. 27
Figure 8: Translation recycling in eukaryotes.. 28
Figure 9: Schematic representation of the integrated stress response (ISR).. 31
Figure 10: Regulatory 5'UTR RNA structures that influence translation initiation by promoting or inhibiting cap-dependent translation. 33
Figure 11: Trans-acting factors interacting with their specific RNA motifs...... 34
Figure 12: Secondary structure and amino-acylation of tRNAs. 36
Figure 13: Modifications within the tRNA anticodon region.. 37
Figure 14: Factors that influence translation elongation...... 39
Figure 15: The detailed steps of the ribosome-associated protein quality control (RQC) pathway.. 40
Figure 16: Histone h4 mRNA localization on the 80S ribosome.. 42
Figure 17: 3'CITE translation initiation used by plant RNA viruses.. 43
Figure 18: CircRNA in cap-independent translation. 45
Figure 19: Translation initiation mediated by IRES...... 47
Figure 20: Composition of bacterial and eukaryotic ribosomes and the common core.. 48
Figure 21: Architecture of the 40S (left site) and 60S (right site) ribosome subunits with new protein nomenclature.. 50
Figure 22: Representation of the different steps of eukaryotic ribosome biogenesis.. .. 51
Figure 23: Timeline of evolution of the concept of specialized ribosomes.. 54
Figure 24: Schematic representation of different types of ribosomes heterogeneity.. ... 54
Figure 25: A revised model for tissue-specific phenotypes of RP mutations.. 58
Figure 26: Main translatomic methods investigating the translatome.. 59
Figure 27: Transcriptional and translational regulation of gene expression...... 63
Figure 28: The implication of the ER in RNA translation.. 66
Figure 29: Homeostasis of Ca^{2+} in the ER...... 67
Figure 30: Structures and functions of the endoplasmic reticulum (ER)...... 69
Figure 31: Nanoscopic level of organization of the ER...... 70
Figure 32: ER-shaping determinants shown in either tubules or sheets...... 71
Figure 33: ER contact sites in mammalian cells. 72

Figure 34: Overview of mammalian ER protein quality control mechanisms...... 74
Figure 35: ER chaperones and foldases implicated in protein quality control. 76
Figure 36: Schematic representation of the three pathways of unfolded protein response (UPR^{ER}) and their downstream effects.. 77
Figure 37: Upstream open reading frames (uORFs) translation regulation of ATF4 under normal and stressed conditions.. 80
Figure 38: Distinct cell fate decisions observed under ER stress...... 81
Figure 39: ER stress-sensing mechanisms.. 83
Figure 40: Illustration of conditions causing lipid bilayer stress (LBS) and unfolded protein response (UPR^{ER}) activation.. 85
Figure 41: Proposed models for sensing lipid ER stress...... 86
Figure 42: Multiple regulatory checkpoints of the UPR^{ER}...... 88
Figure 43: Novel physiological roles of the UPR^{ER}. 91
Figure 44: Small molecules targeting specific UPR^{ER} signaling components.. 93
Figure 45: Overview of the RIBO-seq analysis workflow...... 96
Figure 46: Verification of EWSR1 KD of the ribosome profiling experiment. 97
Figure 47: Quality controls of the ribosome profiling experiment...... 98
Figure 48: DEG upon EWSR1 KD.. 100
Figure 49: GSEA. 101
Figure 50: EWSR1 KD leads to cell cycle arrest and decreased proliferation...... 102
Figure 51: EWSR1 represses the translation of a specific subset of mRNA.. 104
Figure 52: EWSR1 has no effect on global protein synthesis...... 105
Figure 53: EWSR1 KD has no effect on polysomes profiles. 106
Figure 54: Features of EWSR1 TE up targets.. 108
Figure 55: Sequence features of EWSR1 TE up targets are compatible with a model of direct binding of EWSR1...... 109
Figure 56: EWSR1 binds its translational targets...... 112
Figure 57: Tethering of EWSR1 represses translation of a Renilla luciferase mRNA reporter. 113
Figure 58: Repression of translation by EWSR1 does not rely on the miRNA machinery.. 114
Figure 59: Non-tethered FLAG-EWSR1 does not repress the translation...... 115
Figure 60: EWSR1 represses the translation via its CTD. 117
Figure 61: RGG2-ZnF-RGG3 is the region implicated in the repression of translation.. 120
Figure 62: EWSR1 is localized in the nucleus and the cytoplasm...... 122
Figure 63: EWSR1 is present in ribosomal fractions.. 124
Figure 64: EWSR1 binds to total and cytoplasmic mRNA.. 124
Figure 65: EWSR1 binds to the 40S subunit via its CTD. 126
Figure 66: EWSR1 binds to the 40S through RNase-insensitive interaction.. 128

Figure 67: RGG2-ZnF-RGG3 is the CTD region implicated in the interaction with the 40S.. 128
Figure 68: EWSR1 does not interact with the cap or initiation factors. 130
Figure 69: EWSR1 represses cap-independent translation...... 132
Figure 70: EWSR1 represses scanning-independent translation.. 132
Figure 71: P-site occupancy.. 133
Figure 72: GO analysis of EWSR1 TE targets. 135
Figure 73: Lipids classification into three main categories: storage (blue), membrane (green) and signaling (orange).. 136
Figure 74: Quality control of the lipidomic analysis.. 137
Figure 75: Accumulation of lipids upon EWSR1 KD.. 138
Figure 76: EWSR1 binds to mRNA targets associated with lipid metabolism...... 138
Figure 77: Changes of ER shape upon EWSR1 knockdown.. 141
Figure 78: EWSR1 knockdown impairs the activation of the PERK axis of the UPR^{ER}.. 142
Figure 79: EWSR1 promotes expression of ATF4 and its downstream targets.. 143
Figure 80: EWSR1 knockdown impairs the activation of ATF6 axis...... 144
Figure 81: EWSR1 knockdown impairs the activation of IRE1α axis.. 145
Figure 82: EWSR1 knockdown impairs the activation of UPR^{ER} in U2OS and A2780A cell lines...... 146
Figure 83: Knockdown of EWSR1 impairs the activation of the UPR^{ER} under ER stress.. 148
Figure 84: EWSR1 KD leads to increased apoptosis in PA treated condition..... 149
Figure 85: Illustration of the implication of EWSR1 in transcriptional and post-transcriptional regulation. 151
Figure 86: Tethering of FUS or TAF15 repress translation of the Renilla luciferase mRNA reporter.. 156
Figure 87: Non-tethered FLAG-FUS or FLAG-TAF15 do not repress the translation...... 156
Figure 88: Repression of translation by FUS or TAF15 does not rely on the miRNA machinery...... 157
Figure 89: FET proteins bind to ribosomal subunits.. 158
Figure 90: Ct-EWSR1 binds to ribosomal proteins.. 161
Figure 91: Tethering of EWSR1-FLI1 has no effect on the translation of the Renilla luciferase mRNA reporter.. 164
Figure 92: EWSR1-FLI1 does not bind to ribosomal proteins...... 164
Figure 93: EWSR1-FLI1 is not present in ribosomal fractions.. 165
Figure 94: Outcome of the normal or perturbated transcriptional and post-transcriptional mechanisms mediated by EWSR1.. 166
Figure 95: Illustration of main steps of mRNA translation regulation.. 167
Figure 96: Model illustrating the new molecular function of EWSR1 in the repression of translation identified in this work.. 169

Figure 97: Repression of translation by CTD does not rely on PRMT1...... 172
Figure 98: FET proteins bind to PRMT1.. 173
Figure 99: EWSR1 and CTD bind to PRMT1 through RNase-insensitive interaction..... 173
Figure 100: PRMT1 does not affect EWSR1 binding to the 40S...... 174
Figure 101: Knockdown of EWSR1 affects GCN2 axis of the ISR...... 178
Figure 102: EWSR1 KD leads to decreased apoptosis in AAD treated condition...... 179
Figure 103: Model illustrating the new physiological function of EWSR1 in maintaining cellular homeostasis identified in this work.. 180
Figure 104: Annotated plasmid map of the pDONR223 vector.. 201
Figure 105: Annotated plasmid map of the pDEST1899 vector with 3 x FLAG tag...... 202
Figure 106: Annotated plasmid map of the pN-MS2-CP vector...... 203
Figure 107: Annotated plasmid map of the pGEX-4T-1 vector...... 204
Figure 108: Annotated plasmid map of the pGEX-6P-1 vector...... 205
Figure 109: Annotated plasmid map of the pCIneo-TISU-RLuc-6xbsMS2 vector...... 206
Figure 110: Annotated plasmid map of the pCIneo-RLuc-6xbsMS2 vector...... 207

LIST OF TABLES

Table 1: FET proteins are highly related.. 7
Table 2: FUS, EWSR1 and TAF15 mutations identified in ALS, FTD and ET patients.. ... 14
Table 3: Some of EWSR1 gene fusions in cancer.. 21
Table 4: Antibodies used in this work with their source and host species. 193
Table 5: Chemicals used in this work with their source and identifier.. 195
Table 6: Cell lines used in this work with their source and identifier...... 195
Table 7: Palsmids used in this work with their source and identifier...... 196
Table 8: Plasmids sequencing primers. 197
Table 9: Cloning primers used in this work. 198
Table 10: RIP-qPCR primers used in this work...... 200

ABBREVIATIONS

(sno)RNP: small nucleolar RNP
2'-O-me: 2'-O-methylation
3D: three-dimensional
4E-BP: 4E-binding protein
4E-SE: eIF4E-sensitive element
A site: aminoacyl site
aa: aminoacyl
ABCE1: ATP-binding cassette E1
Acc2: acetyl-CoA carboxylases 2
AD: Alzheimer disease
AIP1: ASK1-interacting protein 1
ALS: amyotrophic lateral sclerosis
APH: amphipathic helix
ARE: AU rich element
Arl6IP1: ADP-ribosylation factor-like 6 interacting protein 1
AS: alternative splicing
ASD: autism spectrum disorder
ATAC-seq: transposase-accessible chromatin sequencing
ATCC: American Type Culture Collection
ATF6: activating transcription factor 6
b: bacterial
BAF: BRG1-associated-factor
BAK: BAX and/or BH antagonist or killer
BCL-2: B cell lymphoma 2
BDNF: brain-derived neurotrophic factor
BIBD: basophilic inclusion body disease
BID: BH3-only protein BH3-interacting domain death agonist
BIM: BCL-2-interacting mediator of cell death
BiP: binding immunoglobulin protein
BLI: bio-Layer interferometry
BLIMP1: B lymphocyte-induced maturation protein 1
BRCA1: breast cancer 1 gene
BTE: Barley Yellow Dwarf Virus (BYDV)-like element
bZIP: basic leucine zipper
C: carboxy
Ca^{2+}: calcium
CBP: cap-binding pocket
CBP: CREB-binding protein
CC: coiled coil
CDS: coding sequence
CER: ceramides
CFL1: cofilin 1
ChIP-seq: chromatin immunoprecipitation sequencing
CHOP: C/EBP-homologous protein
circRNA: circular RNA
CITE: cap independent translation element
CL: cloverleaf
CL: co-localization
CLB: cytoplasmic lysis buffer
CNPY2: canopy homolog 2
CNS: central nervous system
Cnx: calnexin
co-IP: co-immunoprecipitation
CP: central protuberance
CPE: cytoplasmic polyadenylation element
CPT: camptothecin
CRAC: Ca^{2+} release-activated channels
CrPv: cricket paralysis virus
Crt: calreticulin
CSFV: classical swine fever virus
CTD: C-terminal domain
D-loop: deoxyuridine loop
DBA: Diamond-Blackfan anemia
DCER: dihydroceramides
DDR: DNA damage response
DG: diacylglycerides
Dgat2: diacyl glycerol acyl transferase 2

DGCR8: DiGeorge Syndrome Critical Region 8
DHC: dihydroceramide
DHPR: dyhydropyridine receptor
DHR: degenerated hexapeptide repeat
DHS: dihydrosphingosine
DP: differential pattern
dsRBD: double-stranded RBD
DTEG: differential translation efficiency gene
DTG: differentially transcribed gene
e: eukaryote
EAD: EWSR1-activation domain
eCLIP: enhanced CLIP
EDEM: ER degradation-enhancing mannosidase
Edem: ER degradation-enhancing α-mannosidase-like protein
eEF1A: eukaryotic translation elongation factor A
eEF2: eukaryotic translation elongation factor 2
eIF3: eukaryotic initiation factor 3
eIF4A: eukaryotic initiation factor 4A
eIF4E: eukaryotic translation initiation factor 4E
eIF5A: eukaryotic initiation factor 5A
EM: electron microscopy
EMCV: encephalomyocarditis virus
EMSA: electrophoretic mobility shift assay
ENCODE: Encyclopedia of DNA Elements
ER: endoplasmic reticulum
ERAD: ER-associated degradation
ERdj: ER-localized DnaJ
ERES: ER Exit Sites
eRF1: eukaryotic release factor 1
eRF3: eukaryotic release factor 3
ERK: extracellular-signal-regulated kinase
Ero1: ER oxidase 1
ERSD: ER storage disease
ES: expansion segment
ESC: embryonic stem cells
ESFT: Ewing sarcoma family tumor
ESyt: extended synaptotagmin
ET: essential tremor
EVI: elongation velocity index
EwS: Ewing sarcoma
EWSAT1: Ewing sarcoma associated transcript 1
EWSR1: Ewing sarcoma breakpoint region 1
FA: fatty acid
FALS: familial-type ALS
FAM134B: family with sequence similarity 134, member B
FLuc: Firefly luciferase
FMDV: foot and mouth disease virus
FOXQ1: Forkhead Box Q1
FTD: frontotemporal dementia
FTLD: frontotemporal lobar degeneration
FUS: fused in liposarcoma
FXS: fragile X syndrome
FXYD6: FXYD domain containing ion transport regulator 6
GADD34: growth arrest and DNA damage-inducible 34
GALNT2: N-acetylgalactosamine transferase 2
GO: Gene Ontology
GRAMD: GRAM domain-containing protein
Grp78: 78-kDa glucose-regulated protein
GSEA: Gene Set Enrichment Analysis
GTP: guanosine triphosphate
HCC: hepatocellular carcinoma
HCV: hepatitis C virus
HD: Huntington’s disease
HDAC: histone deacetylase
HexCer: hexosylceramides
HLH: helix-loop-helix
hnRNP: heterogeneous nuclear ribonucleoprotein particle

Hox: Homeobox
HSC: hematopoietic stem cell
HSP: heat shock protein
HSP72: heat shock protein 72
IBD: Inflammatory Bowel Disease
IDP: intrinsically disordered protein
IDR: intrinsically disordered region
IGF-1: insulin-like growth factor 1
IGF-1R: insulin-like growth factor 1 receptor
IGR: intergenic region
IL-6: interleukin-6
INM: inner nuclear membrane
IP3R: inositol 1,4,5-trisphosphate (IP3) receptor
IPLS: immunoprecipitation low salt
IPMS: immunoprecipitation followed by mass spectrometry
IR: ionizing radiation
IRE: iron responsive element
IRE1: : inositol-requiring enzyme 1
IRF4: interferon regulatory factor 4
IRP1: iron-regulatory protein 1
IRP2: iron-regulatory protein 2
IRS1: insulin receptor substrate 1
ISR: integrated stress response
ISS: I-Shaped Structure
ITAF: IRES-trans-acting factor
JIK: JNK-inhibitory kinase
KD: knockdown
KH: K-homology domain
KO: knockout
KPNA2: karyopherin Alpha2
LAP: liver-enriched activating protein
LARP1: La-related protein 1
LBS: lipid bilayer stress
LC-ESI/MS/MS: liquid chromatography electrospray ionization tandem mass spectrometry
LC: low complexity
LD: lipid droplet
LD: luminal domain
LIP: liver-enriched inhibitory protein
LLO: lipid-linked oligosaccharide
LLPS: liquid-liquid phase separation
LMN: lower motor neuron
lncRNA: long ncRNA
LPC: lysophosphatidylcholine
LPE: lysophosphatidylethanolamine
LSI: Lipidomics Standards Initiative
LSU: large subunit
M: middle
$m^1acp^3\Psi$: 1-methyl-3-α-amino-α-carboxyl-propyl pseudouridine
m^6A: N6-Methyladenosine
m^7G: 7-methylguanylate structure
MALDI-TOF MS: matrix-assisted laser desorption/ionization–time-of-flight mass spectrometry
MAPPIT: mammalian protein–protein interaction trap
MCS: membrane contact site
MERCS: mitochondria-ER contact site
MHC: major histocompatibility complex
miRNA: microRNA
MIST1: muscle, intestine and stomach expression 1
MLV: murine leukaemia virus
MNK1: MAP kinase interacting Ser/Thr kinase 1
MNK2: MAP kinase interacting Ser/Thr kinase 2
MoMLV: Moloney murine leukemia virus
MoRF: molecular recognition feature
MRP: mitoribosomal protein
MS: multiple sclerosis
MS2-CP: MS2 coat protein
MSI: microsatellite instability
MST: microscale thermophoresis
mtDNA: mitochondrial DNA
mtLSU: mitochondrial large subunit

mTORC1: mechanistic target of rapamycin complex 1
mtSSU: mitochondrial small subunit
MYD88: myeloid differentiation primary response 88
MΦ: macrophage
N: amino-terminal
n.s.: not significant
NBD: nucleotide binding domain
NBD1: nucleotide-binding domain 1
NBD2: nucleotide-binding domain 2
NBS: nucleotide-binding site
NDRG2: N-myc downstream-regulated gene-2
NE: nuclear envelope
NEF: nucleotide exchange factor
NF-kB: nuclear factor kB
NF90: nuclear Factor 90
NIFID: neuronal intermediate filament inclusion disease
NLB: nuclear lysis buffer
NLS: nuclear localization signal
NMD: non-mediated decay
NOX2: NADPH oxidase 2
NT: nontreated
nt: nucleotide
NTD: N-terminal domain
ONM: outer nuclear membrane
ORP5: oxysterol-binding protein (OSBP)-related protein 5
ORP8: oxysterol-binding protein (OSBP)-related protein 8
OS9: osteosarcoma amplified 9
Os9: osteosarcoma amplified 9
OSBP: oxysterol-binding protein
p-SILAC: pulsed-SILAC
PA: palmitic acid
PABP: poly(A)-binding protein
PAR-CLIP: photoactivatable ribonucleoside enhanced cross-linking and immunoprecipitation
PARP: poly (ADP-ribose) polymerase
PB: processing body
PC-O: 1-alkyl,2-acylphosphatidylcholine
PC-P: -alkenyl,2-acylphosphatidylcholine
PC: phosphatidylcholine
PCA: Principal Component Analysis
PD: Parkinson disease
PD: pulldown
PDI: protein disulfide-isomerase
PE-O: 1-alkyl,2-acylphosphatidylethanolamine
PE-P: 1-alkenyl,2-acylphosphatidylethanolamine
PE: phosphatidylethanolamine
PERK: PKR-like ER kinase
PERK: protein kinase RNA-like ER kinase
PEV8: porcine teschovirus and porcine enterovirus 8
PG: phosphatidylglycerol
PGC-1α: peroxisome proliferator-activated receptor γ Coactivator
PI: phosphatidylinositol
PIC: pre-initiation complex
PK: proteinase K
PKA: protein kinase A
PKC: protein kinase C
PKR: protein kinase R
PL: phospholipid
PLA: proximity ligation assay
PM: plasma membrane
Pnc: promoter-associated noncoding
PNET: peripheral primitive neuroectodermal tumor
PNS: perinuclear space
post-TC: post-termination complex
POU domain: Pit-Oct-Unc domain
PP2A: protein phosphatase 2A

pre-rRNA: rRNA precursor
pre-TC: pre-termination complex
PrLD: prion-like domain
PRMT: protein arginine methyl transferase
PrxIV: peroxiredoxin IV
PS: phosphatidylserine
PSIV: Platia Stali instestine virus
PTC: peptidyl-transferase center
PTC: premature termination codon
PtdCho: phosphatidylcholine
PtdEtn: phosphatidylethanolamine
PtdIn: phosphatidylinositol
PTE: Panicum mosaic virus-like Translational Enhancer
PTGR: post-transcriptional gene regulation
PTM: post-translational modifications
PTP: permeability transition pore
PTP1B: protein Tyr Phosphatase 1B
PTR: programmed translational read-through
PV: poliovirus
RAP: ribosome affinity purification
RB: retinoblastoma
RBM3: RNA-binding protein 3
RBP: RNA binding proteins
RCP-seq: ribosome complex profiling-seq
RER: rough ER
RG4: RNA G-quadruplex
RGG: Arg-Gly-Gly
RHA: RNA helicase A
RHD: reticulon homology domain
RIBO-seq: ribosome profiling
RIDD: regulated IRE1-dependent decay
RIP: regulated intermembrane proteolysis
RIP: RNA immunoprecipitation
RLuc: Renilla luciferase
RNA-seq: RNA sequencing
RNAPII: RNA Polymerase II
RNAPII: RNA Polymerase II
RNC-seq: ribosome nascent-chain complex-bound RNA sequencing
ROS: reactive oxygen species
RP: ribosomal protein
RPF: ribosome protected fragment
RQC: ribosome-associated protein quality control
RRM: RNA recognition motif
rRNA: ribosomal RNA
rRNA: ribosomal RNA
RT: reverse transcriptase
Rtn: reticulon
RyR: ryanodine receptor
S: Svedberg
S/T: Serine/Threonine
S1P: site-1 protease
S2P: site-2 protease
SBD: substrate binding domain
Scd-1: stearoyl-CoA desaturase-1
SER: smooth ER
SERCA: sarcoendoplasmic reticular Ca^{2+} ATPases
SF1: splicing factor 1
SG: stress granule
sig: significant
SLBP: stem-loop binding protein
SLiM: short linear motif
SMN: survival motor neuron
sncRNA: small ncRNA
SNORA: H/ACA box snoRNAs
SNORD: C/D box snoRNAs
SQSTM1: sequestosome 1
SR: serine-arginine
SR: superresolution
SREBP: sterol regulatory element binding element
SRP: signal recognition particle
ssRNA: single strand RNA
SSU: small subunit

STED: Stimulated Emission Depletion Microscopy
SUnSET: SUrface SEnsing of Translation
SV2: simian virus 2
TAD: transcriptional activation domain
TAF: TBP-associated factor
TAF15: TATA binding associated factor 15
TBP: TATA-binding protein
TC: ternary complex
TCGA: The Cancer Genome Atlas
TCP-seq: translation complex profiling-seq
TE: translational efficiency
TED: Translation Enhancer Domain
TEM: transmission electron microscopy
TES: transcription end site
TF: transcription factor
TFIP11: tuftelin-interacting protein 11
Tg: thapsigargin
TIC: translation initiation complex
TIRAP: TIR domain-containing adaptor protein
TISU: translation initiator of short 5'UTR
TLR: Toll-like receptor
TLR2: Toll-Like Receptor 2
TMH: tandem transmembrane hairpin
TR: translation ratio
TRAF2: tumour necrosis factor receptor-associated factor 2
TRAP: translating RAP
Trn1: transportin 1
tRNA: transfer RNA
TSS: T-Shaped Structure
TSS: transcription start site
TSV: Taura syndrome virus
Tu: tunicamycin
TUDCA: tauroursodeoxycholic acid
TWJ: Three-Way Junction
u: universal
UCHL1: ubiquitin carboxy-terminal hydrolase L1
UGGT: UDP-glucose:glycoprotein glucosyltransferase
uORF: upstream open reading frame
UPP: ubiquitin-proteasome pathway
UPR^{ER}: unfolded protein response
UV: Ultraviolet
V-loop: variable loop
VCP: valosin containing protein
VDC/IE: vincristine, doxorubicin, cyclophosphamide/ifosfamide and etoposide
VEGFA: vascular endothelial growth factor-A
WFS1: Wolfram syndrome 1
XBP1: X-box binding protein 1
XBP1s: spliced XBP1
XTP3-B: XTP3- transactivated gene B protein
Y2H: yeast two hybrid
YBX1: Y-box binding protein 1
YSS: Y-Shaped Structure
YTHDF3: YTH domain family protein 3
ZnF: zinc finger
Ψ: pseudouridylation

1 INTRODUCTION

1.1 The EWSR1 protein

The Ewing sarcoma breakpoint region 1 protein (EWSR1) is a multifunctional protein involved in several steps of DNA and RNA processing and is implicated in maintaining various physiological mechanisms. Structurally, it can be divided in two main regions: the N-terminal domain (NTD) and the C-terminal domain (CTD), dictating its diverse functions. In addition to its implication in the regulation of multiple layers of gene expression, EWSR1 has gained broad attention due its association with neurodegenerative diseases and cancer, notably through its involvement in oncogenic gene fusions. The best characterized fusion, EWSR1-FLI1, is responsible for Ewing sarcoma (EwS), an aggressive tumor of adolescents and young adults. Despite the extensive study of the functions of oncogenes that trigger Ewing sarcoma, it is only recently that a growing number of studies is focusing on the characterization of EWSR1 itself in DNA and RNA processing.

In this chapter, we will start by presenting the FET protein family and its implication in the regulation of gene expression. Next, we will describe the structural organization of EWSR1. Then, we will discuss its involvement in the modulation of transcriptional and post-transcriptional mechanisms. Finally, we will review EWSR1 physiological roles and its association with neurodegenerative diseases, as well as with Ewing sarcoma.

1.1.1 FET proteins

1.1.1.1 An overview

The heterogeneous nuclear ribonucleoprotein particle (hnRNP) proteins are specific RNA binding proteins (RBPs) that impact several steps of nucleic acid metabolism, including transcription, splicing, nuclear-cytoplasmic transport, translation, and degradation[1]. They are generally present in the nucleus but also able to shuttle between the nucleus and the cytoplasm[2]. hnRNP proteins frequently undergo post-translational modifications, including methylation, phosphorylation, ubiquitination and sumoylation, important to regulate changes in their biological activity and subcellular localization[3]. Given their involvement at many stages of mRNA life cycle, it is very crucial to understand their roles in mRNA biogenesis and function.

A notable family of hnRNP proteins is the highly conserved FET protein family, which in vertebrates consists of FUS (fused in liposarcoma), EWSR1 (Ewing sarcoma breakpoint region 1) and TAF15 (TATA binding associated factor 15)[4]. FET proteins are DNA and RNA binding proteins involved in multiple steps of DNA and RNA processing[5]. They are expressed in almost all human fetal and adult tissues and are mainly located in the nucleus but can also translocate

to the cytoplasm[6]. They share a common domain organization, including an N-terminal transcriptional-activation domain (NTD), largely characterized in Ewing's family oncoproteins when fused to various transcription factors[7], and a C-terminal domain (CTD) implicated in different aspects of RNA biogenesis[8]. The CTD contains an RNA recognition motif (RRM), the most conserved region within the FET protein family, a zinc finger (ZnF) domain, RGG-rich domains and a nuclear import and retention signal (nuclear localization signal, or NLS)[9] (**Figure 1**). Although FETs share similarities in their organization and overall sequences, they present specific non-redundant biological and pathological functions. The molecular basis for this specialization is largely unknown but can be explained in part by sequence variations. For instance, TAF15 NTD has a significant negative charge in comparison with EWSR1 and FUS. It has also unique RGG boxes within prevalent YGGDR(G/S)G repeats that are unable to repress transcription due to the conserved aspartate residues[10]. Another layer of specialization could be explained by a divergence in methylation patterns, which might allow differential interactions between FET members and some of their protein partners. Indeed, the aspartate residue in YGGDR(S/G)G repeats of TAF15 confers poor binding to the protein arginine methyl transferases PRMT1 and might contribute to its biological specialization[10]. Interestingly, PRMTs such as PRMT3 and PRMT8 interact only with specific RGG-boxes in EWSR1[11,12]. These observations highlight the important impact of arginine methylation on the regulation of FET function and biology at post-translational level.

1.1.1.2 FET proteins and gene expression regulation

FET proteins are involved in several aspects of gene regulation[13] and interact directly or indirectly with numerous other proteins[14]. The unique qualities of FET family members to bind both DNA and RNA suggest that they function as "nonclassical" transcription factors. Indeed, FET proteins interact with transcription pre-initiation complex (RNA Polymerase II (RNAPII) and TFIID complex)[15] and with several gene-specific transcription factors, suggesting a gene-specific regulatory role in the regulation of gene expression[16]. In addition to transcription, they also affect splicing of pre-mRNAs and can interact with various splicing factors[17–19]. For instance, EWSR1 can regulate alternative splicing (AS) of genes important for DNA damage[20] and FUS binds the alternatively spliced exons of genes associated with neurodegeneration[21].
Early studies have documented important roles of FET proteins in miRNAs processing[8,22] and their ability to co-immunoprecipitate with the Microprocessor[23]. For example, FUS is recruited to chromatin to facilitate Drosha loading and can affect the biogenesis of miRNAs implicated in neuronal function, activity, and differentiation[24].

The fact that FET proteins cycle in and out of the nucleus raises the possibility of their implication in RNA transport. Accordingly, it was reported that EWSR1 can induce the nuclear retention of CFL1 mRNA[25]. Nucleocytoplasmic shuttling of RBPs can be regulated at multiple levels, including post-translational modifications such as phosphorylation, ubiquitination, and arginine methylation[26]. Indeed, it has been reported that arginine methylation and tyrosine

phosphorylation near PY-NLS sequence regulate FET proteins localization[27,28]. Further, it has been suggested that FUS can carry roles in translation repression through its association with polyribosomes[29], as well as in the regulation of mRNA stability [30].
Finally, several studies demonstrated that FET members are implicated in DNA damage repair. Recently, it has been shown that FUS interacts with numerous factors implicated in both replication-dependent and -independent pathways of DNA repair[31], and EWSR1 is important for removing PARP1 from damaged chromatin to ensure cellular and organismal survival[32].

Taken together, these studies emphasize the intimate implication of FET proteins in most aspects of gene expression regulation to maintain various biological functions. Consequently, it is unsurprising that they contribute to various human pathologies, as will be further discussed below.

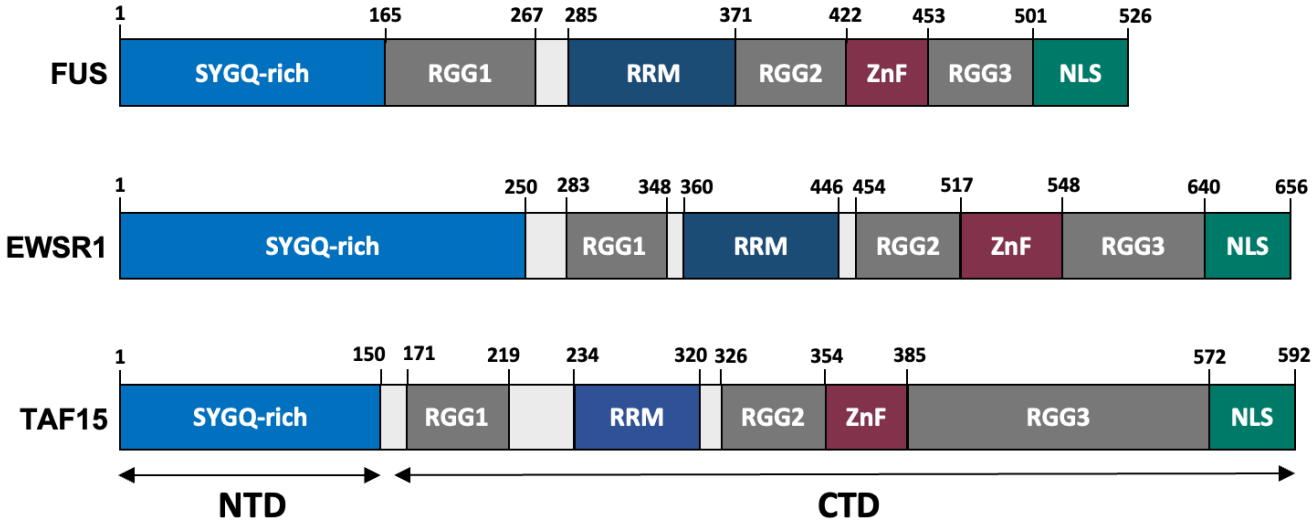


Figure 1: Domain organization of FET proteins. The N-terminal domain (NTD) consists of a SYGQ-rich domain. The C-terminal domain (CTD) is composed of RRM domain, ZnF domain, RGG boxes and a nuclear localization signal (NLS).

1.1.1.3 Self-assembly of FET proteins

FET proteins are intrinsically disordered proteins (IDPs) that are able to oligomerize to form higher-order self-assemblies[9]. IDPs are functional proteins that can be found in all organisms and are characterized by a biased amino acid composition, polyvalency, low sequence complexity and inability to fold into stable secondary and/or tertiary structures[33]. IDPs play an important role in modulating signaling pathways and cellular processes[34]. They contain intrinsically disordered regions (IDRs)[33] which can be classified into molecular recognition features (MoRFs), short linear motifs (SLiMs) and low complexity (LC) sequences[35]. The latter is rich in amino acids having a charged or polar side group, as well as glycine and proline, but

poor in hydrophobes. LC domains with specific enrichment in uncharged polar residues are termed prion-like domains (PrLDs)[36,37] (**BOX1**).

The ability of FET proteins to form fibrous assemblies arises from their LC and RGG domains. Particularly, FET NTD belongs to the LC class and consists of degenerated hexapeptide repeats (DHRs), defined by the prion-like SYGQQS motif. The aromaticity of the LC NTD is essential for the transcriptional and transforming activity of FET proteins[38]. Intrinsic disorder also plays a critical role in "liquid-liquid phase separation" (LLPS) or condensation phenomena (**BOX1**). In particular, LLPS underlies the formation of membrane-less organelles (*e.g.*, Cajal bodies, processing bodies (PBs), nucleolus, stress granules (SGs), centrosomes, and aggresomes)[39]. Phase separation of FET proteins can be dynamically modulated by various post-transcriptional modifications including tyrosine phosphorylation[40], O-GlcNAcylation[41], arginine methylation[42] and arginine citrullination[43]. Under physiological conditions, phase separation is a reversible and critical process that underlines FET molecular functions. For instance, fibers of FUS proteins bind RNA Pol II through its CTD and mutations in FUS LC domain abolish its ability to form fibers, but also to promote transcription *in vivo*[44] (**Figure 2**). In stress conditions, FET phase transition play an important role in the formation of cytosolic membrane-less RNA granules[9,45].

BOX1: Some Definitions

- **Prion-like domains (PrLDs):** PrLDs are domains of low complexity enriched in uncharged polar amino acids and glycine that assemble into higher-order structures by adopting beta-sheet rich conformations. They play a key role in regulating the solubility and folding state of proteins and present a common feature of many human disorders such as Alzheimer's and Parkinson's disease[46].

- **Liquid-liquid phase separation (LLPS):** LLPS is a thermodynamically driven, reversible phenomenon that consists in de-mixing into two distinct liquid phases, with distinct solute concentrations. The equilibrium between mixing and de-mixing depends strongly on the component concentrations, temperature, pressure, pH, and crowding agents. LLPS preferentially involves IDPs/IDRs because of their peculiar conformational properties, however, it has also been observed for ordered, globular proteins[39].

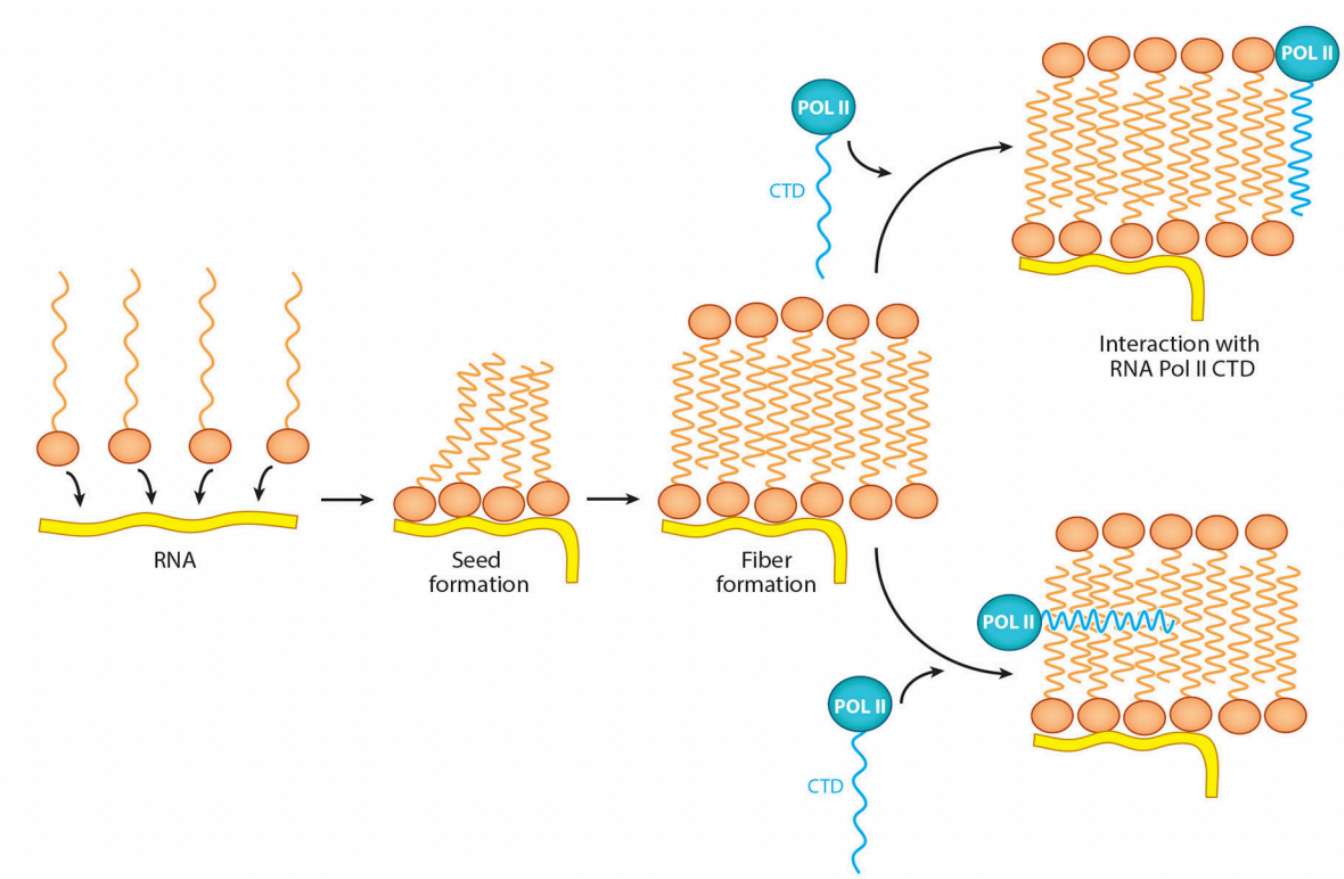


Figure 2: Model for RNA-nucleated assembly of FUS proteins and subsequent recruitment of RNA polymerase II (Pol II). FUS proteins bind to RNA. This complex forms the seed for fiber growth. FUS fibers can be formed by a seed of FUS proteins bound or not to RNA. The CTD of RNA Pol II interacts with this fiber either by intercalating into the growing fiber or by binding alongside the fibrous structure. From[9].

1.1.2 Domain structure and function of EWSR1

1.1.2.1 EWSR1 domain structural organization

The EWSR1 gene spans a ~ 40 kb region on chromosome 22q12.2 with 17 exons interrupted by 16 introns. The gene encodes a 656-amino acid protein and is composed of several domains. The first 7 exons in the N-terminal region encode the EWSR1-activation domain (EAD), and exons 8 to 17 encode the C-terminal domain (CTD)[47]. *EWSR1* is expressed in several isoforms. In addition to EWSα and EWSβ isoforms expressed in Ewing sarcoma family tumor (ESFT), a splicing variant (EWS-b) completely lacking exons 8 and 9 and a brain-specific variant have been reported[48]. EWSR1 is ubiquitously expressed in most human cell types, excluding cardiac muscle cells and melanocytes[6], highlighting its importance in differentiation, development and cellular homeostasis.

1.1.2.2 EWSR1 is a DNA and RNA binding protein

Little is known about EWSR1 nucleic acid recognition specificity. So far, it has been shown that EWSR1 has DNA-binding abilities and binds preferentially to G-quadruplexes with longer loops[49]. Interestingly, the first EWSR1 ChIP-seq showed its association with a subset of actively transcribed genes, often downstream the poly(A)-signal[50]. Most of these genes were involved in RNA regulatory processes.

Consistently with its role as an RNA-binding protein, PAR-CLIP analysis has shown that EWSR1 associates with 3' splice sites in pre-mRNA, preferentially in both G-rich and AU-rich sequences[51]. However, a separate CLIP study in HeLa cells showed enriched RNA binding near EWSR1-regulated exons 5' splice sites [52]. Another group performed EWSR1 CLIP-seq libraries and noted prevalent interaction with many expressed pri-miRNAs, suggesting an important role in enhancing pri-miRNA processing[53].

1.1.2.3 The EAD and CTD

The EAD contains 30 copies of the degenerate hexapeptide repeat with the consensus sequence SYGQQS, in which tyrosines are always conserved[38]. In addition, the EAD contains two autoregulatory regions, IQ and SF1 (splicing factor 1). The IQ domain of EWSR1 is phosphorylated by protein kinase C (PKC) and interacts with calmodulin. Interestingly, PKC phosphorylation of EWSR1 inhibits its binding with its RNA targets, suggesting that the IQ domain may provide a regulatory link between Ca^{2+} signal transduction pathways and RNA processing[54]. The association with SF1 inhibits the EAD and negatively modulates the target genes of EWSR1[55].
Progress in understanding structure/function relations for EAD has been limited by its IDP properties[56], DHR sequence which has not been amenable to mutagenesis and lack of information regarding its cognate partners[38]. However, functional studies of EWSR1 gene fusions, where EAD is perceived as a transcriptional activator, have provided a better acknowledgement of its sequence-function relationship. In concordance with its transactivation activity, the EAD can interact with transcriptional components including the RNA polymerase II subunits, rpb7[57] and rpb5[58], the coactivator CBP[59] and multiple TBP-associated factors (TAFs)[58]. However, its activation function is significantly repressed by FET RGG-boxes[60] *via* intramolecular masking of the EAD[56].

On the other side, the C-terminal domain (CTD) contains an 87-amino-acid RNA recognition motif (RRM), a ZnF motif, three Arg-Gly-Gly (RGG)-rich regions (RGGs) and a nuclear localization signal (NLS)[9] (**Figure 1**). All FET proteins present a ZnF domain with four cysteines coordinating the zinc ion. The RNA-recognition motif (RRM) is the most abundant and conserved RNA-binding domain in higher vertebrates (**Table 1**). Notably, the RRM of FET proteins differs from other hnRNP RRMs because of the extended "KK-loop" between α1 and β2 and the lack

of two aromatic amino acids on β3[9]. The arginine/glycine-rich (RGG) domains, which are intrinsically disordered, present the second common RNA-binding domain (RBD) in the human genome, after the RRM[61]. They play diverse functional roles and influence numerous physiological processes. Therefore, they have been associated with several diseases including neurological and neuromuscular diseases and cancer[62]. FET proteins were shown to bind nucleic acids through the RRM and the RGG2-ZnF-RGG3 motifs. Particularly, the RGG2-ZnF-RGG3 domain of FUS and EWSR1 possesses significant affinity for RNA and ssDNA[9]. Although the individual specific role of each RGG has not been extensively studied, it has been reported that RGG motifs allow self-association in EWSR1[63].

Remarkably, EWSR1 binds with G-quadruplex DNA and RNA *via* Arg residues of its RGGs[49]. In addition, these RGG motifs are targets for methylation by protein arginine methyl transferases (PRMTs)[64]. Indeed, pulldown experiments showed that endogenous EWSR1 protein is a strong substrate for PRMT1 but less for PRMT3[11]. Another report showed that arginine residues of RGG3 are intensively methylated by PRMT8[65]. As mentioned above, EWSR1 engages many transcripts, suggesting “promiscuous” or non-specific binding. Interestingly, a study conducted by *Ozdilek et al.*[66] shed new lights on RGG domains binding features. They showed that FUS RGG/RG domains alone have substantial RNA binding activity as opposed to the RRM and ZnF domains, and that when coupled to adjacent domains (ZnF/RRM), they can achieve affinities approaching that of the full length FUS. They also added that RGG domains are not wholly indiscriminate, as they might prefer specific sequences, such as GC-rich sequences; therefore, RGG domains present a certain degree of flexibility in their recognition of RNA sequences and structures. This addresses new questions about the specificity of EWSR1 RGG domains, which will be important to understand the cellular functions of target RNAs by characterizing their interactions with EWSR1.

Total	55	66	54		49	52	41		
	NTD	55	55	50		49	45	39	% Identity
% Similarity		CTD	70	59	78		63	47	56
			RRM	91	87	85		84	60	59

FUS-TAF15 FUS-EWSR1 EWSR1-TAF15

Table 1: FET proteins are highly related. Amino acid similarity and identity compared over the whole protein (total) and within the NTD, CTD and RRM for FUS, EWSR1 and TAF15. NTD: N-terminal domain, CTD: C-terminal domain, RRM: RNA recognition motif.

1.1.3 EWSR1 functions in DNA and RNA metabolism

1.1.3.1 EWSR1 and transcription

Several studies documented the association of EWSR1 with integral components of the transcriptional pre-initiation complex. Specifically, EWSR1 interacts with the C-terminal domain of RNA Pol II[44] and with different subunits of TFIID complex[58]. These interactions associate EWSR1 to early transcriptional initiation and elongation regulation.
Some reports have linked EWSR1 with several gene-specific transcription factors (activators and repressors) suggesting a more specialized, gene-specific regulatory role in modulating gene expression. Indeed, EWSR1 binds the family of transcription factors containing the POU (Pit-Oct-Unc) domains, widely involved in regulating vertebrate development[67]. For instance, EWSR1 binds Oct-4, a transcriptional regulator of genes involved in maintaining the undifferentiated pluripotent state, through its POU domain and stimulates its transcriptional activity[68]. EWSR1 can also interact with Brn-3a and blocks Brn-3a-mediated activation of Bcl-x promoter, proposing a transcriptional repressor role of EWSR1[69].

Another layer of EWSR1 transcriptional regulation is revealed by its interaction with major chromatin-modifying enzymes such as BRG1-associated-factor (BAF) ATP-dependent chromatin remodeling complex[70] and CBP/p300 proteins. CBP/p300 are large nuclear phospho-proteins that function as cofactors to various transcriptional factors and participate in many physiological processes including proliferation, differentiation, and apoptosis[71]. Intriguingly, EWSR1 notably forms a complex with this coactivator and activates *c-fos*, *Xvent-2*, and *Erb2* promoters, suggesting a potential role of EWSR1 in development[72]. Another study has shown that EWSR1 interacts with CBP and with the hypo-phosphorylated RNA polymerase II to enhance HNF4-mediated transcription[73], suggesting that EWSR1 may function as a co-activator of CBP-dependent transcription factors.

Taken together, these data provide several insights about EWSR1 association with both general and specialized factors for the modulation of transcription. The implication of EWSR1 in transcriptional regulation underlines its importance in maintaining physiological homeostasis, as transcription is a very critical biological process and a principal target of most of signaling pathways in human disorders.

1.1.3.2 EWSR1 and DNA repair and genome surveillance

Although DNA is highly susceptible to chemical modifications by endogenous and exogenous agents, cells are equipped with sophisticated systems to preserve its genomic sequence information, by instigating robust DNA damage response (DDR) pathways[74]. Environmental factors such as ionizing radiation (IR), Ultraviolet (UV), environmental mutagen and metabolic products represent the exogenous sources of DNA damage. On the other hand, replication

errors, DNA base mismatches, spontaneous base deamination and oxidative DNA damage represent endogenous sources of DNA damage[74]. Importantly, it is becoming clear that both transcriptional and post-transcriptional layers of gene expression regulation are involved in stress responses[75].

Many reports have shown the implication of EWSR1 in DNA repair and genome surveillance pathways. Indeed, EWSR1 knockout in mice resulted in disrupted B-cell development, decreased meiotic recombination, extreme sensitivity to ionizing radiation, and premature senescence of embryonic fibroblasts[76]. Additionally, two genetic screens determined *EWSR1* as a gene required for resistance to ionizing radiations (IR)[77], and to the topoisomerase I inhibitor, camptothecin (CPT)[78]. Another study revealed that EWSR1 and EWSR1-FLI1 become phosphorylated at Thr79 in response to either mitogens or DNA-damaging agents[79], highlighting the involvement of EWSR1 in genome stability.

A major control of the expression of critical genes implicated in the DDR is played by mechanisms regulating their splicing profile. This splicing reprogramming can depend on the expression and activity of specific RNA binding proteins (RBPs)[75]. Interestingly, several studies have associated genotoxic stress with EWSR1-mediated regulation of alternative splicing. Indeed, EWSR1 regulates alternative splicing of several genes implicated in DNA repair and genotoxic stress signaling, including *CHEK2*, *ABL1* and *MAP4K2*[20]. Another report indicates that genotoxic stress induces a disruption in the interaction between EWSR1 and the spliceosome-associated factor, YB-1, contributing to co-transcriptional *MDM2* exon skipping[80]. This suggests that genotoxic stress impairs the activity of EWSR1 in coupling transcription to splicing.

To ensure normal mitosis, checkpoints have evolved to induce cell-cycle arrest and detect defects that may have arisen during the steps leading to mitosis. Thus, defective mitosis is a hallmark of human cancer[81]. Strikingly, mounting evidence showed the implication of EWSR1 in checkpoint control and mitosis progression. Firstly, knockdown of EWSR1 orthologs in zebrafish resulted in mitotic defects followed by p53-dependent apoptosis, and embryonic lethality[82]. Secondly, EWSR1 interacts with the BARD1[83], a component of the BRCA1/BARD1 complex[84], which may couple EWSR1 to DNA repair and possibly also to checkpoint control. Thirdly, EWSR1 recruits Aurora B kinase to the midzone through its RGG3 domain[85] and regulates mitotic progression by affecting microtubule acetylation through HDAC6[86]. Interestingly, the *EWSR1* gene was identified as a novel MSI (microsatellite instability) target locus, showing important sensitivity and specificity in detecting mismatch repair-deficient cancers[87]. Accordingly, these results provide a functional connection between EWSR1 functions in DNA repair and cell cycle progression.

Altogether, these results highlight the key role of EWSR1 in homologous recombination, DNA damage response, and maintenance of genome integrity.

1.1.3.3 Splicing

Splicing is the process of removing intronic regions and joining exons. It is catalyzed by the spliceosome, which is a multi-megadalton ribonucleoprotein complex, characterized by a highly dynamic conformation and composition, and is assembled from the U1, U2, U5, and U4/U6 snRNPs and numerous non-snRNP proteins[88].

Increasing data have shown the implication of EWSR1 in the splicing regulation of specific transcripts. Initially, EWSR1 was shown to interact with the branchpoint recognizing protein BBP/SF1[55]. Subsequent studies have highlighted multiple interactions between EWSR1 and the splicing machinery: 1) enhanced mass spectrometric tools identified EWSR1 as a co-purifying protein with splicing complexes assembled on pre-mRNAs[89], 2) yeast two-hybrid screens revealed that EWSR1 interacts with U1C (a component of the U1snRNP, important in the early stages of spliceosome formation)[90] and with tuftelin-interacting protein 11 (TFIP11)[91], involved in alternative splicing regulation[92], 3) EWSR1 was also able to associate with a variety of splicing factors such as SR, TASR-1, TASR-2, and YB1[93,94]. Based on RNA-seq analysis, EWSR1 was shown to alter splicing for many gene products involved in DNA repair and genotoxic stress signaling (*ABL1, CHEK2, and MAP4K2*)[20]. EWSR1 also recruits U1snRNP and U2AF to the flanking splice sites of FAS/CD95 exon 6, which enhances exon inclusion and therefore FAS-mediated apoptosis[52]. Additionally, EWSR1 showed co-localization to speckled nuclear domains[95] and with survival motor neuron protein (SMN), a component of Cajal bodies involved in snRNP biogenesis[96]. These co-localization studies suggest an interesting contribution of EWSR1 in RNA splicing events.

In eukaryotes, transcription is functionally coupled to splicing[97]. Two main mechanisms have been proposed to explain the nature of coupling between transcription and alternative splicing: a) recruitment coupling, by recruitment of splicing factors by the transcription machinery and b) kinetic coupling directed by changes in RNAP II elongation rate that influence alternative splicing choices[98]. Based on current data, EWSR1 proteins probably function as an adapter molecule coupling transcription to RNA splicing for multiple genes. On one hand, it binds to hyper-phosphorylated RNAP II through its EAD and recruits serine-arginine (SR) splicing factors through the RNA-binding domain[94]. Thus, it affects the recruitment coupling mechanism. On the other hand, it has been shown that EWSR1 regulates Cyclin D1 transcripts transcriptionally and by modulating its splicing through the alteration of RNAP II dynamics (phosphorylation and speed) over the *CCND1* gene[99]. Finally, as described above (**see section "1.1.3.2. DNA repair and genome surveillance"**), many reports have also linked genotoxic stress with EWSR1-mediated regulation of alternative splicing (AS).

Altogether, these results underscore the critical implication of EWSR1 in splicing and in connecting transcription to splicing events.

1.1.3.4 EWSR1 and processing of non-coding RNA (ncRNA)

ncRNAs analysis has become a frontier research trend because of their participation in multiple biological and developmental processes[100], and their tightly association with chromosome-rearranged cancers[101]. Based on their sizes, ncRNA can be divided into two categories: small ncRNAs (sncRNAs) and long ncRNAs (lncRNAs). MicroRNAs (miRNAs) are small noncoding RNAs of ~ 22 nucleotides (nt) in length and play important roles in regulating gene expression[102]. The canonical biogenesis pathway of miRNAs-processing requires the RNA binding protein DiGeorge Syndrome Critical Region 8 (DGCR8) and the ribonuclease III enzyme, Drosha[102].

Growing evidence are linking EWSR1 to ncRNA processing. Interestingly, EWSR1 was identified in the complex containing Drosha[23]. EWSR1 deficiency results in elevated Drosha expression, upregulation of miR-29b and miR-18b and decreased expression of Col4a1 and CTGF, which contributed to the deregulation of dermal development in EWSR1 KO mice[22]. Moreover, EWSR1 deficiency leads to increased level of *Mir125a* and *Mir351*, which directly target *Uvrag* resulting in the deregulation of autophagy[103]. EWSR1 can also affect the maturation of let-7g (the first phylogenetically conserved miRNA)[104]. Indeed, EWSR1 depletion in osteosarcoma U2OS cells results in the accumulation of precursor let-7g and down-regulation of mature let-7g[105]. Several studies also showed that EWSR1 regulates pri-miRNA processing *via* distinct mechanisms[53], enhances the recruitment of Drosha to chromatin to regulate miRNA biogenesis[106] and modulates miRNA processing in the spinal cord[107].

lncRNAs are cellular RNAs longer than 200 nucleotides in length, implicated in several regulatory roles and exhibiting various disease associations[108,109]. Although they can interact with other biomolecules including DNA, RNA, and RBPs to modulate their activities, most lncRNAs have unknown or poorly understood functions. Several reports have connected EWSR1 to lncRNA. First, the carboxy-terminal RGG domain of EWSR1 can bind to TERRA, a lncRNA component of telomeric heterochromatin, through the G-quadruplex RNA in a structure-specific manner, leading to transcriptional regulation[110]. Second, a binding site for EWSR1 on the housekeeping ncRNA, 7SL, has been detected, suggesting that EWSR1 may aid in the displacement of HuR and further the downregulation of the tumor suppressor, TP53[111]. Third, a study has shown that primordial, primary and small antral follicles express several lncRNAs and high expression of *FUS, TAF15,* and *EWSR1* components of the paraspeckles, suggesting potential functions of FET proteins in oocyte physiology and development[112].

These findings illustrate the emerging concept of a EWSR1-ncRNA regulatory network in maintaining important and vital cellular functions.

1.1.3.5 EWSR1 and RNA transport and translation

In addition to its localization in the nucleus and cytosol, EWSR1 is exposed on the cell surface of different cells, where its arginine residues are extensively and asymmetrically dimethylated. These findings suggest that EWSR1 shuttles between nucleus, cytoplasm and cell surface and can act as a carrier for export of RNA constituents[64]. Furthermore, EWSR1 sediments to ribosomal fractions when dissociated from Pyk2 protein kinase[113], raising the possibility of its implication in RNA transport and translation. EWSR1 accumulates in the nucleolus[114], a key organelle implicated in ribosomal RNA (rRNA) transcription, rRNA processing and ribosome subunit assembly[115]. In this context, it was reported that EWSR1 is an interacting protein with nucleolar Karyopherin Alpha2 (KPNA2) protein, demonstrated to be crucial for cell growth related to rRNA and protein synthesis under starvation conditions[116]. Moreover, electron microscopy of U2OS osteosarcoma cells with siRNA-mediated depletion of EWSR1 showed disrupted nucleoli[117]. This may suggest a potential role of EWSR1 in rRNA processing and ribosome biogenesis.

Stress granules (SGs) are important players in the modulation of translation. They are dynamic membraneless cytoplasmic RNA granules, formed in response to various stresses and containing translationally stalled mRNAs, multiple RNA-binding proteins (RBPs) and initiation factors. They are cell type- and stress-specific and participate in various biological functions, thus they are involved in many human diseases like cancers, neurodegenerative diseases and viral infections[118]. Emerging evidence has shown that EWSR1 localizes to stress granules (SGs) upon environmental stress[119] and that it is expressed in the cytoplasm of secretory cell types[6]. This suggests potential activities in translational regulation. Moreover, two studies indicated that EWSR1 binds the 3'UTR of PRAS40[120] and cofilin 1 (CFL1)[6] mRNAs and suppresses their protein expression, proposing that it can modulate either mRNA nuclear export or translation, or both. It is very important to mention that many reports have demonstrated the implication of FUS in translational regulation[121], NMD (non-mediated decay)[122] and mTOR-dependent inhibition of translation[29], which might encourage research regarding EWSR1 function in translational regulation, as both proteins are very related.

1.1.4 Physiological roles of EWSR1

The previously described functions of EWSR1 in DNA and RNA metabolism and processing point towards an essential role for the protein in many physiological processes[123]. Indeed, insights on the *in vivo* importance of EWSR1 have been mostly determined by the generation of *EWSR1* knockout (KO) murine models[76]. Homozygous knockout mice were born smaller than their littermates and remained small through weaning[76]. Detailed analysis of surviving mice showed that loss of EWSR1 led to a defect in pre-B lymphocyte development (specifically during the pro-B to pre-B cell transition) and deficient gametogenesis by affecting the formation of XY bivalent and meiotic crossovers[76]. Moreover, EWSR1 loss leads to 1) infertility in both sexes, 2)

development of small testis and ovary compared to littermate wild type (WT) mice and 3) premature cellular senescence that could be explained by decreased lamin A/C expression and function, which may lead to nuclear lamina dysfunction[76]. EWSR1 is also crucial for determining brown fat lineage during development, by binding with YBX1 (Y-box binding protein 1) to activate the transcription of *Bmp7*[124], the earliest known brown adipogenic determination factor. Interestingly, EWSR1 modulates the stability of PGC-1α (Peroxisome proliferator-activated receptor γ Coactivator) protein, thus it regulates mitochondria function and cellular energy homeostasis[125]. Other studies showed that EWSR1 has an important role in maintaining the hematopoietic stem cell lineage[126], and it also regulates autophagy *via* an epigenetic modulation of the level of autophagy-promoter UVRAG[127].

Much knowledge about the role of EWSR1 is derived from studies of oncogenic fusion proteins arising from chromosomal translocations with multiple partners. Besides these oncogenic translocations, an independent role for EWSR1 in cancer is under investigations. Given its above-reported functions, it is unsurprising that EWSR1 is crucial to preserve the genome integrity and protect cells from neoplastic transformation. Indeed, as described in the previous section ("**1.1.3.2. DNA repair and genome surveillance"**), EWSR1 plays a critical role in the DDR and is essential for a proper response to genotoxic agents. It regulates the alternative splicing of genes implicated in oncogenesis, among others *BRCA1* (breast cancer 1 gene: a tumor suppressor)[128], and *ABL1*[20]. Intriguingly, transcriptional properties of EWSR1 in human cancers could be regulated through mitogenic signaling of oncogenic STRAP (a WD40 domain-containing protein). EWSR1 and STRAP are up-regulated in colorectal and lung cancers, and their cooperation block EWSR1-induced p300-mediated transactivation[129]. These data suggest a potential involvement of EWSR1 in tumor progression. Finally, mutations in EWSR1 have been demonstrated to be associated with several neurodegenerative disorders including amyotrophic lateral sclerosis (ALS), and frontotemporal dementia (FTD) (**see below**).
Overall, these observations highlight the role of EWSR1 in diverse physiological functions.

1.1.5 FET proteins and neurodegenerative disorders

Neurodegenerative disorders such as Alzheimer disease (AD), frontotemporal dementia (FTD), amyotrophic lateral sclerosis (ALS), Parkinson disease (PD), Huntington's disease (HD), and multiple sclerosis (MS), are characterized by progressive loss of selectively vulnerable populations of neurons and have variable clinical symptoms and pathological features[130].

Several studies identified mutations causative of neurological disorders in the genes encoding the FET family of proteins, documenting their involvement in the pathology of ALS, FTD, essential tremor (ET) and other neurodegenerative diseases (**Table 2**), where they have been found in cytoplasmic aggregates[131]. The specific mechanism leading to their cytoplasmic mislocalization is under investigation, but some studies are pointing to nucleo-cytoplasmic transport defects[132], and precise post-translational modifications affecting FET proteins in

neurodegenerative disease[28]. Notably, bioinformatics analysis revealed that FET proteins, contain predicted "prion-like" domains[133]. Therefore, when mutated these aggregation-prone FET proteins can form misfolded protein assemblies and act as seeds of aggregation leading to the disease phenotype. This mechanism resembles the replication of infectious prions and is often termed "prion-like"[134].

FET / N.P.	FUS			EWSR1	TAF15	Ref.
ALS	G515C E516V P18S S57del N63S S96del S115N S142N G144-Y149del G156E N159Y G173-G174del G174-G175del G187S G191S G206S Q210H R216C G222-	G223insG G223del G225V G226S G228-G229insG G230delG G230C R234C/L R244C G245V M245V R383C G399V S462F M462F M464I G464I G466VfsX14 R522G R524R/S/T/W	H517D/P/Q R518G/K/del G472VfsX57 R487C R491C G492EfsX527 R495X R495EfsX527 R495QfsX527 G497AfsX527 R502WfsX15 G503WfsX12 G504WfsX12 G507D G509D K510WfsX517 K510E/R S513P R514G/S P525L X527YestX Q519X R512S/G/C/H/L	G511A P522L	A31T M368T R385H G391E R395Q R408C G452E G473E	123
FTD	p.P106L	p.Gly174-Gly175 deletion GG	p.M254V	n.d.	n.d.	87,122,124
ET	p.Q290X	p.M392I	p.R377W	p.R471C	n.d.	119

Table 2: FUS, EWSR1 and TAF15 mutations identified in ALS, FTD and ET patients. del = deletion; ins = insertion; fs = frameshift; X = stop, n.d. = not defined, N.P. = neurological pathologies. ALS: amyotrophic lateral sclerosis, FTD: frontotemporal dementia, ET: essential tremor.

1.1.5.1 Amyotrophic lateral sclerosis (ALS)

Amyotrophic lateral sclerosis (ALS), first described in 1869, by the French neurologist Jean-Martin Charcot, is a fatal motor neuron disorder affecting motor neurons with an incidence of about 1/100,000. It is characterized by progressive loss of the upper and lower motor neurons (LMNs) at the spinal or bulbar level[135].

Over 20 genes have been associated with ALS. *C9orf72*, *SOD1*, *TARDBP* and *FUS* have been reported to be major ALS-related genes[136,137]. More than 50 autosomal dominant *FUS* variants have now been identified in ALS patients (**Table 2**). The majority of these variants are missense

mutations, but also insertions, deletions, splicing, and nonsense mutations have been reported in rare cases[138,139]. Several ALS-associated FUS mutations occur within the nuclear localization signal (NLS) and lead to the redistribution of FUS to the cytoplasm[131]. A more general role for FET proteins in neurodegeneration has been supported by the presence of genetic variants in *TAF15* and *EWSR1* genes in a small number of ALS patients. Notably, none of ALS-FUS cases showed co-accumulation of EWSR1 and TAF15 in the FUS-positive pathological inclusions, moreover, their cellular distribution was not altered[119]. Genomic sequencing in ALS-individuals, of exons (15-18) of *EWSR1* allowed the identification of 2 missense mutations located in the first (G511A) and second (P552L) RGG of EWSR1[123,140] (**Table 2**). Remarkably, transfection of wild type *EWSR1* into primary motor neurons cultured from rat embryos revealed that it is primarily localized to the nucleus, while its ALS-specific variants showed elevated cytoplasmic and neuritic accumulation[123].

1.1.5.2 Frontotemporal dementia (FTD)

Frontotemporal dementia (FTD), one of the most common neurodegenerative conditions after Alzheimer's Disease (AD), is a heterogeneous neurodegenerative disorder, characterized by a series of changes in several neurotransmitter systems, including serotonin, dopamine, GABA and, above all, glutamate. It is characterized by behavioral abnormalities, language impairment, and deficits of executive functions[141] and presents a strong genetic component, and autosomal dominant inheritance in 10%-25% of patients[142]. Selective degeneration of the frontal and temporal lobes is a common feature of FTD, therefore frontotemporal lobar degeneration (FTLD) is often used[143].

The neuropathology underlying clinical FTLD is subclassified based on the identity of the major protein that forms cellular inclusions. Therefore, in addition to FTLD-Tau or FTLD-TDP (with abnormal intracellular accumulation of either tau protein or the transactive response DNA-binding protein MW 43-TDP43, respectively), several other neuropathological depositions have been defined, including FTLD-FET[119], FTLD-UPS (with inclusions of proteins of the ubiquitin-proteasome system) and FTLD-ni (with no inclusions observed)[144]. Furthermore, other uncommon genetic mutations have been described, including valosin containing protein (VCP), sequestosome 1 (SQSTM1) and FUS mutations[145].

FTLD-FUS comprises cases with ubiquitin-positive inclusions (FTLD-U), neuronal intermediate filament inclusion disease (NIFID) and basophilic inclusion body disease (BIBD)[146]. A limited number of FUS mutations (p.P106L, p.Gly174-Gly175deletion GG, p.M254V) have been determined in FTLD patients without concomitant ALS[147] (**Table 2**). It was demonstrated that the pathological inclusions in all these FTLD subtypes also contain EWSR1 and TAF15 and transportin 1 (Trn1), the protein responsible for their nuclear import[119]. Although, it is still debated what causes FET protein translocation in FTLD, several studies have shown that FET species,

accumulating in FTLD, have undergone a common abnormal post-translational modification that may be related to the pathogenesis of FTLD-FET[119]. Hypomethylation of the arginine residues in the arginine-glycine rich region, immediately adjacent to the nuclear localization signal, results in abnormally strong binding of FET proteins to Trn1. This results in a molecular, indissociable and mislocalized complex that forms insoluble inclusions within the cell[28]. This would explain why the inclusions in FTLD-FET contain all of the FET proteins and Trn1, and also suggests that FET proteins are particularly sensitive to alteration of transportin-mediated nuclear import[119].

1.1.5.3 Essential tremor, Parkinson, and Alzheimer diseases

Essential tremor (ET) is a progressive, neurological disease characterized by postural and/or kinetic tremor[148]. Exome sequencing revealed 3 ET variants in FUS: 1) a stop mutation (p.Q290X), 2) (p.M392I) mutation identified as potential ET susceptibility factor and 3) a non-sense mutation, (p.R377W), in one patient with family history of disease. Additionally, a rare p.R471C substitution in a single subject with familial ET was identified after sequencing of the *EWSR1* gene[119] (**Table 2**). This suggests that FET proteins are implicated in ET pathology. Importantly, sequencing specific regions of FUS gene did not identify any novel non-synonymous variant affecting the subjects susceptibility to Parkinson's disease[149], and no mutations in FET genes have been reported in Alzheimer disease patients.

Interestingly, a recent research reported the first implication of EWSR1 in the central nervous system (CNS). Indeed, EWSR1 can play an important role in the neuronal morphology and motor function. Moreover, its deficiency leads to impairments of neuronal morphology, dopaminergic signaling pathways, and motor function in mice[150].

Collectively, these studies show that mutations in FET proteins alter their cellular distribution and lead to aberrant regulation of their expression and activity, contributing to disease onset and progression.

1.1.6 EWSR1 and sarcomas

1.1.6.1 Sarcomas

Sarcomas are rare malignant tumors of soft tissue and bone (World Health Organization classification)[151]. Soft tissue sarcomas can be formed in the body's muscles, joints, fat, nerves, deep skin tissues, and blood vessels, while bone tumors are found throughout the bones of the body and in the cartilage[152]. Sarcomas are a heterogeneous group of more than 100 cancer subtypes, comprising 1-2% of adult cancers worldwide. They represent 6-15% of childhood cancer (< 15 years) and 11% of adolescent and young adult cancers (15-29 years)[153].

Although the pathogenesis of many sarcomas is poorly understood, some sarcomas arise in patients with cancer predisposition syndromes. For instance, patients with Li Fraumeni syndrome, with germline mutations in TP53, have increased susceptibility for osteosarcoma and rhabdomyosarcoma[154], others with mutations in the retinoblastoma (RB) gene are at increased risk for osteosarcoma[155]. Importantly, about a third of sarcomas, such as Ewing sarcoma, arise from characteristic chromosomal translocations, leading to fusion proteins, which are often the primary driver of disease pathogenesis[151] (**see below**).

1.1.6.2 EWSR1 and Ewing sarcoma

- **Description of Ewing sarcoma:**

In 1921, James Ewing described a group of bone tumor notably different from that of osteogenic sarcoma[156], he called "diffuse endothelioma of bone", which was subsequently named Ewing sarcoma.

Ewing sarcoma (EwS) is the second most frequent type of bone tumor following osteosarcoma. EwS primary tumors can virtually originate from any anatomical sites but are located mainly in bones and soft tissues and arise preferentially in adolescents and young adults[157,158], with a reported incidence of 1-3 cases per million of population per year[159]. These tumors are often aggressive and may metastasize to additional areas of the body, such as lungs, bone and bone marrow[160,161]. It has been reported that four to six alternating cycles of VDC/IE (vincristine, doxorubicin, cyclophosphamide/ifosfamide and etoposide) have been used as the current chemotherapy for patients with localized and metastatic disease[162]. Unfortunately, even patients who do successfully recover, often suffer from late effects of therapy[163]. Therefore, it is very crucial to develop treatments that especially target the biological drivers of EwS[162,164].

Unlike other sarcomas presenting some lineage-specific differentiation, Ewing sarcoma histologically appeared as uniformly undifferentiated small round cells, containing an oval hyperchromatic and granulated nuclei and a narrow cytoplasm within a sparse intercellular stroma. When first reported by James Ewing, endothelial origin for the sarcoma was first proposed on the basis of its cellular morphology and the rareness of stroma[156]. Since then, different theories regarding ES progenitor[165] have been advanced, but increasing evidences have supported the theory of mesenchymal stem cells as putative cells of origin[166]. Specific candidate progenitors include bone marrow-derived mesenchymal stem cells[167], neural crest-derived stem cells[168], bone progenitors[169], and osteochondrogenic progenitors[170].

- **Genetics:**

Genetically, EwS is mainly characterized by *FET-ETS* gene fusions, associating a member of the *FET* family with one of the members of the *ETS* family of transcription factors (TFs) (*ERG, FLI1, ETV1*, *ETV4*, and *FEV*)[171–174] (**Figure 3a,b**), a family characterized by a conserved DNA-

binding domain that recognizes GGAA or GGAT motifs[175]. Particularly, *FET-ETS* gene fusions encode fusion proteins that retain the N-terminal transactivation low complexity region of FET proteins to the C-terminal DNA-binding domain of ETS TFs. Although, these fusions have long been considered to result from a balanced translocation (*i.e.*, the phenomenon of transposition of chromosomal material without the loss of genetic material)[176], recently, it was reported that *EWSR1*-derived gene fusions can also arise in the context of chromoplexy[177] (*i.e.,* the phenomenon of complex genome restructuring, characterized by multiple inter- and intra-chromosomal translocations and deletions[178]).

In ~85% of Ewing sarcoma (**Figure 3b**), a somatic reciprocal t(11;22)(q24;q12) chromosomal translocation results in a fusion gene, *EWSR1-FLI1,* involving *EWSR1* and friend leukemia virus integration 1 genes *(FLI1)*[179] (**Figure 3a**). Both of EWSR1 and FLI1 domains are required for the oncogenic function of EWSR1-FLI1[180] fusion, in which the weak N-terminal transcription activation domain of FLI1 is replaced by the stronger activation domain of EWSR1. The remaining 15% include *EWSR1* or *FUS* fused to other *ETS* members (*ERG, FEV, ETV1 and ETV4*)[171–174,181–185]. Recently, TAF15 was discovered to be implicated in EwS[186].

Besides *EWSR1-ETS* gene fusions, EWSR1 can also be fused to non-ETS proteins in the cases of "EwS-like" tumors (**Table 3***,* see[187–189] for detailed "EwS-like tumors" descriptions). Ewing-like sarcomas lack the pathognomonic molecular hallmark of Ewing sarcoma, they morphologically resemble classical EwS, and so far, have been classified in four main types: i) CIC-rearranged sarcomas, ii) BCOR-rearranged sarcomas, iii) sarcomas with a fusion between EWSR1 and a gene not belonging to the ETS family members, and iv) unclassified round cell sarcomas[189].

EWSR1-FLI1 includes a set of highly related isoforms or subtypes with many breakpoints occurring in a variety of introns of *EWSR1* and *FLI1* genes[190–192], with the fusion of *EWSR1* exons 1 to 7 and *FLI1* exons 6 to 9 being the most common isoform. A disruption of processing of *EWSR1-FLI1* fusion RNA alters the expression of the fusion protein[193]. CD99, a heavily O-glycosylated transmembrane, is a diagnostic marker for Ewing sarcoma that has been reported to affect migration, invasion and metastasis of tumor cells[194,195]. Recently, BCL11B and GLG1 were identified as high specific markers for *EWSR1-FLI1*-positive Ewing sarcoma[196].

- **EWSR1-FLI1 regulates epigenetic, transcriptional, and post-transcriptional mechanisms:**

Despite EwS remarkable paucity of somatic mutations (low mutational rate of 0.15 mutations/Mb compared to other malignant neoplasms[197]), it presents a strong propensity to progress and metastasize, suggesting powerful adaptation properties of EwS cancer cells. Various studies have endeavored to determine the role of EWSR1-FLI1 fusion protein in Ewing sarcoma oncogenesis[198,199]. Interestingly, *EWSR1-FLI1* gene fusion can transform mouse cells such as NIH3T3[200], C3H10T1/2[180] and primary mesenchymal stem cells[201]. Genomic approaches

showed that thousands of genes are dysregulated by EWSR1-FLI1 in Ewing's sarcoma cell lines, linking this fusion protein to a complex transcriptional program[202]. Indeed, EWSR1-FLI1 activates genes implicated in proliferation, cell differentiation, and cell survival and suppresses others implicated in apoptosis and cell cycle[203].

Recent data suggest that EwS-induced transformation is driven by reprogramming the epigenome (DNA methylation or histone marks)[204] and transcriptional heterogeneity[205]. Thus, these mechanisms are now recognized as critical oncogenic hubs in EwS tumorigenesis[206]. Strikingly, EWSR1-ETS fusions have been shown to impact the actions of polycomb group, histone deacetylase (HDAC) and nucleosome remodeling complexes that modulate gene expression by altering chromatin structure[207]. EWSR1-FLI1 has a major impact on chromatin states and remodeling events. It represses target genes by displacing wild type ETS factors from mesenchymal enhancers[208] and interacts with the BAF complex, which contributes to target gene activation[70]. Additionally, it is now emerging that EWSR1-FLI1 can actively remodel 3D chromatin architecture, creating oncogenic transcription hubs within super-enhancer regions[209].

Although, EWSR1-FLI1 is crucial for the growth and survival of Ewing sarcoma and transformation *in vivo* and *in vitro*[210], several lines of evidence suggested that alone it cannot fully explain the Ewing sarcomagenesis. This implies parallel pathways and cooperating mutations. Indeed, BMI-1 and Forkhead Box Q1 (FOXQ1) have been identified as novel cooperative transcription factors interacting with EWSR1-FLI1 to maintain the transformation[211,212]. In addition, many studies have identified insulin-like growth factor 1 (IGF-1) and its receptor (IGF-1R)[213] and CD99[214] (also called MIC2) as parallel pathways contributing to the oncogenesis in Ewing sarcoma. Additional reports indicated that EwS/PNET (peripheral primitive neuroectodermal tumor) show INK4A deletion[215] and p53 alteration[216], highlighting the importance of these parallel genetic alterations in the *in vivo* biologic behavior of EwS/PNET. It is also important to note that a part of EWSR1-FLI1 function comes from an interference with wild-type EWSR1 function, adding another layer to explain Ewing sarcomagenesis[82,217].

EWSR1-FLI1 not only binds to DNA, but also directly interacts with the transcriptional apparatus: RNA polymerase II[94], hsRBP7[57], CREB-binding protein (CBP)/p300[218], and RNA helicase A (RHA)[219]. Moreover, it was reported that EWSR1-FLI1 binds RNA[70], interacts with SF1 and U1C splicing factors[220,221] and affects alternative splicing[221]. Indeed, several studies demonstrate that EWSR1-FLI1 generates the cyclin D1b oncogenic isoform by modulating the rate of RNA polymerase activity[222] and leads to preferential splicing of *ARID1A-L* by modulating the splicing of ARID1A, a specific protein component of the BAF complex[223]. Furthermore, EWSR1-FLI1 modulates the expression of microRNAs (miRNAs) and long non-coding RNAs, which contributes to EwS tumorigenicity and therefore provides a potential therapeutic target[224]. For instance, lncRNA Ewing sarcoma associated transcript 1 (EWSAT1) has been identified as a downstream target of EWSR1-FLI1 that facilitates tumor progression[225], and another report showed that EWSAT1 is implicated in the proliferation and invasion of glioma by sponging miR-

152-3p[226]. Other studies identified high expression of pncCCND1_B RNA, a previously uncharacterized CCND1 promoter-associated noncoding (pnc) transcript[227] and HULC lncRNA in Ewing Sarcoma, which correlates with EwS aggressiveness[228]. Finally, multiple studies have connected EWSR1-FLI1 to DNA damage response, cell cycle checkpoint control[229,230] and R-loops formation[231].

Taken together, these findings illustrate the diversity and the complexity of EWSR1-FLI1 roles. They suggest a model whereby EWSR1-FLI1 might drive tumorigenesis by deregulating epigenetic processes and transcriptional and post-transcriptional programs. This stresses on the importance of investigating the precise molecular functions of EWSR1-FLI1, which might offer therapeutic opportunities for Ewing sarcomagenesis.

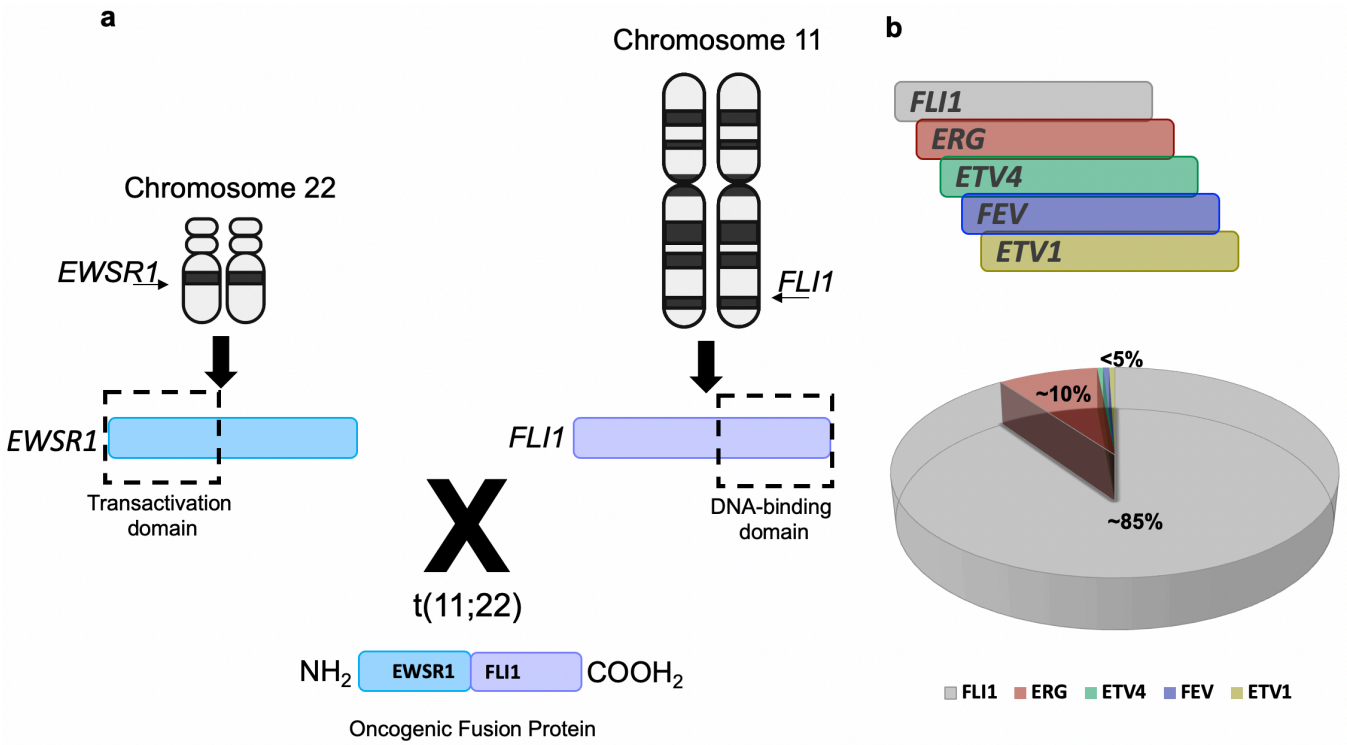


Figure 3: Schematic representation of the *EWSR1-FLI1* fusion gene. (**a**) The EWSR1-FLI1 rearrangement is caused by the fusion of the N-terminal transactivation domain of the *EWSR1* gene with the C-terminal DNA binding domain of the *FLI1* gene. Schemes of chromosomes 11 and 22 were created with BioRender.com. (**b**) About 85% of *EWSR1-ETS* fusions involve the *FLI1* gene. *ERG* is the next most common fusion partner accounting for about 10 % of cases. T: translocation.

Fusion Protein	TF partner	Translocation	Tumor type
EWSR1-FLI1	FLI1	t(11;22)(q24;q12)	Ewing sarcoma/ESFT
EWSR1-ERG	ERG	t(21;22)(q22;q12)	
EWSR1-ETV1	ETV1	t(7;22)(p22;q12)	
EWSR1-FEV	FEV	t(2;22)(q33;q12)	
EWSR1-ETV4	ETV4	t(17;22)(q21;q12)	
EWSR1–WT1	WT1	t(11;22)(p13;q12)	Desmoplastic small round cell tumor
EWSR1–ERG	ERG	t(21;22)(q22;q12)	
EWSR1–DDIT3	DDIT3	t(12;22)(q13;q12)	Myxoid liposarcoma
EWSR1–ATF1	ATF1	t(12;22)(q13;q12)	Clear cell sarcoma
EWSR1-CREB1	CREB1	t(2;22)(q33;q12)	
EWSR1–NR4A3	NR4A3	t(9;22)(q22;q12)	Extraskeletal myxoid chondrosarcoma
EWSR1–ATF1	ATF1	t(12;22)(q13;q12)	Angiomatoid fibrous histiocytoma
EWSR1-CREB1	CREB1	t(2;22)(q33;q12)	
EWSR1-NFATC2	NFATC2	t(20;22)(q13;q12)	ESFT-like
EWSR1-ZNF278	ZNF278	t(22;22)(q12;q12)	Askin-like, CD99 neg.
EWSR1-POU5F1	POU5F1	t(6;22)(p21;q12)	Myoepithelial tumor of soft tissue
EWSR1-PBX1	PXB1	t(1;22)(q23;q12)	
EWSR1–PBX3	PBX3	t(9;22)(q33.2;q12)	
EWSR1-ZNF444	ZNF444	t(19;22)(q13;q12)	
EWSR1-KLF17	KLF17	t(1;22)(p34.1;q12)	
EWSR1-ZNF384	ZNF384	t(12;22)(p13;q12)	Acute lymphoblastic leukemia
EWSR1-CREB3L1	CREB3L1	t(11;22)(p11;q12)	Low-grade fibromyxoid sarcoma/sclerosing epithelioid fibrosarcoma

Table 3: Some of EWSR1 gene fusions in cancer. TF: transcription factor, ESFT: Ewing sarcoma family tumor, t: translocation.

1.2 mRNA Translation

mRNA translation is a ubiquitous process that occurs across all prokaryotic and eukaryotic kingdoms. It consists in the translation of the genetic codons from processed and mature mRNA into polypeptide chains by ribosome translocation. It is conceptually divided into four stages, initiation, elongation, termination, and ribosome recycling, and commonly involves three main players: the mRNA (genetic template), the ribosome (assembly machinery), and the aminoacyl transfer RNAs (tRNAs).

In this chapter, we will start by presenting the translation as a highly regulated layer of the complex regulatory pathways of gene expression. Next, we will describe the different steps of cap dependent translation and their regulation by different mechanisms. Then, we will briefly discuss the alternative translation mechanisms. We will subsequently review diverse emerging concepts about ribosomes and their importance in the regulation of translation. Finally, we will cover the principles of the main translatomics methods implicated in the study of the translatome.

1.2.1 Gene expression: a highly regulated process

Gene expression is a highly complex process comprising many steps, each with elaborate regulation (**Figure 4**) to properly convert the encoded information in a gene into a functional product. It is regulated at many levels by a diversity of participating factors, such as transcription factors (TFs), RNA binding proteins (RBPs) and miRNAs. Indeed, the human genome encodes about 2,000 TFs[232], more than 1,542 RBPs[233] and approximately 2,600 mature miRNAs[234], which underlies the tremendous complexity of transcriptional and post-transcriptional regulations. Additional processes that add further complexity to gene regulation have been identified, including alternative pre-mRNA splicing, alternative translation initiation and termination. Moreover, once a protein is produced, it can further undergo post-translational modifications, including phosphorylation, acetylation, ubiquitination, and SUMOylation, affecting its interaction with other proteins and its catalytic activity to give rise to phenotypes[232]. Gene expression regulatory pathways are further complicated by feedback mechanisms and diverse coupling between individual processes. These are connections which have emerged as an economic strategy to respond robustly to environmental cues[235].

The emergence of high-throughput methods and next-generation sequencing in the past two decades have allowed tremendous discoveries in gene expression regulation. For example, chromatin immunoprecipitation followed by sequencing (ChIP-seq) and assay for transposase-accessible chromatin sequencing (ATAC-seq) are used at the genomic levels to understand the contextual behavior of DNA-binding proteins or epigenomic markers. On the other hand, next-generation RNA sequencing (RNA-seq) and mass spectrometry-based proteomics can respectively profile the transcriptome and the proteome of cells even at single-cell resolution[236].

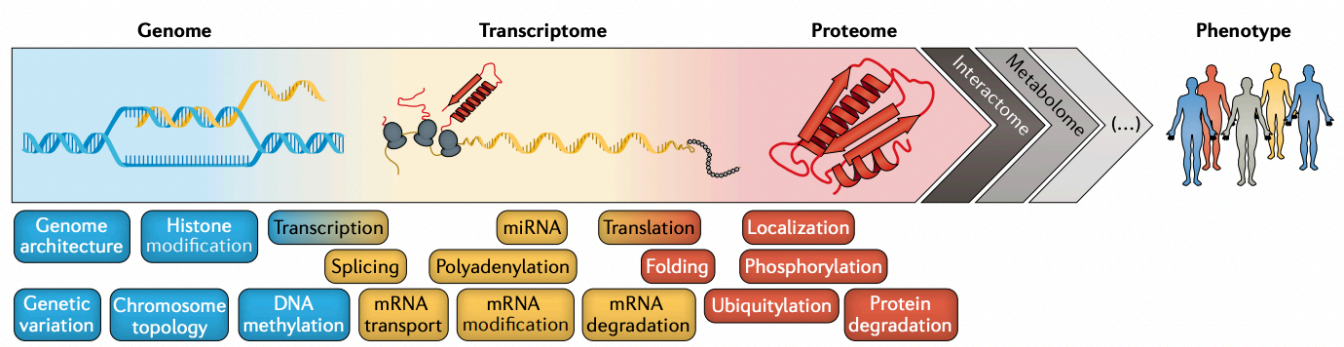


Figure 4: Overview of gene expression pathways. Processes conferring regulatory control are indicated at the bottom. From[236].

1.2.2 Cap-dependent translation in eukaryotes

1.2.2.1 Initiation step

Most cellular mRNAs are translated by cap-dependent mechanism. This canonical initiation of translation is dependent on the presence of both the 5'cap structure (7-methylguanylate structure, m^7G) and the 3'poly (A) tail on mRNAs and requires a plethora of eukaryotic initiation factors (eIFs) for the attachment of ribosomes on mRNA. The 5'cap allows the specific recruitment of a tripartite complex eIF4F, which comprises i) the eIF4E cap-binding protein, ii) an RNA helicase eIF4A and iii) a large platform protein eIF4G. This binding permits the further recruitment of the so-called 43S pre-initiation complex (PIC), comprising the small 40S ribosomal subunit, the ternary complex (TC) containing eIF2-GTP-Met-$tRNA^{Met}$, the small translation factors eIF1, eIF1A, and eIF5 and the multi-subunit eIF3 factor[237]. The recruitment of the 43S PIC to the 5'cap is possible through a protein-protein interaction between eIF4G and eIF3[238]. eIF4G also interacts with PABP, which promotes mRNA circularization[239]. However, mRNA circularization is not observed for all messenger mRNAs[240]. Then, the PIC undergoes a 5'→3' sliding that consumes ATP to reach the AUG start codon; this step is called "scanning". Since 5'UTRs often contain secondary structures that might block the scanning step, the RNA helicase eIF4A is employed to unwind them. The scanning process ends at AUG start codons, embedded in an optimal context GCC(A/G)CCAUGG called the Kozak sequence[241]; the complex is then called 48S PIC. After 48S complex formation, eIF5 and eIF5B promote the hydrolysis of eIF2-bound GTP thereby inducing the displacement of eIFs, which allows the joining of the 60S subunit and the formation of an elongation-competent 80S ribosome. The assembled 80S complex is now ready to accept the appropriate aminoacyl-tRNA into the A (aminoacyl) site of ribosome and synthesize the first peptide bond[237] (**Figure 5**).

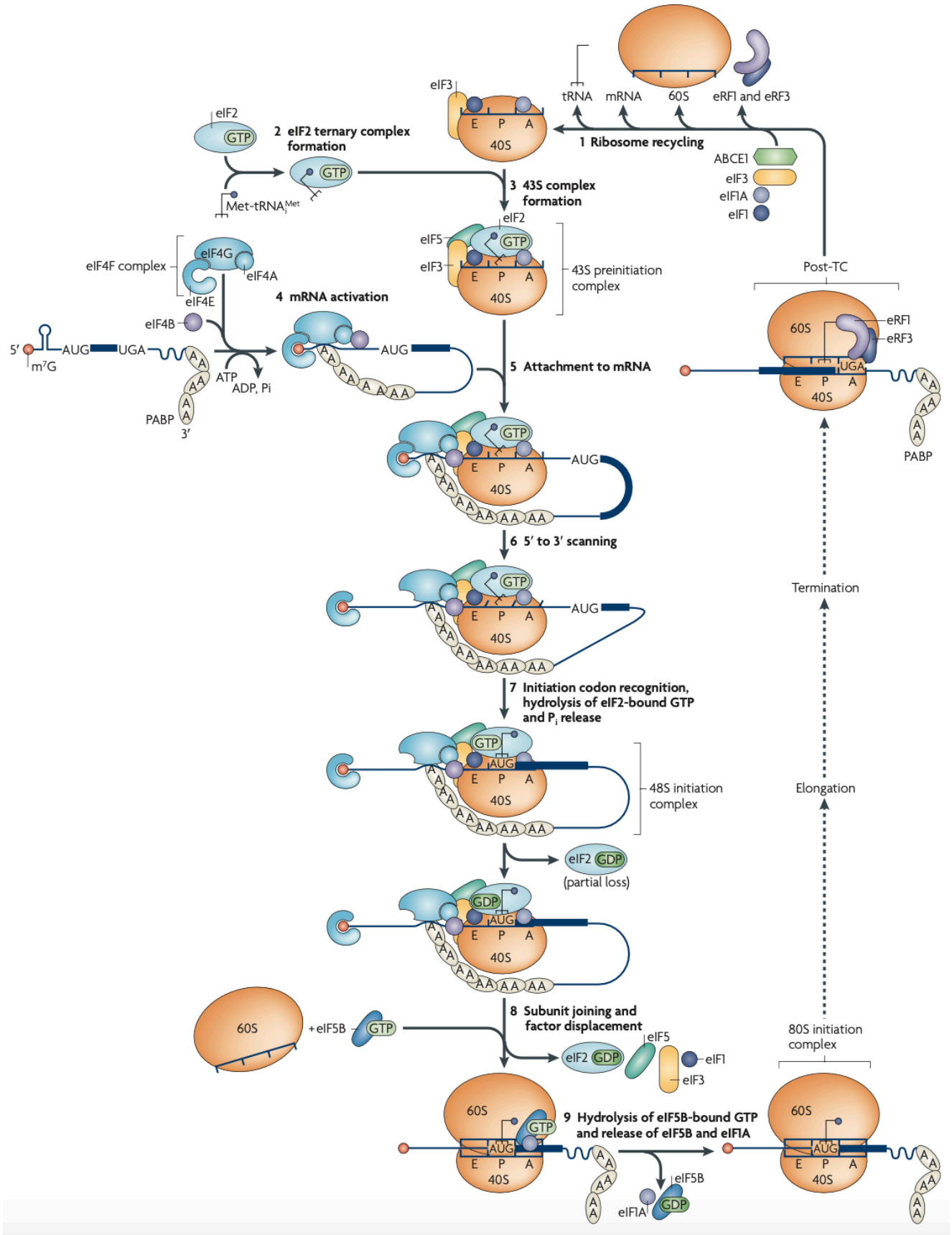

tRNA
mRNA
60S
eRF1 and eRF3
eIF3
E P A
40S
1 Ribosome recycling
ABCE1
eIF3
eIF1A
eIF1
eIF2
GTP
2 eIF2 ternary complex formation
Met-tRNA$_i^{Met}$
3 43S complex formation
eIF2
eIF5
eIF3
43S preinitiation complex
Post-TC
eIF4F complex
eIF4G
eIF4A
eIF4E
eIF4B
4 mRNA activation
5′
m^{7}G
AUG
UGA
ATP
ADP, Pi
PABP
3′
5 Attachment to mRNA
60S
eRF1
eRF3
UGA
6 5′ to 3′ scanning
Termination
7 Initiation codon recognition, hydrolysis of eIF2-bound GTP and P$_i$ release
48S initiation complex
eIF2 GDP
(partial loss)
Elongation
8 Subunit joining and factor displacement
+eIF5B
eIF5
eIF1
eIF3
80S initiation complex
9 Hydrolysis of eIF5B-bound GTP and release of eIF5B and eIF1A
eIF5B
eIF1A

Figure 5: Model of the canonical pathway of eukaryotic translation initiation, elongation, termination, and ribosome recycling. Post-TC: post-termination complex, m^7G: 7-methylguanylate structure. From[242].

1.2.2.2 Elongation step

Translation elongation is conserved in all kingdoms of life, with key steps shared between eukaryotes and bacteria. It is a multistep sequential process encompassing aminoacyl-tRNA (aa-tRNA selection), peptide bond formation, and translocation (*i.e.,* the process of shifting of mRNA-tRNA complexes within the ribosome). The aa-tRNA selection relies on the correct Watson-Crick base complementary between the codon of the mRNA in the A site and the anticodon of the incoming tRNA[243]. Interestingly, the ribosome undergoes a series of major structural rearrangements during translocation [244] (**see below**).

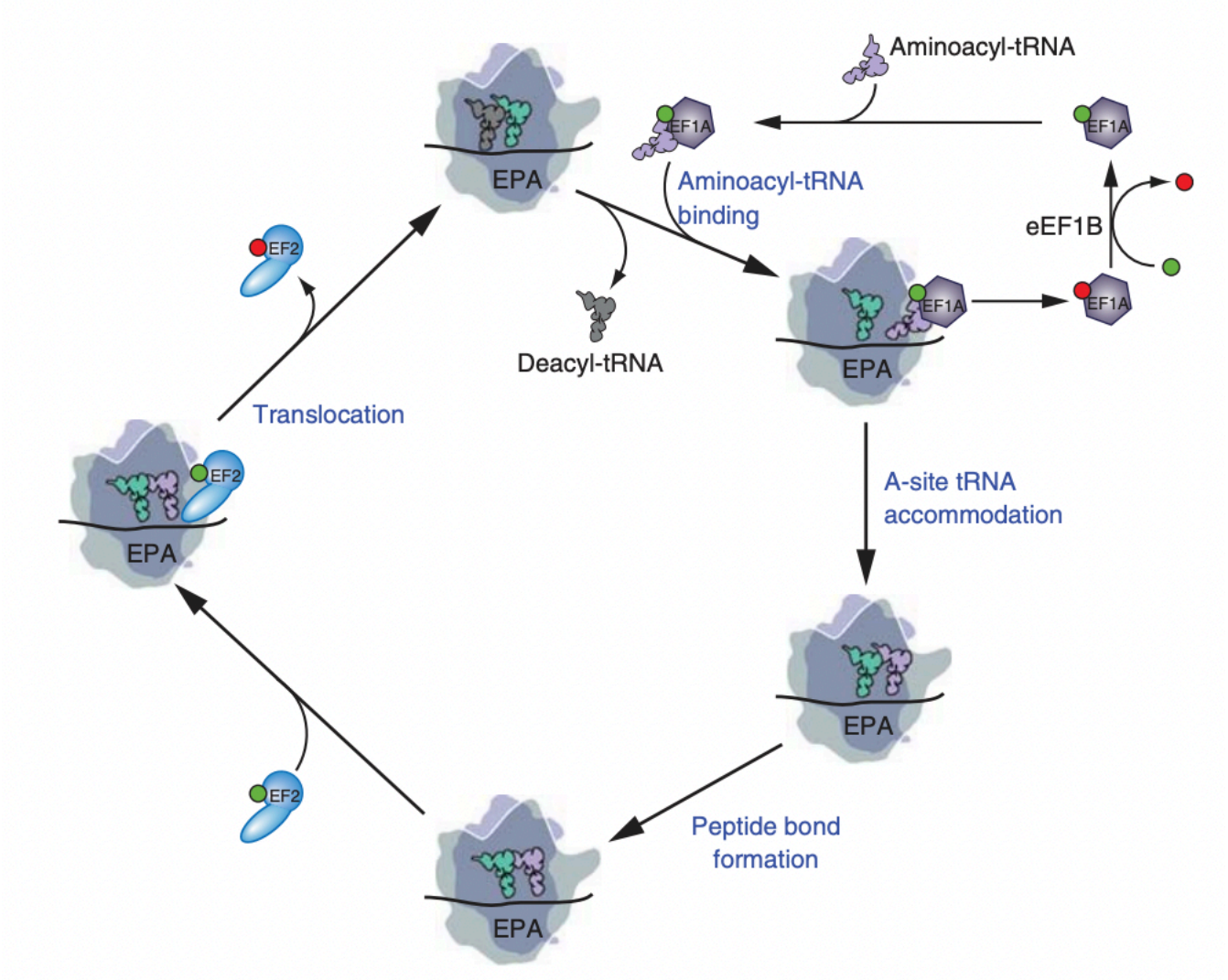


Figure 6: Model of eukaryotic translation elongation. The large ribosomal subunit is drawn transparent to allow the visualization of tRNAs, factors, and mRNA binding to the decoding center. Green ball represents GTP and red ball is GDP. E: exit site, P: peptidyl site, A: aminoacyl site. From[245].

Following translation initiation, the 80S eukaryotic ribosome is positioned on the mRNA, with the AUG start codon base-paired with the anticodon of $tRNA^{Met}$ in the P-site. Elongation commences with the delivery of the cognate elongating aminoacyl-tRNA to the decoding site (A site) of the ribosome with the eukaryotic translation elongation factor, eEF1A. eEF1A is activated upon binding to GTP to form eEF1A•GTP•aminoacyl-tRNA complex. Following GTP hydrolysis and release of an eEF1A•GDP binary complex, the aa-tRNA is accommodated into the A site, and the eEF1A•GDP is recycled to eEF1A•GTP by the exchange factor eEF1B. During peptide bond formation, the A- and P-site tRNAs shift from "classic" to "hybrid" states: the acceptors ends of tRNAs move to the P and E sites and the anticodon ends stay in their original A and P sites[245]. This rotated state of the ribosome undergoes further conformational changes in preparation for translocation[246]. Then, eEF2•GTP binds in the A-site and promotes translocation of the tRNAs into the canonical P and E sites. Finally, eEF2•GDP is released, and the next cycle of elongation commences with release of the deacylated tRNA from the E site[245] (**Figure 6**).

1.2.2.3 Termination step

Termination of mRNA translation occurs when the ribosome reaches the stop codon in the A-site. It is mediated by the eukaryotic release factors eRF1 and eRF3. eRF1 consists of 1) an amino-terminal domain (N) responsible for recognition of the stop codon in the A site[247], 2) a middle domain (M), with a conserved GGQ motif that induces the release of the nascent polypeptide from peptidyl-tRNA in the ribosomal P site[248], and 3) a carboxy-terminal domain (C) for binding to eRF3 and ABCE1, it also contains a mini-domain affecting stop codon specificity[249] (**Figure 7a**).

eRF3 consists of (1) C-terminal region comprising a canonical guanosine triphosphate (GTP)-binding domain (G) and two β-barrel domains (2 and 3), which have strong structural homology to GTP-binding translation factors such as EF-Tu, eEF1A, and the carboxy-terminal region of the ribosome rescue factor Hbs1, and 2) a non-conserved N-terminal domain, dispensable for termination, that binds to the poly(A)-binding protein (PABP) and UPF3b (the nonsense-mediated decay (NMD) factor)[250] (**Figure 7b**). Although, free eRF3 is unable to bind GTP because of the disorder of Switch I and Switch II regions (contained in its G domain), it undergoes conformational changes upon binding to eRF1 and gains its ability to bind GTP[251].

The pre-termination complex (pre-TC) contains peptidyl-tRNA in the P site. eRF1 enhances the eRF3 binding to GTP, leading to the formation of a stable eRF1/eRF3•GTP complex, which binds to the A site of the pre-TC, where eRF1 recognizes the stop codon. Then, the M domain of eRF1 dissociates from eRF3's Switch I/Switch II elements and accommodates in the peptidyl-transferase center (PTC), after GTP hydrolysis by eRF3. This promotes the peptide release. eRF1 and eRF3-GDP can remain bound with the post-termination complexes (post-TCs)[250] (**Figure 7c**).

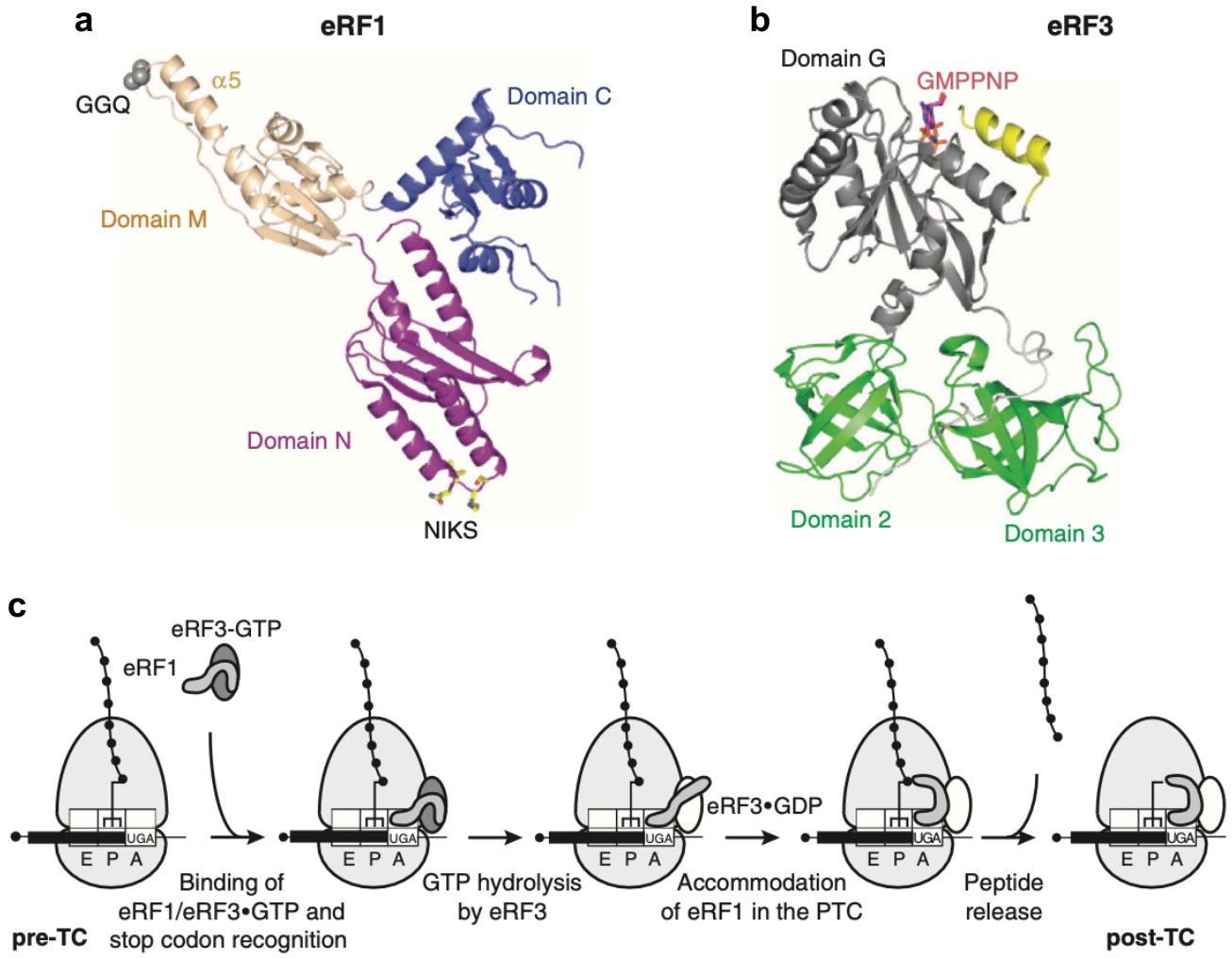


Figure 7: Translation termination in eukaryotes. Ribbon representations of (**a**) human eRF1 and (**b**) *Schizosaccharomyces pombe* eRF3. (**c**) Schematic representation of the different steps of translation termination process in eukaryotes. Post-TC: post-termination complex. From[250].

1.2.2.4 Recycling step

Post-termination ribosomes recycling is initiated by the conserved protein ATP-binding cassette E1 (ABCE1) and depends on the presence of eRF1 in the A site. ABCE1 protein consists of two nucleotide-binding domains (NBD1 and NBD2) which form sandwich-like nucleotide-binding sites (NBSs) for two ATP molecules, a helix-loop-helix (HLH) motif in NBD1 and a unique extra N-terminal region FeS domain containing two $[4Fe\text{-}4S]^{2+}$ clusters, connected to the NBD core by a characteristic antiparallel β-sheet called the cantilever arm along with a flexible linker, the cantilever hinge[250] (**Figure 8a,b**).

ABCE1 binds to eRF1 on the post-TC. Cycles of binding of ATP to the NBD sites in ABC proteins, hydrolysis of ATP and release of ADP, induce conformational changes in these proteins that are thought to generate power strokes, resulting in splitting the post-termination ribosome. This yields a 60S subunit, eRF1, ABCE1, and a 40S subunit bound to mRNA and deacylated tRNA[250] (**Figure 8c**).

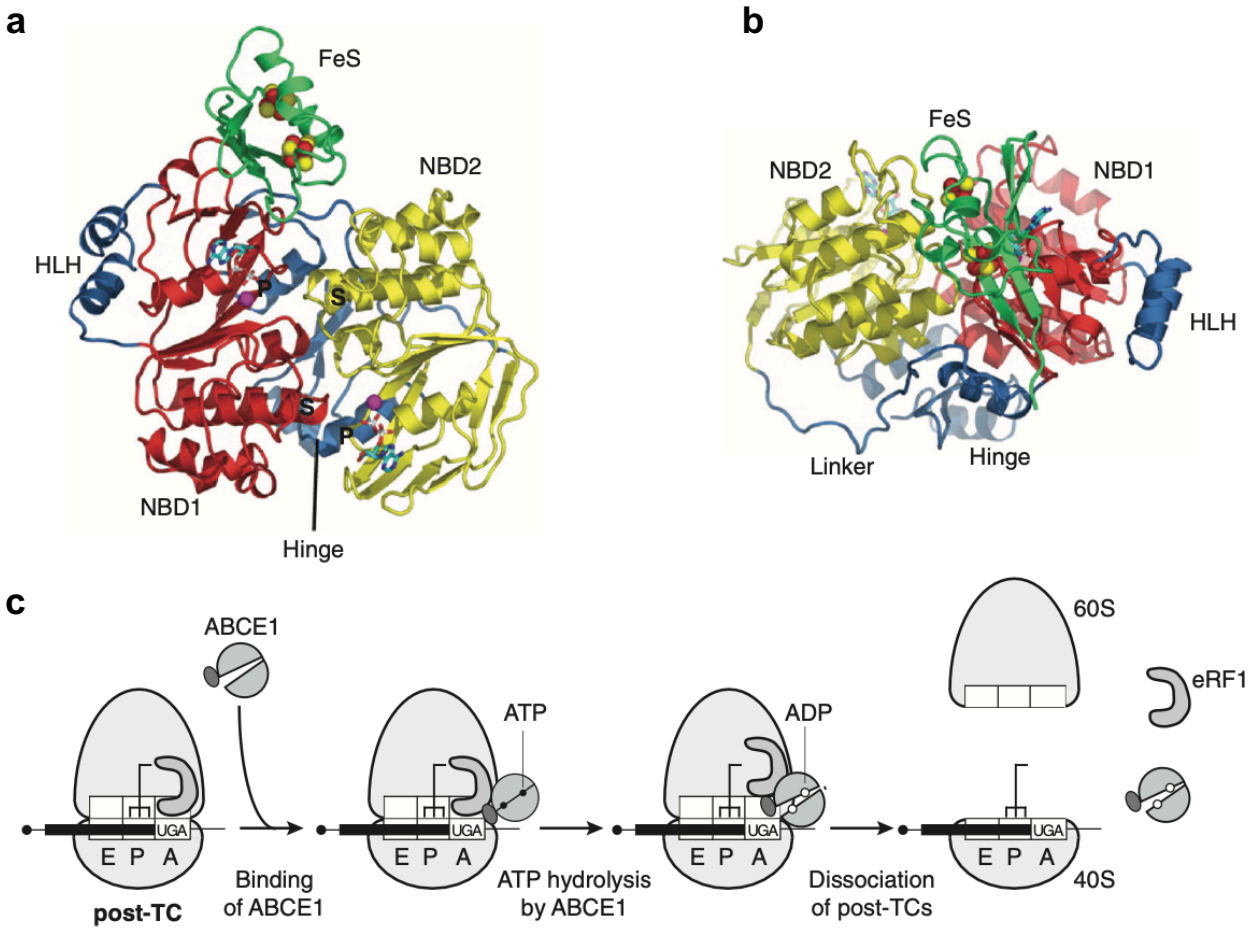


Figure 8: Translation recycling in eukaryotes. (**a, b**) Ribbon representations of ABCE1 from *Pyrococcus abyssi.* (**a**) top and (**b**) front views. (**c**) Schematic representation of the different steps of translation recycling process in eukaryotes. Post-TC: post-termination complex. From[250].

1.2.3 Regulation of translation steps

1.2.3.1 Regulation of translation initiation

- **Regulation of translation initiation by uORFs:**

Genome-wide sequencing of 5'UTRs reveals that ~50% of mammalian mRNAs, have at least one upstream open reading frame (uORF) fully residing within the 5'UTR or partially overlapping with the main coding sequence (CDS). Interestingly, uORFs are important regulators of translation initiation and mRNA level[252].

Under normal cellular conditions, uORFs typically reduce protein expression by 30-80%[253]. Several factors influence suppressive capacity of uORFs on CDS translation, such as the number and length of the uORFs, the distance between an uORF and the downstream CDS, and the start codon of uORF and its context[254]. Under stress conditions, the reduced availability of the ternary complex (TC) results in the scanning ribosomes to bypass inhibitory uORFs, which allows the translation of certain stress-responsive transcripts, such as ATF4[255]. In the other hand, some uORFs, such as uORF1 of GCN4, present a positive role in the translation of the downstream CDS. GCN4 is a transcriptional activator of amino acid biosynthetic genes. GCN4 mRNA translation is activated in response to low TC levels, through a re-initiation mechanism involving four uORFs (uORFs 1-4). Indeed, functional interaction between the N-terminal domain (NTD) of eIF3a and sequences 5' of uORF1 is important to promote scanning and re-initiation of the 40S ribosomal subunit after the termination of translation[256].

In addition, uORFs may direct start codon selection of the main CDS to generate different protein isoforms. This can be exemplified by *CEBPB*, which has four AUG initiation codons encoding three different protein isoforms (liver-enriched activating proteins (LAP* and LAP) and liver-enriched inhibitory protein (LIP)) and one uORF[257]. Therefore, uORF-mediated translation can determine cell fate through balancing the expression of protein isoforms.

- **Regulation of translation initiation by initiation factors activities:**

Regulation of translation initiation can also be mediated by the control of the availability of active eIF2 and eIF4F by reversible protein phosphorylation.

Phosphorylation of the α subunit of eukaryotic initiation factor eIF2 is a main mechanism for translational control, which causes a reduction in global protein synthesis with selective translation of specific genes including ATF4. Without this phosphorylation, eIF2 is incapable of forming the ternary complex. There are four different mammalian Serine/Threonine (S/T) eIF2α kinases, acting as early responders for different exogenous and endogenous stresses: 1) GCN2 (EIF2AK4), induced in response to nutritional stresses, 2) PKR-like ER kinase (PERK) activated by ER stress, 3) protein kinase R (PKR) (EIF2AK2) participating in antiviral defense pathway

involving interferon and 4) heme-regulated eIF2α kinase (HRI) (EIF2AK1) activated by heme deprivation in erythroid cells (**Figure 9**). These kinases have critical roles in alleviating environmental stress, as their dysfunctions are linked with pathologies in multiple organs. They constitute the sensors of the integrated stress response (ISR), an evolutionarily conserved intracellular signaling network activated to maintain cellular homeostasis[258].

Furthermore, the mRNA cap-binding protein, eukaryotic translation initiation factor 4E (eIF4E) of the EIF4F complex has a central role in the regulation of translation initiation. Its activity is regulated *via* two mechanisms. First, the translational repressor 4E-binding protein (4E-BP) inhibits the formation of the eIF4F complex, by binding to eIF4E, thus, preventing eIF4E-eIF4G interaction. When 4E-BP is phosphorylated by the mechanistic target of rapamycin complex 1 (mTORC1), 4E-BP/eIF4E interaction is reduced, which allows the formation of eIF4F complex at the mRNA cap and the stimulation of translation[259].
Second, eIF4E activity can also be regulated by phosphorylation of its sole phosphorylation site (Ser 209), by MAP kinase interacting Ser/Thr kinase 1 (MNK1) and MNK2, downstream of the extracellular-signal-regulated kinase (ERK) and the p38 MAPK signaling pathways. MNK1/2 are recruited to the eIF4F complex by binding eIF4G's C-terminus, where they phosphorylate eIF4E. Therefore, eIF4E integrates inputs from the mTOR and ERK signaling pathways for translational control linked to several cellular processes, including cell cycle progression, cell survival, cell motility and tumorigenesis[259].

Other eIFs (eIF1, eIF2β, several eIF3 subunits, eIF4G, eIF4B, eIF4H, eIF5 and eIF5B) have been reported to be phosphorylated. In many cases, they presented increased phosphorylation under conditions of activation of translation. However, there is no solid evidence that translation activation is caused by these phosphorylations[242].

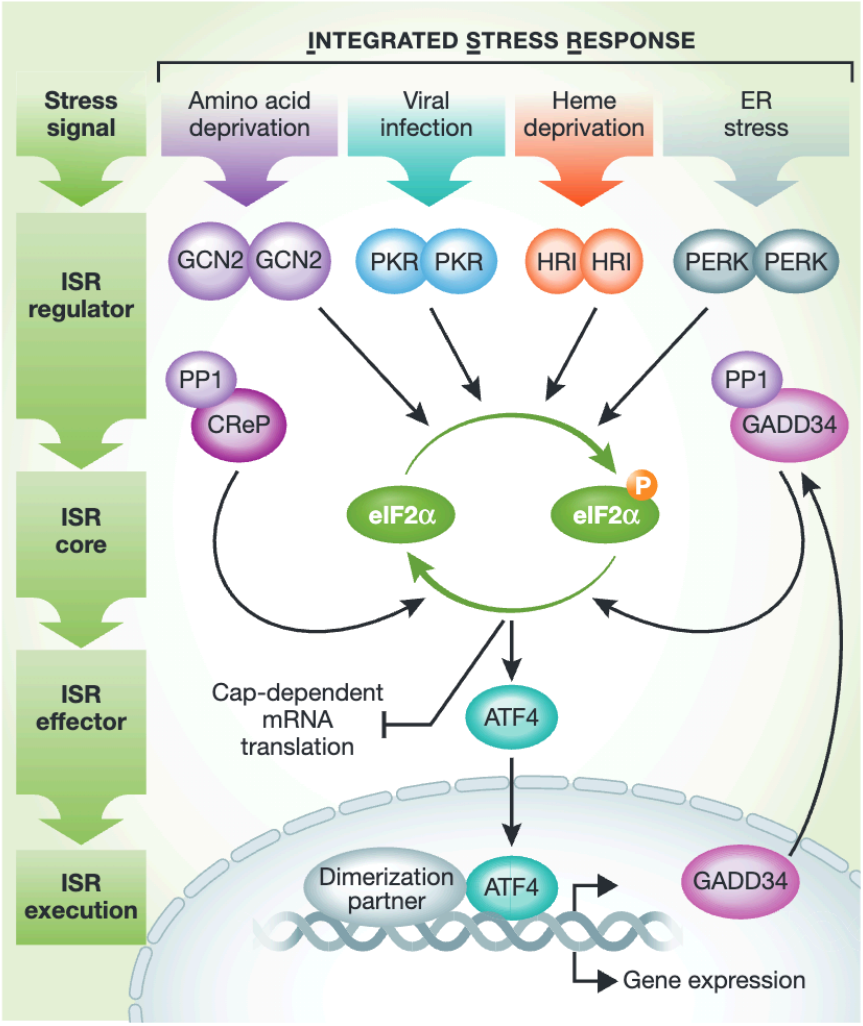


Figure 9: Schematic representation of the integrated stress response (ISR). Various stress signals activate the four ISR regulators: PERK, PKR, HRI, and GCN2 kinases. This converges on the phosphorylation of eIF2α, the core of ISR, leading to global attenuation of cap-dependent translation and the preferential translation of ISR-specific mRNAs, such as *ATF4*. ATF4 controls the expression of genes implicated in cellular adaptation. CReP and GADD34 dephosphorylate eIF2α, which terminate the ISR. From[258].

- **Regulation of translation initiation by 5'UTRs mRNA structures:**

UTRs have especially expanded in length during evolution from invertebrates to humans. Whereas 3'UTR length increased during eukaryotic evolution, the 5'UTR conserved a median length of approximately 53-218 nucleotides. As they are not coated with translating ribosomes, UTRs can interact with a plethora of regulatory factors[260]. Interestingly, various mRNA structures have been identified in the 5'UTR to regulate the initiation of translation (**Figure 10**).

Typically, eukaryotic translation starts at the 5' end of the mRNA, harboring the 5' cap and a UTR to allow the entry of the ribosome. On the one hand, some mRNAs completely lack a 5'UTR,

for instance all mRNA species in mammalian mitochondria are leaderless[261]. Additionally, a sizable number of genes with basic cellular functions, present an extremely short 5'UTR, known as translation initiator of short 5'UTR (TISU), with 12 nt median length and undergo scanning-free initiation[262]. On the other hand, some 5'UTRs have complex structures and can block translation initiation. This can be exemplified by 5'UTR structural element, the iron responsive element (IRE), regulating mRNA translation important for ion homeostasis[263]. In low-iron conditions, iron-regulatory protein 1 (IRP1) or (IRP2) binds to IRE stem-loop, prevents the 43S pre-initiation complex from associating with the mRNA, which represses translation initiation[264,265] (**Figure 10a**). The DEAD-box RNA helicase eukaryotic initiation factor 4A (eIF4A) plays an important role in unwinding 5'UTR complex structures to allow ribosome scanning. This activity is sensitive to both local RNA structures and sequence motifs[266]. For example, the 5'UTRs of eIF4A-sensitive mRNAs contain (CGG)4 motifs[267] (**Figure 10b**). In addition to eIF4A, other helicases present redundant function in translation, such as the DExH-box protein DHX29, which rescues the unwinding of structured 5'UTRs in the absence of eIF4A activity. In addition, ribosome profiling analysis showed that the eIF4A cofactor, eIF4B, stimulates translation of many mRNAs with heightened propensity for 5'UTR secondary structure[268]. These observations indicate that RNA helicases and initiation factors target specific mRNAs with structured regions to allow ribosome scanning and translation initiation.

RNA G-quadruplex (RG4) structures in mRNAs are also important for mRNA processing and translation regulation. Indeed, RG4s in 5'UTRs lead to translation repression by preventing mRNA binding to the 43S pre-initiation complex or by slowing down translation scanning. A stable RNA G-quadruplex element situated relatively proximal to the 5'UTR, within the first 50 nt, of the *NRAS* proto-oncogene contributes to repress translation[269]. In addition, scanning inhibition can be increased by recruitment of RG4-stabilizing proteins such as FMRP, which is suggested to repress the translation of *pp2acβ* mRNA[270] (**Figure 10c**). Another RBP, CCHC-type Zinc Finger Nucleic Acid Binding Protein (CNBP/ZNF9) preferentially binds G-rich elements in the target mRNA coding sequences and destabilizes RG4 structures to promote translation[271]. These studies highlight the diverse functions of tertiary stable structures in regulating the translation.

Higher-order structures mRNA, such as pseudoknots, can also affect the translation. Pseudoknots are complex intramolecular RNA structures, originally defined as structures formed by base-pairing of a single-stranded region of RNA in the hairpin loop to complementary nucleotides elsewhere in the RNA chain. They are increasingly recognized in viral and cellular RNAs and present various functions in their gene expression[272]. PKR is an interferon IFN-γ-inducible protein kinase, which is activated depending on dsRNA produced during replication of viruses or cellular stress and inhibits translation by activating eIF2α phosphorylation. Interestingly, *IFN-γ* mRNA uses its 5'UTR conserved pseudoknot structure to activate PKR and to regulate its own translation yield[273] (**Figure 10d**). Finally, RNAs can form complexes with trans-acting long non-coding RNAs (lncRNAs) for post-transcriptional regulation. For example,

antisense *Uchl1* lncRNA increases ubiquitin carboxy-terminal hydrolase L1 (UCHL1) protein synthesis, through a base pairing with *Uchl1* mRNA 5'UTR[274] (**Figure 10e**).

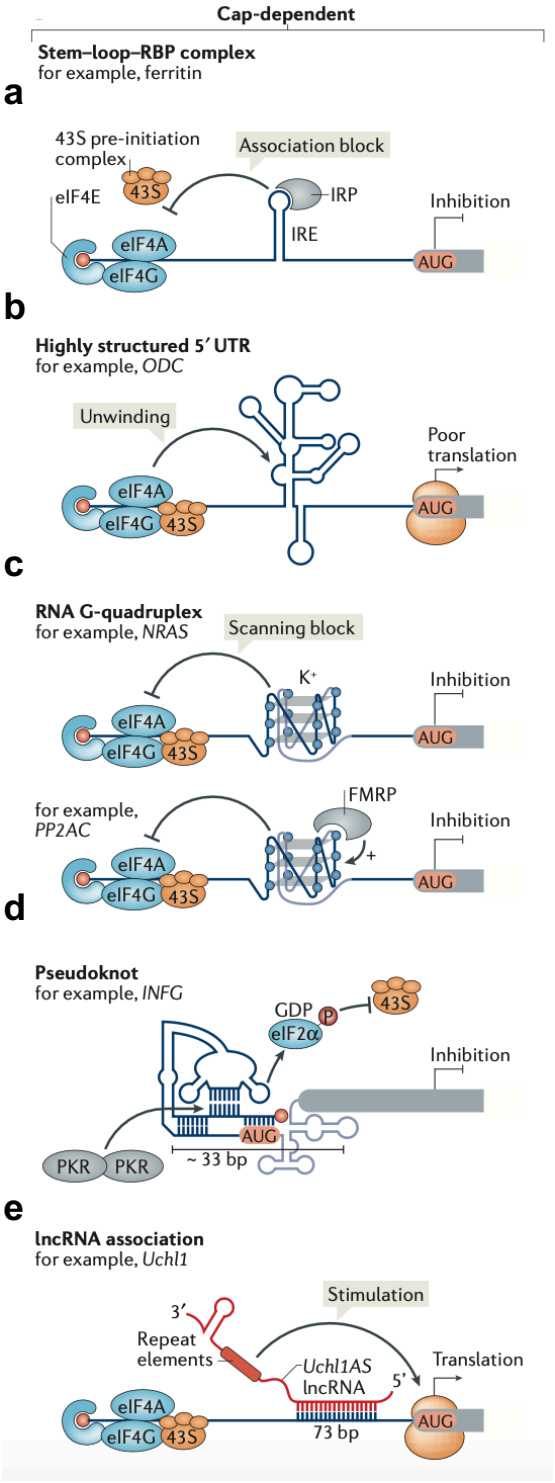


Figure 10: Regulatory 5'UTR RNA structures that influence translation initiation by promoting or inhibiting cap-dependent translation. Adapted from[275].

- **Regulation of translation initiation by RBPs:**

In addition to the regulation of global translation through changes in the phosphorylation status of eIFs, the recruitment of *trans*-acting factors, such as RNA binding proteins (RBPs), to *cis*-acting elements in the mRNA has a major impact on mRNA-specific translation[276]. To date, over 1500 RBPs have been identified with diverse cellular functions[233]. They bind RNA by RNA binding domains (RBDs), which recognize specific RNA sequences or structures. Basic mammalian canonical RBDs include RNA recognition motif (RRM), K-homology domain (KH domain), zinc finger binding domain (ZnF) and double-stranded RBD (dsRBD)[277]. Interestingly, cooperative binding of different RBPs also expands RNA sequence recognition to function in a specific manner[278]. The most common used RNA motifs include TOP motifs, uORFs, internal ribosome entry segments (IRES), cytoplasmic polyadenylation elements (CPEs), AU rich elements (ARE), and microRNA target sites[276] (**Figure 11**).

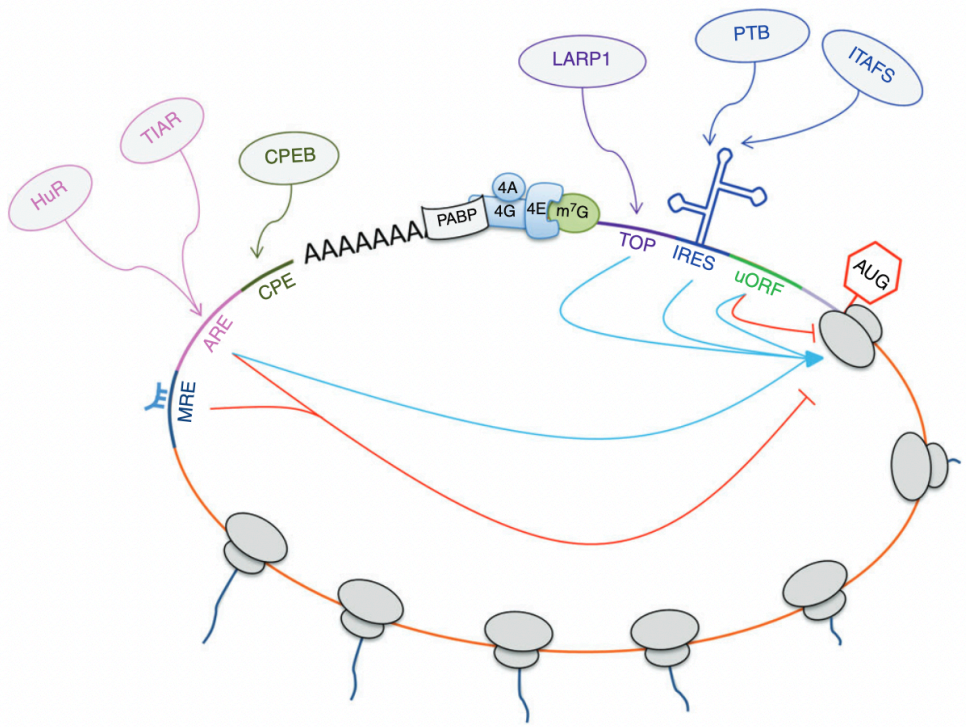


Figure 11: *Trans*-acting factors interacting with their specific RNA motifs. Schematic representation of the binding of several RBPs to their sequence-specific mRNA to regulate the translation. Terminal oligopyrimidine motifs: TOPs, internal ribosome entry sites: IRESs, miRNA-responsive: MREs, AU-rich: AREs, and cytoplasmic polyadenylation: CPEs. From[276].

Mammalian RBPs can have both positive and negative effects on the mRNA translation by interacting with specific RNA motifs. For instance, La-related protein 1 (LARP1) regulates the translation of mRNAs that contain 5' TOP motifs. A recent crystallographic data demonstrated that the DM15 region of LARP1 binds specifically to the m^7GTP cap and the first cytidine of TOP

mRNAs, thus it prevents eIF4E from interacting and blocks the assembly of eIF4F complex on TOP-containing mRNAs[279]. LARP1 had also a higher affinity for mRNAs containing TOP than eIF4E, which is consistent with its role as a repressor of translation initiation[280]. Another example of RBPs is that of the poly(A)-binding proteins (PABPs), first identified as a protein protecting mRNA poly(A) tails from deadenylation[281]. However, it presents several roles in the regulation of translation especially in cap-dependent translation initiation. Indeed, PABP can interact with the eIF4F cap-binding complex *via* the eIF4G subunit, leading to the stabilization of bound mRNAs into a closed loop conformation. In addition, PABP enhances both pre-initiation complex assembly and post-termination ribosome recycling[282,283].

- **Regulation of translation initiation by miRNAs:**

miRNAs are widely expressed and modulate transcriptional and translational programs. Therefore, they affect both physiological and pathological processes like cell differentiation, proliferation, apoptosis, and tumor growth. Interestingly, miRNAs can repress translation at the initiation, post-initiation and elongation steps[284]. For instance, a study showed that a motif (MC) within the Mid domain of Ago proteins presents a significant similarity to the m^7G cap-binding domain of eIF4E and is required to repress translation initiation of a specific subset of mRNAs; likely by precluding the recruitment of eIF4E[285]. Other studies demonstrated that the CCR4-NOT deadenylation complex can inhibit mRNA translation independently of its deadenylation activity[286] and that miRISC inhibits 43S scanning by impairing eIF4F function[287].

- **Regulation of translation initiation by ribosome concentration:**

The second hypothesis, other than the specialized ribosome hypothesis (**see below; 1.2.5.4 Ribosome heterogeneity**), to explain tissue-specificity of ribosomopathies (**see below; Ribosomopathies**) is the ribosome concentration hypothesis. It was originally proposed by Lodish[288] but was recently re-popularized by Mills and Green[289]. The ribosome concentration hypothesis postulates that low concentrations of ribosomes only marginally attenuate the translation of mRNAs with high initiation rates (*e.g.,* mRNAs encoding RPs or hemoglobin), while translation of poorly initiated (*e.g.,* mRNAs encoding hormones, transcription factors, certain Hox proteins, and GATA1 mRNAs) is greatly attenuated[289,290]. Thus, the sensitivity of specific cells to RP mutations can be the result of a selective reduction in the translation of specific mRNAs with low translational efficiencies. mRNA-specific features, such as 5' untranslated region (UTR) length or structural elements, open reading frame (ORF) length, Kozak context, and internal ribosome entry site (IRES) elements can influence the translation initiation rate. Therefore, they can influence specific gene expression because of even modest changes in ribosome concentration[291].

1.2.3.2 Regulation of translation elongation

- **Regulation of translation elongation by aminoacyl-tRNA abundance:**

Transfer tRNAs (tRNAs) are small non-coding RNAs, which allow decoding of codons into amino acids. Their secondary structure resembles a cloverleaf with four stem-loops: 1) the anticodon loop that recognizes the codons, 2) the deoxyuridine loop (D-loop) close to the 5' end of tRNA, 3) the TΨC stem-loop close to the 3' end and 4) the variable loop (V-loop)[292]. They are catalyzed by aminoacyl-tRNA synthetases (**Figure 12**). Interestingly, the availability of aminoacyl-tRNAs (aa-tRNAs) influences the rate, efficiency and accuracy of translation, and many efforts were made to quantify and characterize tRNA pools for a better understanding of their landscape in cells under different conditions[293]. Several studies showed that tRNA pools differ among tissues in humans[294], between proliferating and differentiating cells[295], as well as during different cell cycle stages[296]. This highlights that the variation of tRNA landscapes under different conditions can lead to gene expression specificity, by modulating translation elongation parameters.

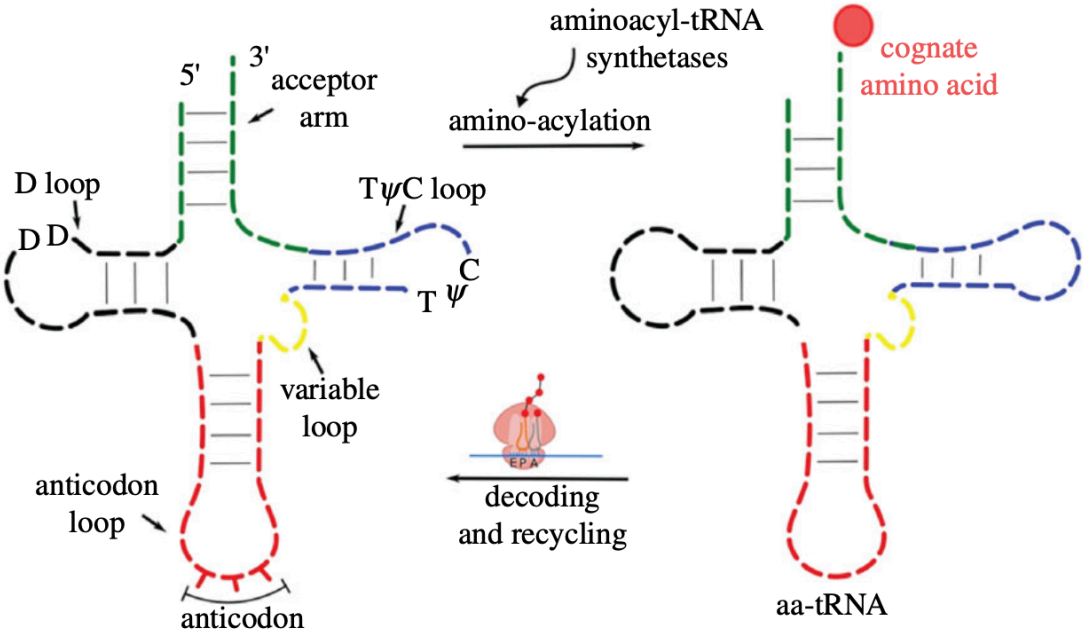


Figure 12: Secondary structure and amino-acylation of tRNAs. D: dihydrouridine; T: thymidine; ψ: pseudouridine; C: cytidine. Aminoacyl-tRNA synthetases enzymes conduct the amino-acylation of tRNAs. During translation, tRNAs undergo de-aminoacylation and re-enter the pool of free tRNA to undergo another aminoacylation cycle. From[293].

- **Regulation of translation elongation by tRNA modifications:**

tRNAs and their anticodon stem loops are known for their abundant modifications that influence their function and structure[297] (**Figure 13**). Indeed, several modifications identified in the anticodon loop can affect the decoding efficiency of tRNAs. For example, loss of U34 modifications in the anticodon wobble position, in yeast and *C. elegans*, leads to ribosome

pausing events[298]. Another modification, m5C38 mediated by DNMT2 influences the accuracy of translation[299].

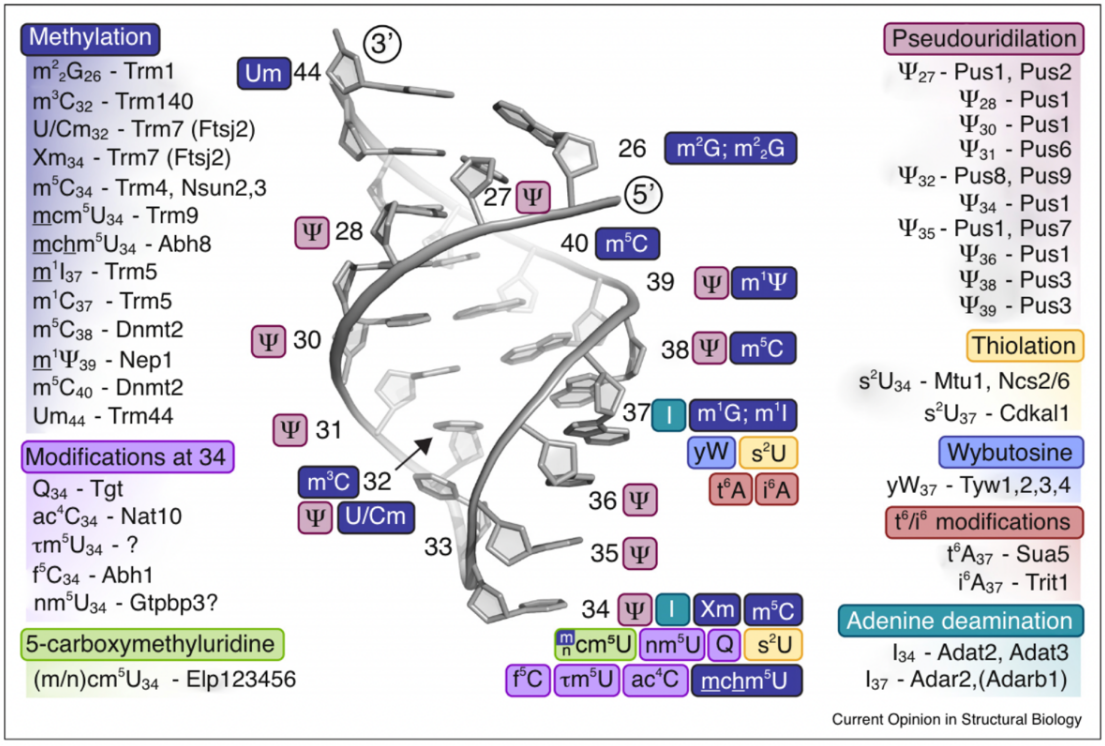


Figure 13: Modifications within the tRNA anticodon region. tRNA anticodon is depicted in cartoon representation with individual modifications. From[297].

- **Regulation of translation elongation by codon usage bias and amino acid composition:**

There are 61 sense codons for only 20 amino acids, which explains why each amino acid can be encoded by multiple synonymous codons. Codon usage bias, the phenomenon of preferential use of certain synonymous codons, is found in all genomes. Notably, “preferred or optimal codons” are preferentially used synonymous codons, whereas “nonoptimal or rare codons” are less used. Mammals prefer C/G at wobble positions, but budding yeast prefer A/U codons, probably due to different mutation biases in each organism. Several studies demonstrated that codon usage influences elongation speed, translation efficiency, initiation, and termination, as well as accuracy and co-translational protein folding[300]. Strikingly, the development of ribosome profiling has uncovered several interesting aspects of codon usage and translation. A novel analysis of the ribosome profiling data of four organisms revealed a correlation between tRNA concentrations and the putative codon-decoding rate, both in prokaryotes and in eukaryotes, supporting the idea that tRNA levels influence translation efficiency[301]. Interestingly, dwell times, defined as the time spent by a ribosome on a specific position of a transcript are codon-specific[302,303], for instance, they can be influenced by codon pairs in mouse liver[304].

Translation elongation rates can also be influenced by the amino acid composition of the synthesized peptide. Indeed, a recent study in yeast showed that negatively charged proteins are synthesized at up to ~2-fold higher rates, compared with positively charged proteins and that amino acids with small side chains are associated with faster elongation[305].

- **Deregulation of translation elongation during stress and disease:**

Although highly expressed genes are encoded predominantly by optimal codons, growing evidence are showing regulatory roles for non-optimal codons in translation dynamics under stress conditions. For example, in *E. coli*, genes encoding amino acid biosynthetic enzymes use codons adapted to starvation-induced tRNA pools[306]. A similar mechanism was demonstrated in mammalian cells, where non-optimal codons were used in promoting ubiquitin-proteasome pathway (UPP) mRNAs translation during amino acid starvation[307]. Codon optimality and tRNA levels are also involved in tumors. A recent study analyzing the tRNA abundance of more than 8,000 tumor samples from The Cancer Genome Atlas (TCGA), together with their paired mRNA-seq and proteomics data, showed an altered tRNA landscape in different cancerous tissues, mainly affected by cellular proliferation state[308].

- **Ribosome stalling and pausing:**

Ribosomes dwell times vary along the mRNA and depend on several parameters such as codon optimality, peptide-bond formation efficiency, availability of tRNA and elongation factors (**Figure 14a**), amino acid charge (**Figure 14c**) and limitation (**Figure 14b**) and nascent chain properties and poly(A) stretch (**Figure 14d**). When the dwell time is prolonged, the ribosome stops during elongation leading to a phenomenon referred to as “ribosome stalling”. Although ribosome stalling can be regulatory, stalled ribosomes can be signals that alert for the presence of defective mRNAs or altered physiological states, eliminated by the ribosome-associated protein quality control (RQC)* (**see below**). The emergence of several techniques helped to map ribosomal stop sites for better understanding of causes and fates of stalled ribosomes and showed the enrichment of certain motifs and the implications of some factors in the regulation of stalling[309].

A well-studied example of ribosome stalling is that of ribosomes stall when they encounter poly(A) (**Figure 14d**). Biochemical and structural approaches in mammalian systems solved the structure of the ribosome translating a poly(A) stretch. In this work, they showed that poly-lysine, encoded by poly(A) favors a consequent rearrangement of the conformation at the decoding center. The reconfigured decoding center clashes with incoming aminoacyl-tRNA, thereby preventing elongation[310]. In addition, limitation of some amino acids can cause ribosomes to stall during elongation (**Figure 14c**). A recent study described that a selective loss of arginine tRNA charging regulated translation through ribosome pausing at two of six arginine codons (**Figure 14b**). However, limitation for leucine, resulted in little or no ribosome pausing. These

observations reflect an amino acid- and codon-specific stalling effect in elongation[311]. Other features such as charge and structure of the nascent polypeptide, slow release of ribosomes at the stop codons were identified as causes of ribosomes pausing and stalling[312,313]. Interestingly, the eukaryotic initiation factor 5A (eIF5A), the sole protein in the eukaryotic proteome to undergo the post-translational modification known as *hypusination*[314], presents important function during translation elongation. Indeed, it was described to affect the elongation of poly-proline motifs[315]. Once present in a protein, a poly-proline stretch could slow down elongation, creating a ribosome stall. In this case, eIF5A recognizes the stalled conformation and the hypusine residue can reach the PTC, therefore stabilizing the prolines and facilitating the peptide-bond formation[316]. Recently, eIF5A function in elongation was also confirmed in mammalian cells[317].

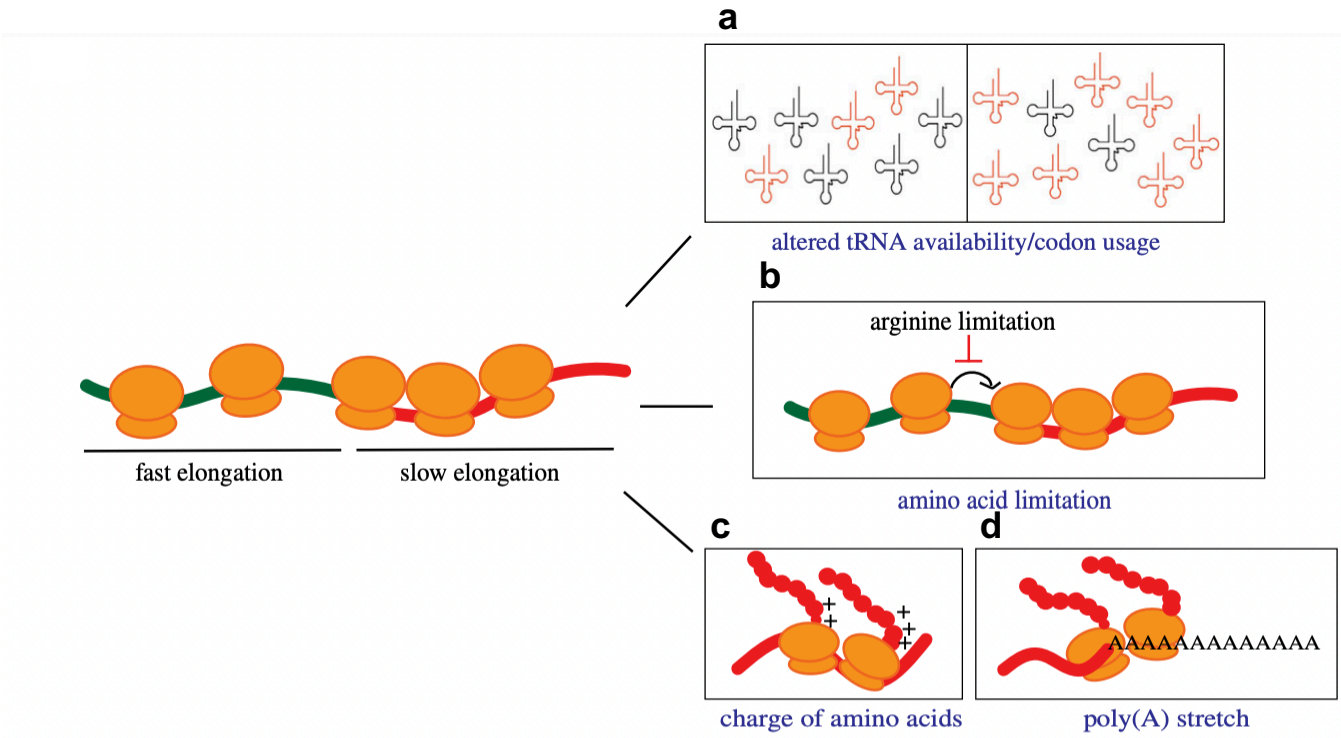


Figure 14: Factors that influence translation elongation. Adapted from[293].

*Ribosome-associated protein quality control (RQC)

Faithful interpretation of the genetic code is essential to generate functional protein products, participating in all areas of cellular physiology[318]. Hence, every step of protein production is subject to quality control to degrade aberrant mRNAs[319], detect mutant or damaged rRNAs and ribosomes[320], ensure appropriate tRNA amino-acylation[321], monitor the kinetic proofreading during codon-anticodon recognition[322] and ensure the proper folding of nascent polypeptides during and after protein synthesis[323].

Particularly, stalling of ribosomes results in the production of truncated polypeptides with deleterious effects on cells, which in eukaryotes, are eliminated by a dedicated surveillance mechanism known as ribosome-associated protein quality control (RQC). The RQC consists of two sequential steps, each sensing a unique defect. During the first step of ribosomal rescue, ribosomes stalled at the mRNA 3' end are sensed by Hbs1 in yeast (or HBS1L and GTPBP2 in mammals) and Dom34 (pelota (PELO) in mammals). In addition, recognition of stalled ribosomes can involve the E3 ubiquitin-protein ligase Hel2/ZNF598, Asc1/RACK1 and Slh1. Then, Rli1 (ATP-binding cassette protein subfamily E member 1 (ABCE1) in mammals) is recruited by Dom34/PELO, leading to the separation of the 40S and 60S ribosomal subunits. With the rescue reaction, released 40S ribosomal subunits can be recycled, and the truncated mRNA is degraded by the 5'-3' exoribonuclease Xrn1 and the exosome complex to repress mRNA translation and keep low levels of proteins. Another product of the rescue reaction is the aberrant 60S subunit attached to an obstructing nascent-chain-tRNA conjugate. Sensing of the obstructed 60S ribosomal subunits is mediated by the RQC complex subunit Rqc2 (NEMF in mammals) that recruits E3 ligase Ltn1/listerin and stabilizes its binding to the 60S. Then, the ubiquitin (Ub) chain is polymerized by Ltn1 on the nascent polypeptides. This signals the recruitment of the AAA ATPase Cdc48 (VCP or p97 in mammals) and its cofactors. This recruitment also involves the Rqc1 subunit of the RQC complex. Finally, Cdc48 extracts nascent polypeptides from the 60S subunit and delivers them to the proteasome for degradation. This occurs after these polypeptides have been released from the conjugated tRNA by Vms1 (ANKZF1 in mammals), which is a paralog of eRF1[324] (**Figure 15**).

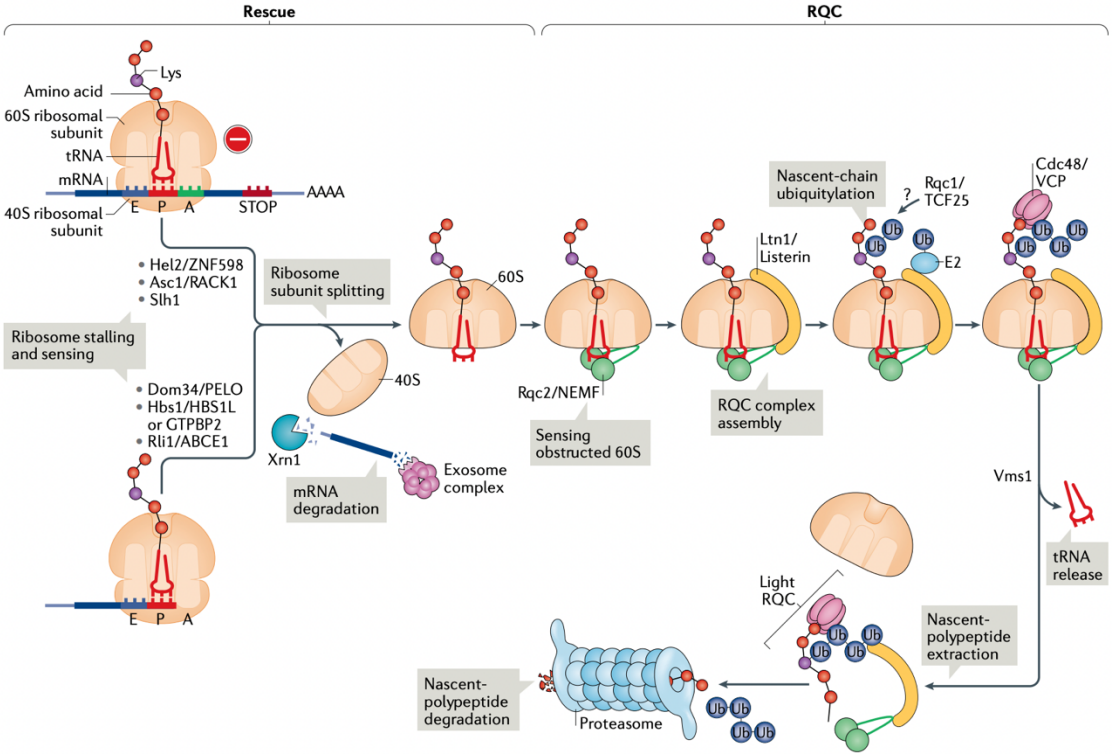


Figure 15: The detailed steps of the ribosome-associated protein quality control (RQC) pathway. From[324].

1.2.3.3 Regulation of translation termination

- **Regulation of translation termination by *trans*-acting factors:**

Translation termination can be regulated by various *trans*-acting factors. Numerous observations suggest that NMD (*i.e.,* the surveillance mechanism of destruction of mRNAs harboring premature termination codons (PTCs)[325]) factors influence termination, possibly in an organism-specific manner. For instance, ATP hydrolysis by UPF1 is important for efficient termination at PTCs in yeast[326], whereas human UPF1 does not appear to bind eRF1 and eRF3a directly and had no effect on termination[327]. eIF5A can also affect the termination, as ribosome profiling in eIF5A-depleted *S. cerevisiae* showed impaired termination[328]. In addition, retroviral reverse transcriptase (RT) of Moloney murine leukemia virus (MoMLV) interacts with the C-terminal domain of eRF1, occluding the binding of eRF3 to eRF1 and thereby promoting stop codon read-through[329] (**see below**).

- **Regulation of termination by *cis*-acting RNA elements:**

Stop codon read-through is the phenomenon that a ribosome goes past the stop codon of a transcript and continues translating in the same frame, adding a peptide extension to that instance of the protein. Although, it can alter the protein function, it can be beneficial under certain circumstances[330]. Readthrough depends on particular mRNA sequences and structures that can be located close to the stop or lying hundreds of nucleotides downstream from it[331]. For example, in murine leukaemia virus (MLV), a pseudoknot downstream from the *gag* stop codon can promote read-through[332]. Although read-through is common in viruses as it increases functional versatility in their compact genome, a handful of eukaryotic wild-type genes were shown to exhibit this process[333]. In the case of the *Drosophila hdc* stop codon, an 80-nucleotide long sequence downstream of UAA can form a secondary structure that can function in heterologous mRNAs. This readthrough is necessary for *hdc* function as a branching inhibitor during tracheal development[334]. Vascular endothelial growth factor-A (*VEGFA*) mRNA in mammalian endothelial cells can also undergo programmed translational read-through (PTR) to generate VEGF-Ax. A cis-acting element in the *VEGFA* 3'UTR directs the PTR. In addition, heterogeneous nuclear ribonucleoprotein (hnRNP) A2/B1 binds this element and promotes the read-through[335].

- **Regulation of translation termination by post-translational modification:**

Components of the translation termination apparatus can be post-translationally modified to regulate the termination of translation, but functional consequences of these modifications remain largely uncharacterized. For instance, a study showed that an uncharacterized JmjC 2OG oxygenase, Jmjd4, optimizes translational termination via lysyl hydroxylation of the stop codon recognition domain of eRF1[336].

1.2.4 Alternative translation initiation mechanisms

Although eukaryotic mRNAs were historically known to rely exclusively on cap-dependent translation, there is now emerging evidence that they also undergo non-canonical modes of translation initiation[337].

- **Translation of H4 mRNA by tethering mechanism:**

Histones are very conserved proteins in the eukaryotic kingdom, massively produced during the S phase of cell cycle. Their 5'UTRs are very small, ranging from 20 to 60 nt, with the mouse histone H4-12 presenting a 9 nt 5'UTR and their 3' end lack a poly (A) tail. Instead, they end in a highly conserved 26 nt sequence, containing a 16 nt hairpin structure recognized by the stem-loop binding protein (SLBP)[338,339]. This structural organization do not conform to the conventional scanning-initiation model. Indeed, the ORF of Histone H4 mRNA harbors two structural elements critical for translation initiation: 1) the eIF4E-sensitive element (4E-SE) allowing the recruitment of eIF4E without the need of the cap and 2) the Three-Way Junction (TWJ), located 19 nucleotides downstream from the AUG codon and sequestering the cap by forming a Cap-Binding Pocket (CBP). Therefore, Histone H4 mRNA adopted a ribosome tethering mechanism, combining canonical features (cap-dependent) and viral strategy (lack of scanning)[340]. The interaction between an AGG triplet located upstream of the TWJ and the loop of eukaryotic helix h16 of the 18S ribosomal RNA leads to the correct positioning of the AUG codon in the P-site, but also stabilizes the 48S binding on this codon[340] (**Figures 16a,b**).

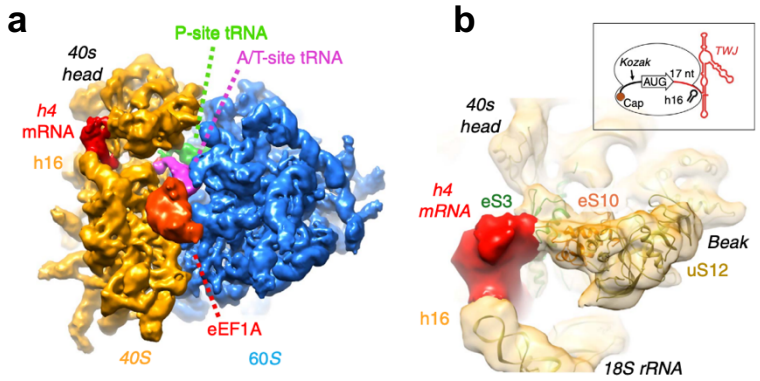


Figure 16: Histone *h4* mRNA localization on the 80S ribosome. (**a**) Overview of the *h4*/80S complex stalled in the pre-translocation state. h4 and eEF1A are in red, A/T-site tRNA in magenta, P-site tRNA in green, 60S ribosomal subunit in blue and 40S ribosomal subunit in orange. (**b**) Localization of *h4* mRNA between the tip of ribosomal helix h16 and proteins eS3 and eS10. TWJ: Three-Way Junction. From[340].

- **Translation by 3' cap independent translation elements (3'CITEs) in plant RNA viruses:**

Another example of atypical translation initiation occurs in the plant viral mRNAs harboring 3'CITEs. Cap independent translation elements (CITEs) have been identified in the 3' ends of

members of the *Tombusviridae* and *Luteoviridae* plant virus families lacking both 5' cap and 3' poly(A) elements. To date, seven different classes of 3'CITEs have been categorized, based on their different RNA structures: 1) Translation Enhancer Domain (TED), 2) Barley Yellow Dwarf Virus (BYDV)-like element (BTE), 3) Panicum mosaic virus-like Translational Enhancer (PTE), 4) T-Shaped Structure (TSS), 5) I-Shaped Structure (ISS), 6) Y-Shaped Structure (YSS) and 7) CXTE[341] (**Figure 17**). They generally regulate the formation of the translation initiation complex (TIC) by various mechanisms. The eIFs bound to the viral 3'CITEs are thought to be brought to the 5'end of translation initiation complex by a "kissing-loop" interaction, based on sequence complementarity between the 3'CITE and the 5'end[341]. Some 3'CITEs bind eukaryotic translation initiation factors. For example, BTE interacts with eIF4G[342] and PTE with eIF4E[343].

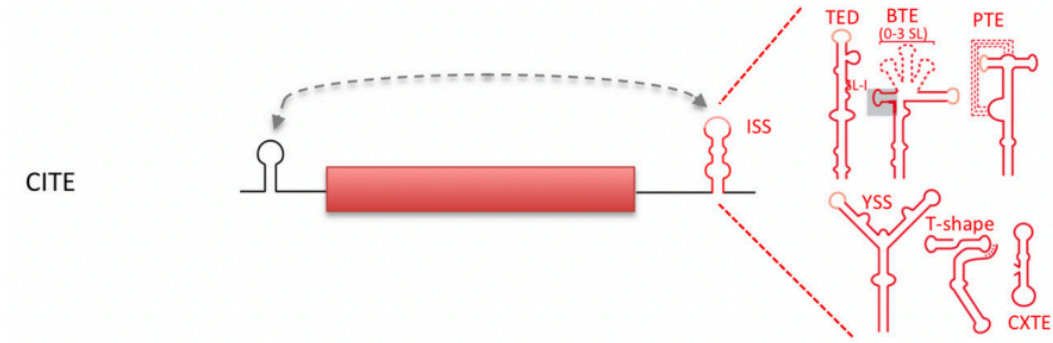


Figure 17: 3'CITE translation initiation used by plant RNA viruses. Lighter-shaded loops of 3'CITEs structure show sequences known or predicted to base-pair to the 5'end of the plant viral genome. ISS: I-Shaped Structure, TED: Translation Enhancer Domain, BTE: Barley Yellow Dwarf Virus (BYDV)-like element, PTE: Panicum mosaic virus-like Translational Enhancer, YSS: Y-Shaped Structure. Adapted from[710].

- **Translation regulation by adenosine methylation (m^6A):**

RNA molecules undergo a vast number of modifications. N^6-Methyladenosine (m^6A) has been considered as the most abundant mRNA internal modification, regulating multiple steps of gene expression, including splicing, export, stability, and translation[344]. The mechanisms by which m^6A regulates the translation are dependent on the binding of reader proteins to protein factors implicated in the translation and m^6A modification locations in different RNA regions. Interestingly, the m^6A reader protein, human YTHDF1, directly promotes the translation of methylated mRNAs, as YTHDF1 knockdown led to reduced translation efficiency of its target transcripts[345].

In vitro reconstitution approaches and translation assays showed that mRNAs containing m^6A in their 5'UTRs are translated in a cap-independent manner, through a direct binding of eukaryotic initiation factor 3 (eIF3) to nucleotide m^6A, allowing the recruitment of the 43S complex in the absence of eIF4E[346]. However, when present in the mRNA coding sequence, m^6A modifications induce aberrant tRNA selection, therefore causing a reduction in elongation dynamics[347].

- **Translation initiation mediated by eIF3:**

The eukaryotic initiation factor 3 (eIF3) is the largest and most complex eukaryotic initiation factor (800 kDa, approximately), comprising 13 subunits (eIF3a-m). eIF3a, eIF3b and eIF3c were conserved through evolution, whereas eIF3e, eIF3f and eIF3h were not. This protein is important in translation initiation and termination, and in ribosomal recycling. Therefore, its deregulation is associated with different pathological conditions, including cancer[348]. eIF3 is a multitasking factor coordinating the progress of most of the initiation steps, including scanning, start codon recognition, and termination[349].
Interestingly, several studies described non-canonical roles of eIF3 in translation. For instance, the m^{6}A reader, YTHDF3, can join YTHDF1-eIF3 complex through its binding to YTHDF1 and promote the translation[350]. In addition, the m^{6}A writer METTL3 forms a complex with eIF3, augmenting the translation of some mRNAs involved in tumor progression and apoptosis[351].

- **Translation regulation of circular RNA (circRNA):**

Circular RNAs (circRNAs) are closed noncoding RNAs, without 5' and 3' ends, that can regulate gene expression by associating with RNA binding proteins and miRNAs. They are usually stable because of their resistance to the degradation mediated by exonuclease and can be translated into proteins[352]. At first, circRNAs were thought to be noncoding. However, recent studies are suggesting their association with polyribosomes IRES elements and m^{6}A modified nucleotide sequences for cap-independent translation initiation. Indeed, circRNAs with IRES, such as circZNF609, can be associated with polyribosomes and generate proteins by IRES-mediated translation mechanisms[353] (**Figure 18a**). Moreover, circRNAs bearing the m^{6}A motif "RRACH" (R = G or A; H = A, C or U) in their 5'UTRs can associate with YTH domain family protein 3 (YTHDF3; m^{6}A reader) and recruit the translation initiation complex to initiate the translation[354] (**Figure 18b**).

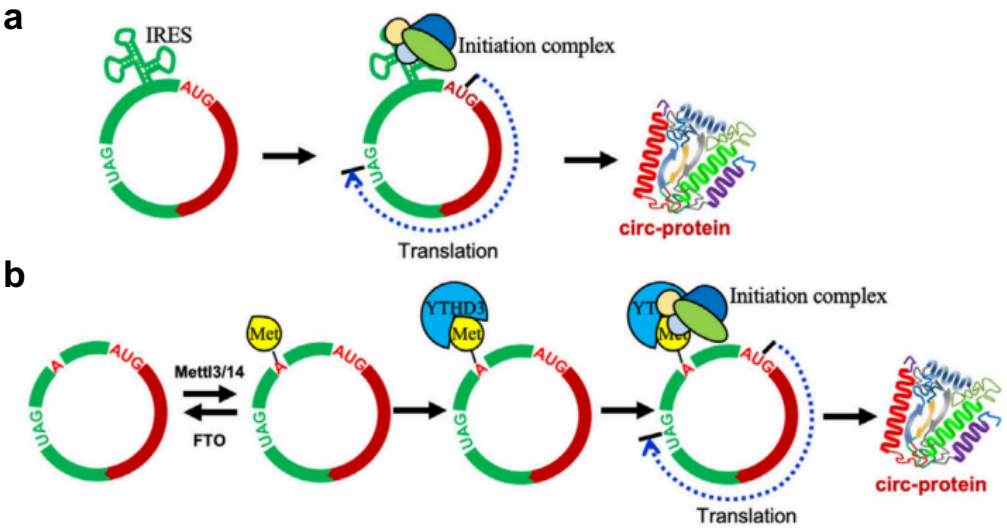


Figure 18: CircRNA in cap-independent translation. (**a**) CircRNAs with IRES or (**b**) with m^6A modifications that associate with YTHD3, recruit translation initiation complex to initiate the translation. CircRNA: circular RNA, IRES: internal ribosome entry site. From[352].

- **Translation by internal ribosome entry site (IRES):**

Internal ribosome entry site (IRES) elements are one of the cap-independent mechanisms, responsible for 3-5% of all translation initiation events[355]. They were discovered originally in the late 1980s, in 5'UTRs of Poliovirus (PV) and encephalomyocarditis virus (EMCV); two picornavirus family members targeting the host translation factors to shut down its translation[238].

Later, cellular IRES mRNAs were discovered. They function when canonical cap-dependent translation is compromised under many physiological, pathophysiological and stress conditions. Such conditions include endoplasmic reticulum (ER) stress, hypoxia, nutrient limitation, mitosis, and cell differentiation. Cellular IRESs structures are complex and include stem loops and pseudoknots. Nevertheless, no common sequence and/or structural motifs have been detected among them. Cellular IRESs can be located in the 5'UTR, downstream of the initiation codon or in the coding regions of the message and are likely to interact with components of the translational machinery (IRES-*trans*-acting factors (ITAFs), canonical initiation factors, and 40S ribosomal subunits)[356].

There has been an increased focus on viral IRESs, which can be classified into four types/classes based on their eIFs recruitment, their structure, and their mode of action.

- Type 1: PV IRES is the best characterized among the type 1. The PV 5'UTR is 743 nt-long, with the first 100 nucleotides being a cloverleaf (CL) structure required for replication of both positive and negative viral RNA strands. It also contains RNA structural elements that have been grouped into six domains (I–VI). PV IRES requires all the eIFs except eIF4E to assemble the PIC upstream of the coding region[238]. The PV RNA genome encodes a viral protease 2A that cleaves eIF4G, leading to the loss of the eIF4E-binding, thereby inhibiting cap-dependent translation[357]. An AUG codon, at positions 586-588 is the ribosome entry site. Then, the ribosome scans the RNA to reach a second AUG codon at position 743 and initiates the translation. PV IRES also requires another ITAF, PCBP2, a viral replication and translation *trans*-acting factor[358] (**Figure 19a**).

- Type 2: IRESs from type 2 are found in *Picornaviridiae* family in the genera of *Cardiovirus,* such as encephalomyocarditis virus (EMCV), and *Aphtovirus* such as foot and mouth disease virus (FMDV). IRESs of these two viruses have a comparable size (439 nt for EMCV and 445 nt for FMDV) and same domain organization (domains I-V), whereas their nucleotides sequences are not identical[359,360]. Domain I is required for viral replication. Domain II contains a pyrimidine tract, recruiting PTB in EMCV and FMDV IRESs, and Domain III encompasses two conserved motifs GNRA and RAAA required for IRES

structural and functional organization[238]. Both IRESs have two in-frame AUG codons: EMCV only uses the second AUG for translation initiation, but FMDV uses both[361]. IRESs of this type are independent of eIF4E and do not require further scanning step[242] (**Figure 19b**).

- Type 3: Type 3 IRESs were described in the *Flaviviridae* family-like hepatitis C virus (HCV), classical swine fever virus (CSFV) and some members of the *Picornaviridae*, such as porcine teschovirus and porcine enterovirus 8 (PEV8) or simian virus 2 (SV2)[362]. HCV, belonging to the *Flaviviridae* family, is the best characterized member of this type. Its 5'UTR is highly conserved and contains four domains (I-IV)[363]. Domain I contains two binding sites for miR-122, activating HCV replication[364]. Domain II is a long hairpin encompassing two subdomains: IIa containing a multinucleotide bulge hinge and IIb harboring an E motif internal loop and the apical hairpin[365]. Domain III is more structured with six subdomains (a-f). Domain IV is a stem loop structure containing the AUG start codon and is unfolded when the IRES is located in the ribosomal mRNA channel. IRES spans domain II and III, orchestrating the 40S recruitment. After this recruitment, the HCV IRES interacts with eIF3, which itself interacts with the ternary complex, leading to the formation of the 48S[238]. Interestingly, HCV translation does not require scanning, suggesting that eIF3 may have a role of "mediator" of translation[238,366] (**Figure 19c**).

- Type 4: The *Dicistroviridae* family contains in its viral genome two open reading frames (ORF1 and ORF2). Translation from the second ORF is driven by an IRES, localized between the two ORFs, in the intergenic region (IGR). Type 4 IRESs are found exclusively in the IGR of the *Dicistroviridae* family. Therefore, they are called IGR. These IRES present a simple mode of action[238]. Indeed, i) they do not need any eIFs, ii) do not use the initiator Met-tRNA^{Met} and iii) they initiate translation on a non-AUG start codon[367–369]. The most characterized member in this type is cricket paralysis virus (CrPV) but others have also been investigated such as Platia Stali instestine virus (PSIV), or Taura syndrome virus (TSV). CrPV IRES harbors three domains, each domain containing an essential pseudoknot structure named PKI, PKII, and PKIII. Domains I and II are required for ribosome recruitment. Domain III containing the pseudoknot PKI, is important for efficient translation initiation as PKI mimics a codon-anticodon complex to establish the correct reading frame in the viral messenger upon interaction with the ribosome[238] (**Figure 19d**).

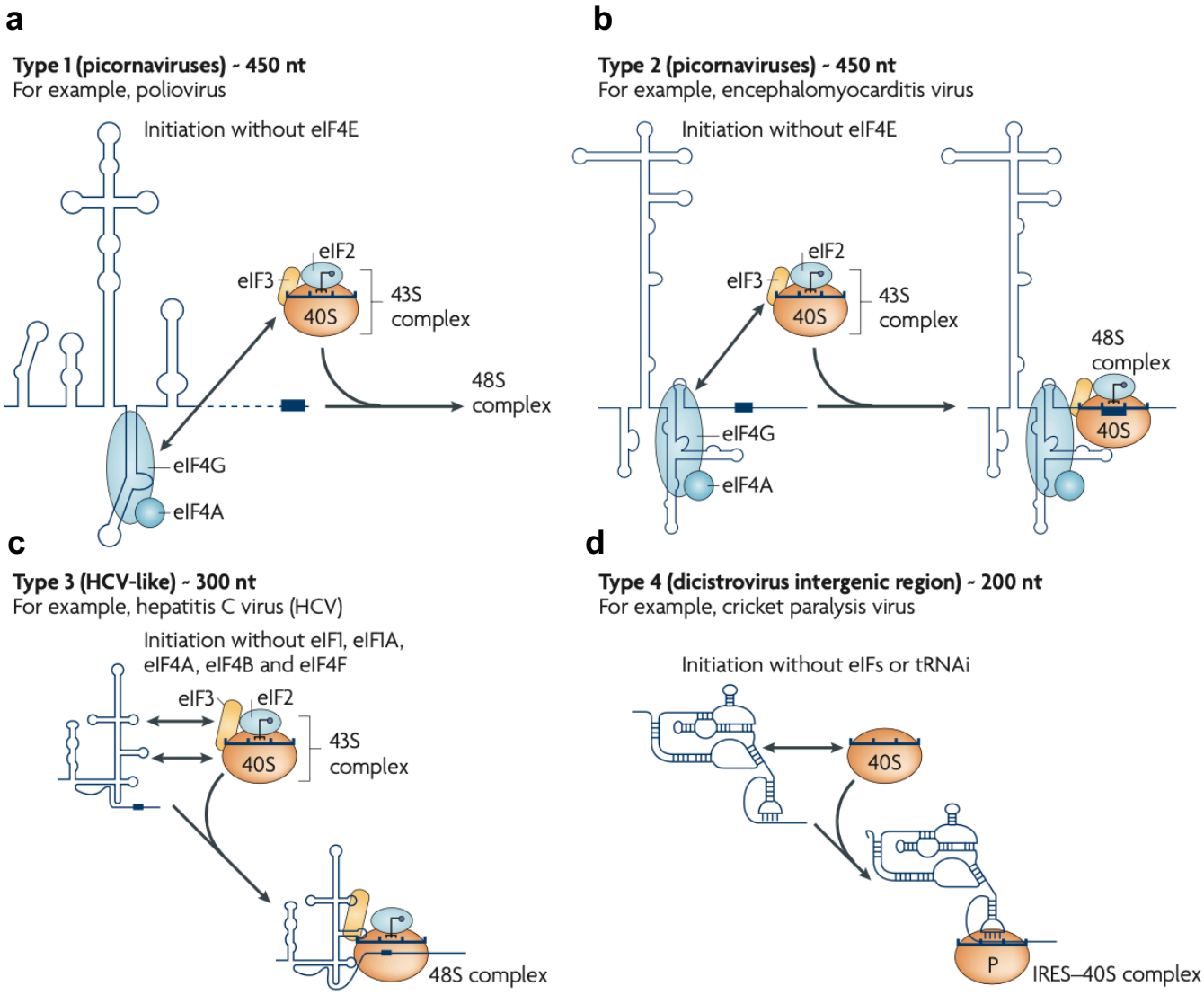


Figure 19: Translation initiation mediated by IRES. The four types of IRES use different mechanisms, based on non-canonical interactions with eukaryotic initiation factors (eIFs) and/or 40S subunits. (**a**) Type 1, (**b**) Type 2, (**c**) Type 3, (**d**) Type 4 IRESs. IRES: internal ribosome entry site. From[242].

1.2.5 Ribosomes

1.2.5.1 Ribosome's structure and organization

In the 1950s, George Palade first described the ribosome, when he observed dense intracellular particles using electron microscopy (EM). Then in 1974, he was awarded the Nobel Prize for this discovery, along with Albert Claude and Christian de Duve[370].

At the simplest level of description, the ribosome is a ribonucleoprotein assembly that synthesizes the proteome in all living cells. It consists of two unequal subunits: the large subunit

(LSU) and the small subunit (SSU), in which the former is about twice the size of the latter[371]. Ribosome components size is given in Svedberg units (S), which is a measure of the rate of sedimentation during centrifugation[370]. Ribosome subunits contain ribosomal RNA (rRNA) and ribosomal proteins (RPs)[372]. Functionally, LSU, which houses the peptidyl transferase center (PTC), catalyzes the formation of peptide bond, whereas SSU serves as the decoding center to bring together mRNAs and tRNAs in order to translate the genetic code[373].

Despite its universality, the compositions of the subunits of different species are not identical (**Figure 20**). Specifically, the 40S subunit (SSU) of eukaryotes comprises 18S rRNA and 33 different RPs (referred to as RPS or S), while the 60S subunit (LSU) contains 5S rRNA, 5.8S rRNA, 25S/28S rRNA, and 46 (in yeast)/47 (in human) RPs (referred to as RPL or L). Together, 40S and 60S subunits constitute the 80S ribosome, which ranges in size from 3.5 MDa in lower eukaryotes to 4.0 MDa in higher eukaryotes[371,374]. In contrast, ribosomes from bacteria and archaea consist of a large (50S) (containing 23S and 5S rRNAs) and a small (30S) subunit (containing a 16S rRNA), which interact with 60-70 RPs[375]. Together, they constitute the 70S ribosome of approximately 2.5 MDa[374].

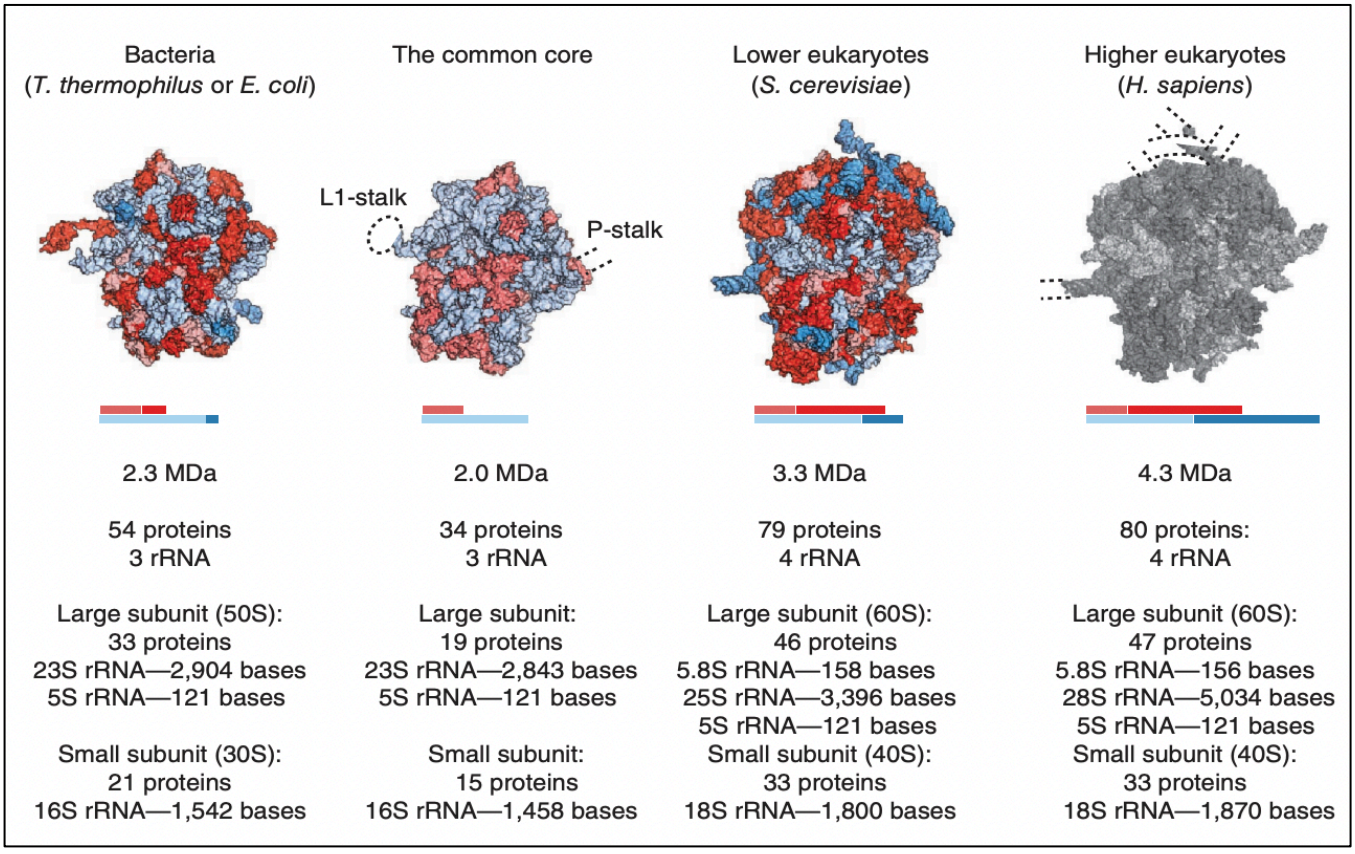


Figure 20: Composition of bacterial and eukaryotic ribosomes and the common core. The common core is formed by RNA (light blue) and proteins (light red) and is shared between bacterial and eukaryotic ribosomes. Both extensions and insertions in conserved proteins are presented in red and extension segments in ribosomal RNA are in blue. Dashed lines around the core show positions of flexible stalks of the ribosomes. The yeast 80S structure is shown in gray (instead of the human ribosome structure) and the dashed lines indicate the positions of human-specific long rRNA expansion segments. From[376].

Despite the conservation of many ribosome components across the three kingdoms of life (Bacteria, Archaea and Eukarya), their composition varies considerably between them, reflecting the functional divergence of ribosomes. Indeed, throughout evolution, the ribosome has encountered a concomitant increase in both size and complexity. Apart from the core, which refers to structurally conserved part of 70S ribosomes (from *T. thermophilus* and *Escherichia coli*) and 80S ribosomes (from *S. cerevisiae*), each of the ribosomes contains its own set of specific moieties[376] (**Figure 20**).

Remarkable advances have been made in full-ribosome crystallography, which helped researchers to better understand ribosomal complexes and their functions. As described previously, the 70S and 80S ribosomes are asymmetric assemblies. Each ribosomal component is present in the ribosome as a single copy, except for P-stalk proteins, presenting several copies. Early genetic data, corroborated by structural studies, demonstrated that the bacterial and eukaryotic ribosomes share a common structural core, consisting of 34 conserved proteins (15 in the small subunit and 19 in the large subunit) and approximately 4,400 RNA bases. This common core bears the major functional centers of the ribosomes, such as the decoding site, tRNA-binding sites and the peptidyl transferase center[376]. Besides this core, ribosomes from different species contains domain-specific proteins, insertions and extensions of conserved proteins and expansion segments of rRNAs. These rRNA and ribosomal proteins envelop the core from the solvent side and are therefore accessible for potential interactions with molecular partners[377,378].

The 30S and 40S ribosomal small subunits have similar structural landmarks known as the “head”, “body”, “platform” and “beak”. On the subunit interface, are located the mRNA- and the three tRNA-binding sites (A, P and E). The 50S and 60S subunits also share similar overall crown-like shapes, including the “central protuberance (CP)”, “L1-stalk” and the “L7/L12-stalk” (“P-stalk” in eukaryotes). As ribosomal proteins from *Escherichia coli* were the first to be described, their archaeal and eukaryotic homologues were assigned *E. coli* names:

- Prefix “u” (for universal) designate ribosomal proteins shared in all three domains (bacteria, archaea, and eukaryotes).
- Prefix “b” (for bacterial) designate bacterial proteins without eukaryotic homologues.
- Prefix “e” (for eukaryote) designate eukaryotic proteins without bacterial homologues[374,376] (**Figures 21a,b**).

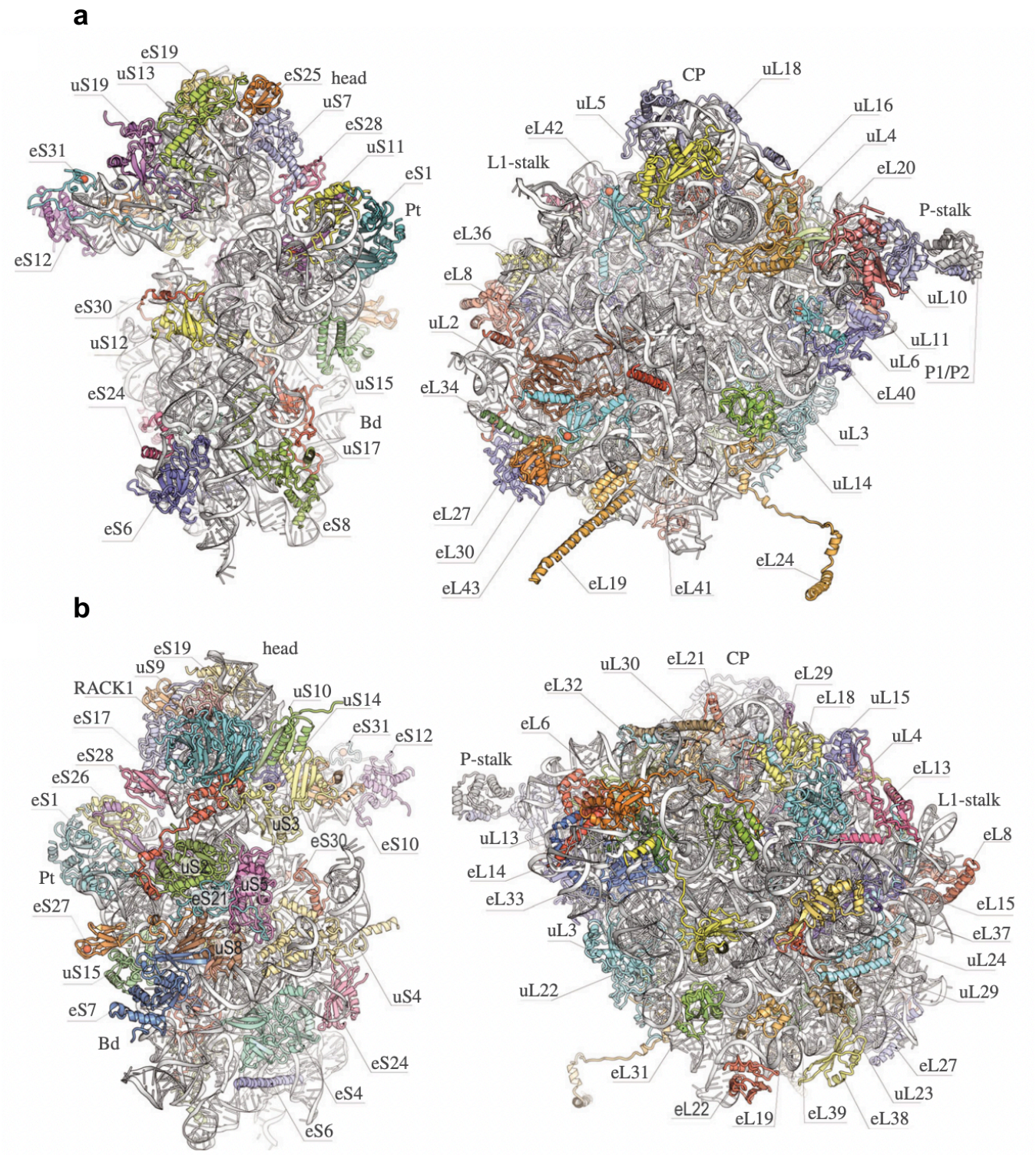


Figure 21: Architecture of the 40S (left site) and 60S (right site) ribosome subunits with new protein nomenclature. (**a**) Interface and (**b**) solvent side views of the 40S and 60S subunits. Body: Bd, platform: Pt of 40S, central protuberance: CP. U: universal, e: eukaryote. From[374].

1.2.5.2 Ribosome biogenesis

Ribosome biogenesis is a fundamental process providing cells with the molecular factories that are important for cellular protein production. It is one of the most energetically demanding of cellular activities, which comprises processing and folding of pre-rRNAs and their concomitant assembly with the ribosomal proteins (**Figure 22**).

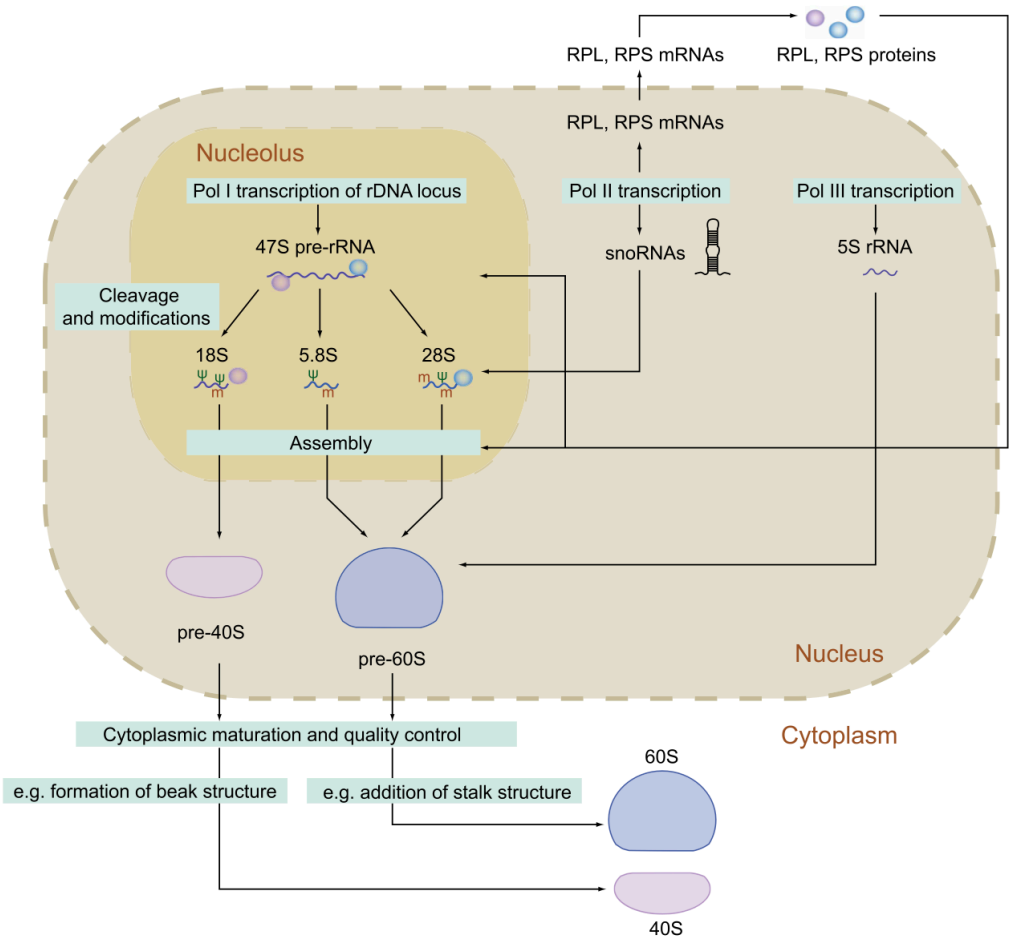


Figure 22: Representation of the different steps of eukaryotic ribosome biogenesis. Ψ: Pseudouridylation and m: 2'-O methylation of ribose. From[380].

This process is tightly regulated according to the cellular environment; hence it is linked to cellular processes, including growth and cell division. Eukaryotic ribosome biogenesis relies on a large number (> 200) of non-ribosomal factors and the coordinated activity of all three RNA polymerases. The plethora of non-ribosomal factors include various families of energy-consuming enzymes, notably ATP-dependent RNA helicases, AAA-ATPases, GTPases, and kinases[379].

Briefly, in mammals, Pol I transcribes the nucleolar ribosomal DNA loci, generating the 47S rRNA precursor (pre-rRNA). The latter is subjected to specific processing into three mature

rRNAs: the 18S, 5.8S and 28S. Pol III transcribes the 5S and Pol II transcribes loci encoding ribosomal proteins of the large and the small subunits, as well as the snoRNAs, which are involved in adding modifications to rRNAs, such as pseudouridylation (Ψ) and 2'-O methylation of ribose (m). Subsequently, modified rRNAs are assembled with imported cytoplasmic ribosomal proteins (RPLs and RPSs) into the pre-60S (28S rRNA, 5.8S rRNA, 5S rRNA, and RPLs) and the pre-40S (18S rRNA, and RPSs). These pre-subunits are exported into the cytoplasm, where they are subjected to additional maturation and quality control steps, including the addition of the stalk structure for the 60S subunit and the formation of the beak structure for the 40S subunit. Before the subunits become competent for translation, they also undergo further quality control steps[380] (**Figure 22**).

1.2.5.3 Types of ribosomes and their localization

Ribosomes are mainly located in the cytoplasm, where they can be 1) associated with the ER to facilitate the translation of luminal, secretory, transmembrane and a subpopulation of cytosolic proteins, or 2) free in the cytosol to synthesize cytosolic proteins[381].

Beyond these "cytosolic eukaryotic ribosomes", additional ribosomes were found within semiautonomous eukaryotic organelles of bacterial ancestry: mitochondria (ribosomes are referred to as "mitoribosomes"), as well as chloroplasts (ribosomes are referred to as "chlororibosomes") for the Archaeplastida kingdom[382]. Specifically, mammalian mitoribosomes sediment as 55S particles, with mitochondrial rRNAs transcribed from mtDNA genes. They are constituted of a 28S small subunit (mtSSU), formed by a 12S rRNA and 29 mitoribosomal proteins (MRPs), and 39S large subunit (mtLSU), formed by a 16S rRNA and 50 MRPs[383].

Mitoribosomes primarily locate in the matrix of the organelle and are attached to the inner membrane to facilitate co-translational insertion of highly hydrophobic nascent polypeptides. They are essential for the synthesis of polypeptides, predominantly implicated in facilitating ATP production aerobically and oxidative phosphorylation. Owing to evolution, mitoribosomes differ in composition, function, and structure from ribosomes of their bacterial ancestors, but also among different species. Interestingly, they have been involved in human pathologies, including cardiomyopathies and developmental abnormalities, cancer, and hearing loss[384].

1.2.5.4 Ribosome heterogeneity

As we described previously, gene expression can be regulated at multiple levels (transcription, splicing, translation, mRNA and protein stability). Analysis of ribosome occupancy, mRNA and protein levels by genomics and proteomics technologies revealed that protein abundance is better predicted by ribosome occupancy than by transcript expression levels[385]. This suggests that translational control has an important role in the regulation of gene expression, particularly of a plethora of mRNAs underlying tissue-specific developmental processes[386]. Furthermore, we

have already described that translational regulation is an area of intense research that can be regulated through diverse mechanisms, including proteins binding to UTRs, uORFs and RNA structures within the mRNA.

Interestingly, recent work revealed that the ribosome itself has become a part of this regulatory system. Indeed, although ribosomes have been considered as uniform molecular factories with little regulatory capacity, an exciting literature on functional ribosome heterogeneity is extending rapidly. Thus, functional ribosome heterogeneity is defined as "*variations in ribosome composition that influence its activity, thereby changing the output of translation*"[290]. This results in the generation of "specialized ribosomes", which comprise specific protein and/or rRNA components and are tailored to translate specific groups of mRNAs[387]. Cell investment in translational regulation through ribosome modification could be beneficial. Indeed, this layer of translational regulation has the advantage to occur extremely quickly, while transcriptional regulation occurs within minutes. It can also allow the generation of different cellular outputs, depending on different stimuli[290].

The conceptual origins of ribosome specialization can be traced back to the 1950s, when Palade observed differences in ribosome size and shape and when Francis Crick discussed his "one gene-one ribosome-one protein" hypothesis. Later in 2002, the theory of "ribosome filter hypothesis" emerged. This model considers ribosomal subunits as regulatory elements or filters, which mediate interactions between specific mRNAs and components of the translation machinery, resulting in differential rates of mRNA translation[388]. It was then proposed the existence of a "ribosome code", analogous to "histone code", based on the discovery that different ribosomal proteins paralogs have distinct roles in specific translation regulation[389]. After the emergence of the concept of "specialized ribosomes" in 2011, numerous studies have identified different heterogeneous ribosomes across multiple species and model systems[387] (**Figure 23**).

Six sources of generating ribosome heterogeneity have been described: 1) substitution of RP paralogs, 2) differential stoichiometry of RPs, 3) ribosome associated factors, 4) post-translational modification of RPs, 5) rRNA variation and 6) rRNA modifications[387] (**Figure 24**).

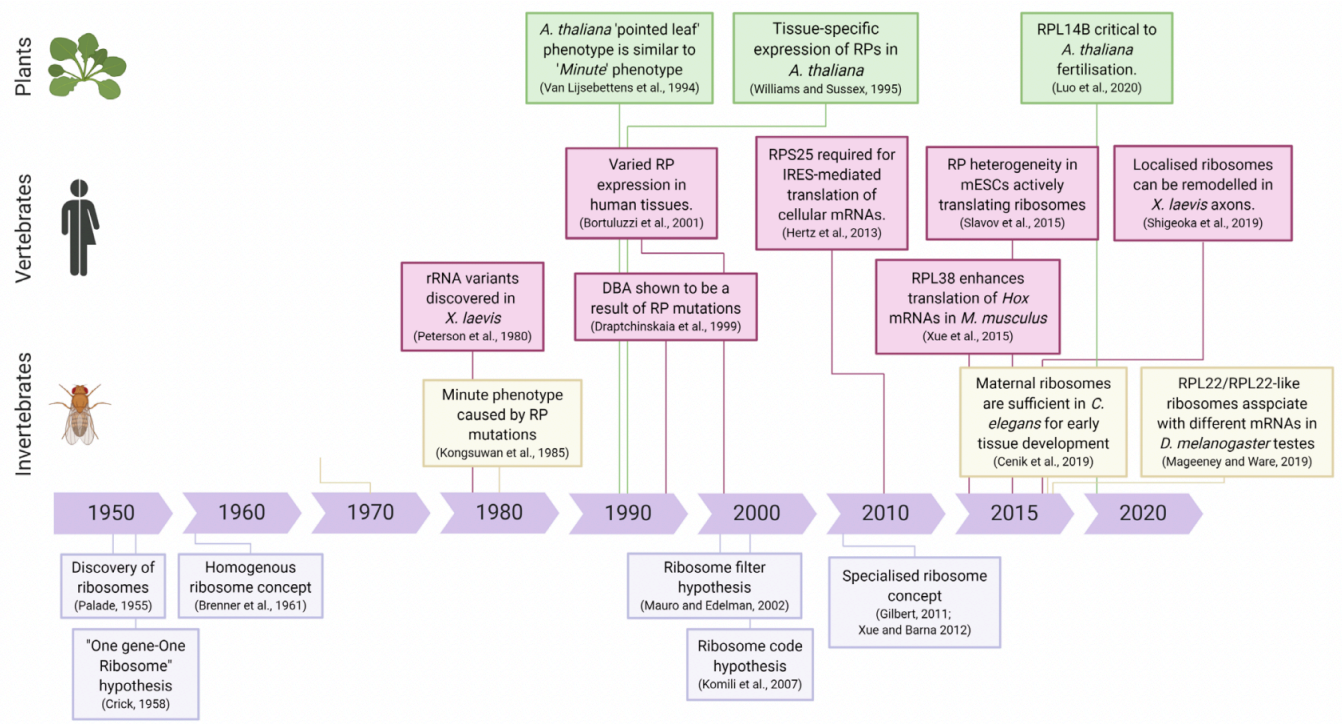


Figure 23: Timeline of evolution of the concept of specialized ribosomes. From[387].

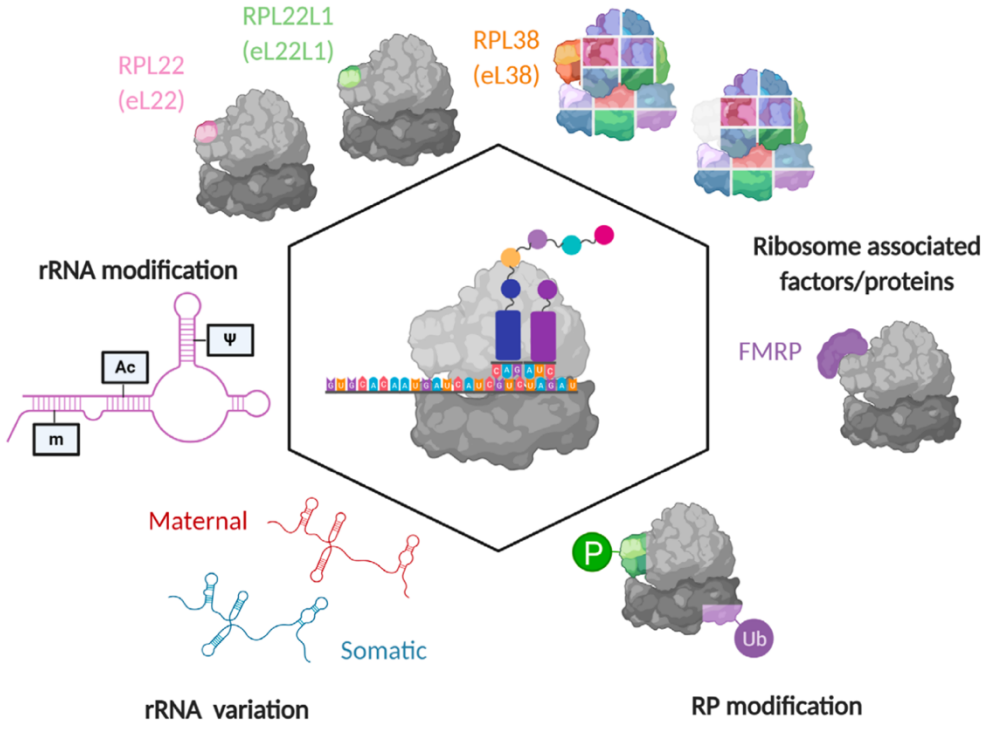


Figure 24: Schematic representation of different types of ribosomes heterogeneity. RP: ribosomal protein. From[387].

- **rRNA heterogeneity:**

The best example of rRNA heterogeneity was shown with the malaria parasite *Plasmodium*, expressing three different forms of rRNA over different phases of its life cycle[390]. rRNA heterogeneity can also arise from ribosome biogenesis (**see Figure 22**). In eukaryotes, the 5.8S, 18S, and 28S rRNA are processed from a single transcript, whereas the 5S is transcribed separately. Interestingly, rDNA loci are present in many hundreds of copies and different rRNA variant alleles can be expressed, which provides a source of ribosome heterogeneity[391].

Another layer of ribosome heterogeneity can be resumed by eukaryotic rRNAs expansion segments (ESs), which exhibit high sequence and length variability between species[392] and tissues[393]. A study reported that the maternal-type of 5.8S, 18S, and 28S rRNA sequences of zebrafish embryos, differ from those of the somatic-type. In addition, *in silico* analyses suggested that expansion segments (ES) in 18S rRNA, may interact preferentially with specific mRNA genes[394]. This would be a classic case of the regulatory model put forth by the ribosome filter hypothesis. Another study on ribosome-mRNA interaction showed that large expansion segments 7, 15, and 27 of human 28S rRNA present much higher density and numbers of mRNA matches than in other 28S rRNA parts[395]. Moreover, a structural report revealed an interaction between ES9S, on the 40S small ribosomal subunit and Hoxa9 5'UTR IRES-like element[396]. These experiments showcase the critical importance of mRNA-rRNA binding for accurate translational control.

Strikingly, rRNA can be post-transcriptionally modified, which constitutes another prominent source of ribosome heterogeneity. The two most abundant modifications that account for 90% of all rRNA modifications in human ribosomes[397] are: i) 2'-O-methylation (2'-O-me) of ribose (which can occur on any nucleotide) and ii) pseudouridylation (Ψ) (the isomerisation of uridine to pseudouridine). These 2'-O-methylations and pseudouridines are mostly installed by 2 classes of small nucleolar (sno)RNPs, termed C/D box snoRNAs (*SNORDs*) and H/ACA box snoRNAs (*SNORAs*), respectively[398]. Other identified rRNA modifications include acetylation and methylation of the nucleotide base. rRNA modifications distributions are not random and most of them occur in conserved regions of the two ribosomal subunits, which imply their influence on the structure and the function of the ribosome[399]. Indeed, recently, RiboMeth-seq demonstrated that dynamic changes in rRNA 2'-O-methylation in human cells, at distinct sites, occur in response to upstream signaling pathways, such as *MYC* oncogene expression[400]. In addition, Dyskerin (a pseudouridine synthase) mutation or knockdown resulted in reduced translational fidelity and IRES-mediated translation in yeast and mammalian cells[401]. Whilst these modifications are a clear source of ribosome heterogeneity, the precise mechanisms by which they may influence ribosome function remain unknown.

Finally, a growing number of evidence suggests that cancer cells harbor a specialized class of ribosomes, "onco-ribosomes", that confer preferential translation of oncogenic and prosurvival

genes, modulating cellular functions under cancer progression[402]. While mutations in ribosomal protein (RP) genes (*e.g.,* RPS15 (uS19) C'-terminal mutations in chronic lymphocytic leukemia[403]) are known drivers of oncogenesis, oncogenic rRNA variants have been shown to be implicated in the tumorigenesis. *Babaian et al.*, identified a cancer-specific single-nucleotide variation, affecting more than 45% of patients with colorectal carcinoma (CRC), in 18S rRNA at nucleotide 1248.U, at the decoding core of the ribosomal peptidyl (P) site. Particularly, this is the site of a conserved RNA modification, 1-methyl-3-α-amino-α-carboxyl-propyl pseudouridine ($m^1acp^3\Psi$). Therefore, a subset of CRC tumors is called hypo-$m^1acp^3\Psi$, unlike normal control tissues[404].

- **RP heterogeneity:**

Although the catalytic functions of the ribosome are conducted by the rRNA, most ribosomal proteins (RPs) are essential for its function: some RPs promote structural integrity, others present important roles in the binding of initiation, elongation, and release factors. However, functions of most RPs remain unclear[290]. Multiple studies have shown that varying the complement of core RPs can give rise to the formation of heterogeneous ribosomes. Indeed, ribosomes lacking specific RPs are present in cells. For instance, RP content from mouse embryonic stem cells (ESCs) varied in ribosomes isolated from distinct translational pools[405]. In addition, RNA regulons embedded in Homeobox (Hox) 5'UTRs facilitate ribosome recruitment and require RPL38 for their activity[406]. This study provided strong evidence that this regulation relied on RPL38's presence on the ribosome, which is important in lending support to the specialized ribosome hypothesis, as some RPs act in an extra-ribosomal capacity[407]. IRES translation can also be mediated by specialized ribosomes. In yeast and mammalian cells, RPS25 from the 40S subunit is essential to initiate translation from two viral IRESs: CrPV and HCV[408]. Subsequent work showed that cells lacking RPS25 present defective initiation from certain cellular IRESs[409].

Eukaryotic genome duplication created paralog pairs of RPs, showing high sequence similarity/identity but conferring vastly distinct effects upon cellular processes. Therefore, ribosomes can also be specialized through integration of different paralogs. Recently, a study, using translatome analysis, showed that yeast lacking *RPL1b* exhibited a downregulation of mitochondrial proteins, required for normal mitochondrial function. This function was not compensated by the paralogous *RPL1a*[410]. Another example is present in mice, where RPL22 is thought to down-regulate the expression of its paralog, Rpl22l1[411]. *Rpl22/eL22* knockout mice are viable but exhibit abnormal T cell development[412], whereas knockout of its paralogue *Rpl22l1/eL22l1* is embryonic lethal[413]. This suggests that different RPs are expressed at different times and places during development. Additional examples of RP paralog heterogeneity in vertebrates are discussed in[414].

So far, we have focused on RPs with the understanding that they are present in equimolar ratio (1:1 ratio) on the ribosome; implying that a subset of ribosomes in the cellular pool exhibits the same incorporation of distinct RPs. However, P proteins which are highly acidic RPs, are present in the stalk of the 60S subunit, formed by a pentameric complex, with two copies of RPLP1 and RPLP2 heterodimers bound to the C-terminal region of RPLP0[415]. Disrupting ribosomal P complex in human cells leads to stress-induced autophagy[416].

Another layer of heterogeneity can be introduced by post-translational modifications of RPs. Proteomics studies have identified several modifications in yeast, *Arabidopsis* and human cells[380]. These modifications include acetylation, methylation, hydroxylation, phosphorylation and O-GlcNAc. However, the functional significance of modified ribosomes was less described. For many years, the most well-known RP modification was the phosphorylation of RPS6, but its functional relevance remains poorly understood[417]. One recent example is the link between the phosphorylation of RPS15 and the pathogenesis of Parkinson's disease[418]. Another example of modulating protein synthesis is ubiquitination. Ribosomes with polyubiquitinated L28 present higher protein synthesis rate compared with ribosomes with monoubiquitinated L28[419].

Besides RPs and rRNAs composition, a plethora of factors can also bind to ribosomes and contribute to their heterogeneity. One of the most well-known ribosome-associated factors is FMRP that binds directly to ribosomes and translationally represses specific group of mRNAs by stalling ribosome translocation[420].

1.2.5.5 Ribosomopathies

Ribosomopathies are heterogeneous disorders caused by mutations in RP genes, rRNA processing genes, or ribosome biogenesis factors, leading to haploinsufficiency of the protein concerned, therefore, impacting ribosome abundance, function, or both. They can be inherited or sporadic disorders[289].

These conditions typically manifest themselves in specific cell and tissue types, strengthening the idea that ribosome specialization has tissue-specific importance. The mechanism(s) through which these mutations lead to disease manifestations remain an area of discussion. As expected, modest reduction in overall protein synthesis have been documented in a majority of studies[421]. This may lead to the possibility that impaired translation of global or specific mRNAs in certain tissues gives rise to tissue-specific phenotypes[422,423]. Other studies have suggested that ribosomopathy phenotypes occur independently of effects on protein synthesis and are not a consequence of ribosome specialization. Instead, they can be attributable to aberrant p53 activation. Recently, *Mills et al.* suggested a revised model (**Figure 25**) for tissue-specific phenotypes of RP mutations, governed by tissue-specific differences in the effectiveness of compensatory processes, rather than being simply and broadly explained by loss of specialized ribosomes or increased translation rates in affected tissues. They suggested that p53 activation

arises after such compensatory processes are overwhelmed, mediating cell cycle arrest and apoptosis[289].

Examples of ribosomopathies include Diamond-Blackfan anemia (DBA)[424], Shwachman-Diamond syndrome[425], Treacher Collins syndrome[426], chromosome 5q syndrome[427], North American Indian childhood cirrhosis[428], and isolated congenital asplenia[429].

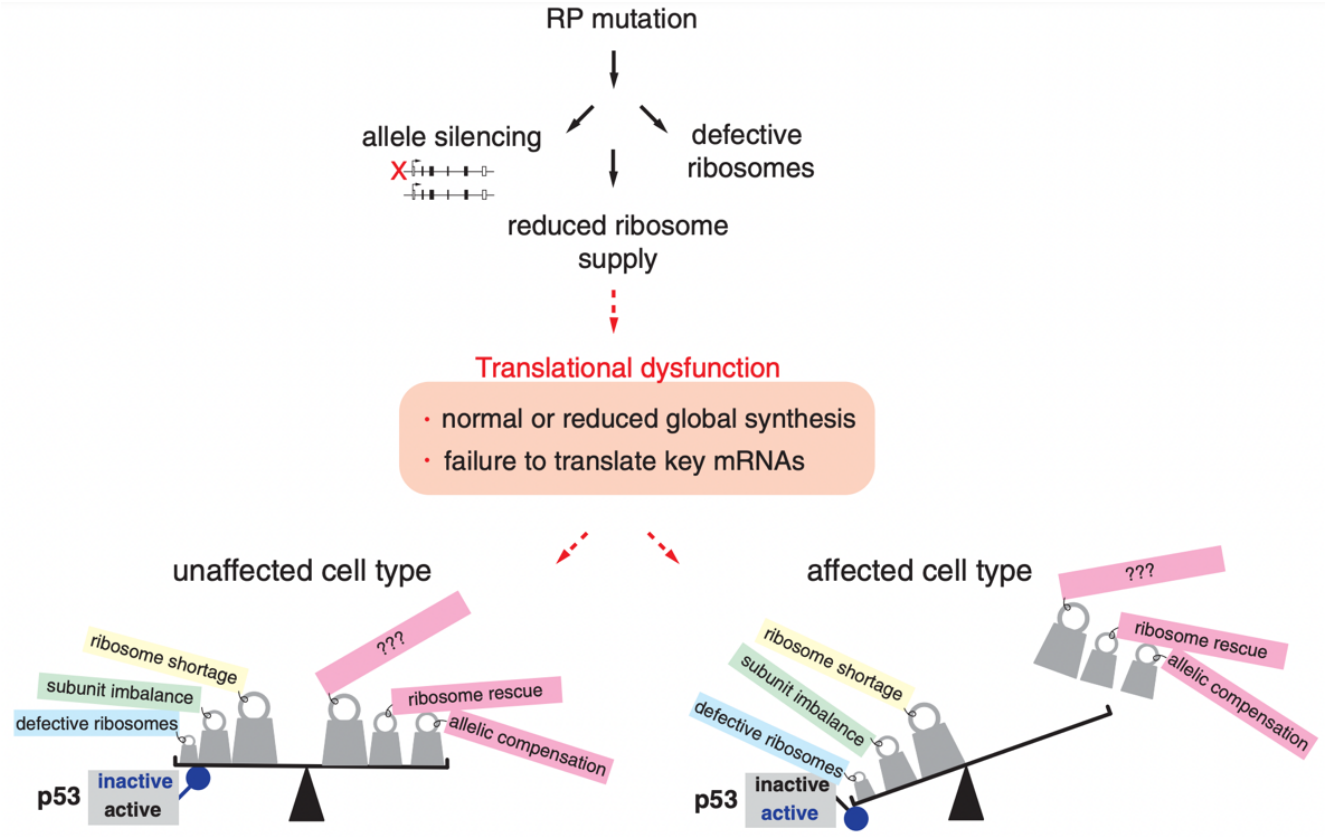


Figure 25: A revised model for tissue-specific phenotypes of RP mutations. RP mutations may cause allele silencing or defective ribosomes. Together these processes lead to translational dysfunction, involving reduced or normal global protein synthesis levels and failure to translate certain key mRNAs. The impact of this dysfunction is balanced in unaffected cells (left), by compensatory processes (pink 'weights'), like allelic compensation and ribosome rescue. However, in affected cells (right) balancing mechanisms are overwhelmed. From[289].

1.2.6 Methods of translatome profiling

As there has been an appreciation for the role of translation in the regulation of gene expression, development of methods inferring the translatome (*i.e.,* the entirety of mRNAs associated with ribosomes for protein synthesis) become mandatory. Here, we briefly review the four main methodologies for translatome analysis, with a focus on the ribosome profiling approach (**Figure 26).**

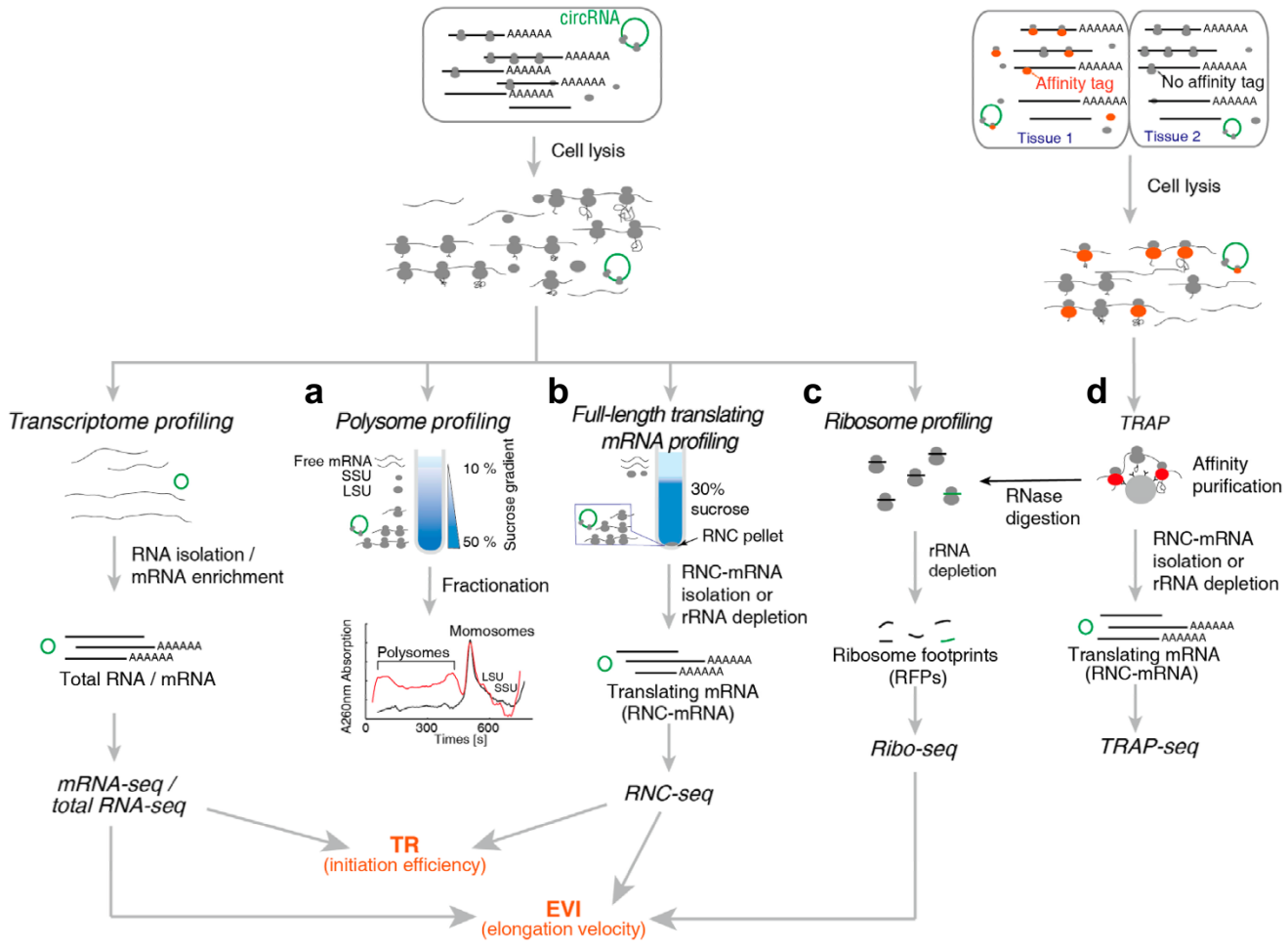


Figure 26: Main translatomic methods investigating the translatome. Overview of the principles of (**a**) polysome profiling, (**b**) ribosome nascent-chain complex-bound RNA sequencing (RNC-seq), (**c**) ribosome profiling and (**d**) translating ribosome affinity purification (TRAP). Translation ratio: TR, elongation velocity index: EVI, SSU: small subunit, LSU: large subunit. From[431].

1.2.6.1 Polysome Profiling

Polysome profiling, developed in the 1960s, is a technique of separation, based on sucrose gradient centrifugation, of actively translated mRNAs bound by several ribosomes (polysomes) from the "free" mRNA, the small (40S) and the large (60S) ribosomal subunits and the 80S monosomes. After sucrose gradient centrifugation and due to their density, mRNAs bound by more ribosomes sediment faster in the sucrose gradient, whereas free RNA and proteins float on the top of it[430]. The mRNAs in the fractions are then analyzed by northern blot or reverse-transcription quantitative polymerase chain reaction, or on a global level using cDNA microarrays, or with RNA-seq. In addition, polysome profiling allows monitoring of proteins associated with translation machineries by immunoblot-blot analysis and/or proteomics[431] (**Figure 26a**).

The technique was commonly used in studying translational changes. Indeed, analysis has been carried out to study diverse stress conditions and drug action. For example, it was used to investigate the translatome under different physiological conditions in yeast[432], but also in mammalian cells to interrogate the response to diverse stress conditions, like apoptosis[433], ER stress[434], hypoxia[435] and viral infection[436]. It also revealed that the antiproliferative effect of metformin is mediated by selective translational suppression of mRNAs encoding cell-cycle factors *via* the mTORC1/4E-BP pathway[437]. The main drawbacks of the polysome profiling include i) requirement of specialized and expensive equipment and large sample size, ii) contamination of polysomal fractions with non-ribosomal complexes and iii) necessity of more elaborate precipitation steps to isolate quality RNA for microarray/RNA-seq analysis[430].

1.2.6.2 RNC-seq

Profiling of full-length translating mRNA, or ribosome nascent-chain complex-bound RNA sequencing (RNC-seq), may better reflect protein synthesis and abundance[438]. Briefly, the technique starts by loading of cell lysate on 30% cushion, then all the translating mRNA bound to ribosomes are separated from free mRNA and other cellular components by ultracentrifugation. Finally, the RNC-mRNA can be recovered from the pelleted sedimented RNC. Higher RNC recovery rate can be reached by optimizing the centrifugation and the sucrose cushion. The technical difficulty of RNC-seq consists of separation of intact RNC, as they are fragile, which may lead to ribosome dissociation and mRNA degradation, and therefore to subsequent biased analyses of RNC-mRNAs[431] (**Figure 26b)**.

1.2.6.3 TRAP-seq

Ribosome affinity purification (RAP) or translating RAP (TRAP) is a unique approach, allowing to evaluate translational regulation in a cell-specific manner[439]. It requires the generation of genetically modified cells/organisms that contain an affinity-tagged ribosomal protein of the large

60S ribosomal subunit. Cell specific incorporation of the tagged RPs can be directed by a tissue-specific promotor, such as the Gal4-UAS system in *D. melanogaster* or the Cre-lox system in mice, which allows affinity tagging of the ribosomes of specific populations of cells. Cells are then collected and tagged ribosomes are purified by affinity selection (beads or columns), allowing capturing of only expressed ribosomes in the cells of interest. Finally, isolated RNA from the captured ribosomes can be quantitatively measured with qPCR, microarrays and/or RNA-seq[430,440] (**Figure 26d**).

In comparison to ribosome and polysome profiling, TRAP does not require large amounts of starting material. Interestingly, ribosomes isolated by TRAP-seq should not be contaminated with non-ribosomal mRNPs, as the technique does not require ultracentrifugation. However, TRAP-seq is time-consuming, costly and needs a stably transfected cell line to produce tagged ribosomes. Additionally, expression of tagged ribosomes can alter their structure and properties, thus the system is no more under physiological conditions[431].

1.2.6.4 Ribosome profiling (RIBO-seq)

Ribosome profiling, first described in 2009, is a deep-sequencing-based tool to monitor translation *in vivo,* at near nucleotide resolution[309]. The technique exploits the classical molecular method of ribosome footprinting and is based on deep sequencing of ribosome protected fragments (RPFs, *i.e.,* footprints of ~30 nucleotides) that persist after treatment of cell lysate with nuclease (typically, RNase I or micrococcal nuclease) (**Figure 26c**). This enables to define the exact location of translating ribosomes and to measure their densities along all RNA molecules present in a cell[441].

Ribosome profiling requires: 1) collection of a physiological sample, 2) inhibition of translation to freeze translating ribosomes, 3) nuclease digestion of ribosome-bound mRNAs, isolated by size, to generate RPFs, 4) isolation of ribosomes and ribosome footprints, 5) conversion of these ribosome footprints to a strand-specific library for next-generation sequencing and 6) mapping of these fragments to the appropriate reference genome[441].

Typically, ribosome footprints are characterized by precise positioning between the start and the stop codon of a gene. Changes in RPFs number can be used as a proxy to quantify translational regulation of the encoded protein. However, mRNA abundance of the transcript is likely to affect the probability of ribosome occupancy, and thus, translational regulation. This explains the importance of constructing parallel libraries of mRNA-seq for measuring mRNA abundance and determining the translational efficiency (TE) for each mRNA (**Figure 27a**). TE is defined as the number of ribosomes per transcript (*i.e.,* the ratio of the RPFs over mRNA counts within a gene's CDS). Therefore, differential translation efficiency genes (DTEGs), which are translationally regulated, are genes presenting changes in the number of RPFs independently of the changes in mRNA read counts between conditions (**Figure 27c**). However, differentially transcribed gene

(DTG), which are transcriptionally regulated, present concordant changes in RPFs and mRNA counts (**Figure 27b**). Beside DTEG and DTG, a gene can be regulated transcriptionally and translationally, thus, categorized as translationally intensified or buffered depending on the direction of the regulation[442] (**Figure 27d,e**).

Since the development of the technique by the Weissman lab in the yeast *Saccharomyces cerevisiae*[309], it became widely adapted to different cellular systems and organisms, including a variety of bacteria, yeast, parasitic protozoa, zebrafish, flies, nematodes, mice, rats, plants, viruses to understand translational control under diverse conditions. For instance, ribosome profiling approach was used to:

- study the translational control under oxidative stress or during yeast meiosis.
- decipher translational response of mammalian cells to heat shock and proteotoxic stress.
- investigate the impact of *trans*-acting factors (*e.g.,* RBPs and miRNAs) on translation[430,441].

Furthermore, ribosome profiling allowed the discovery of valuable translational features such as uORFs, non-AUG initiation codons, novel coding transcripts, codon usage bias, translational pausing landscape, measurement of elongation rates and the study of co-translational folding[431]. However, notable weaknesses and caveats should be considered when using ribosome profiling. It is costly and requires a large amount of starting material. The technique mainly analyzes the CDS, leaving the UTRs, implicated in the regulation of translation, with no efficient analysis. Recently, variants of ribosome profiling, such as translation complex profiling (TCP-seq) introduced in yeast[443] and ribosome complex profiling (RCP-seq)[444], have been introduced to capture footprints from all ribosome-associated mRNAs. Another drawback is the shortness of RPF sequences. This requires exaggerated sequencing throughput to obtain sufficient coverage of medium-abundance mRNAs[431].

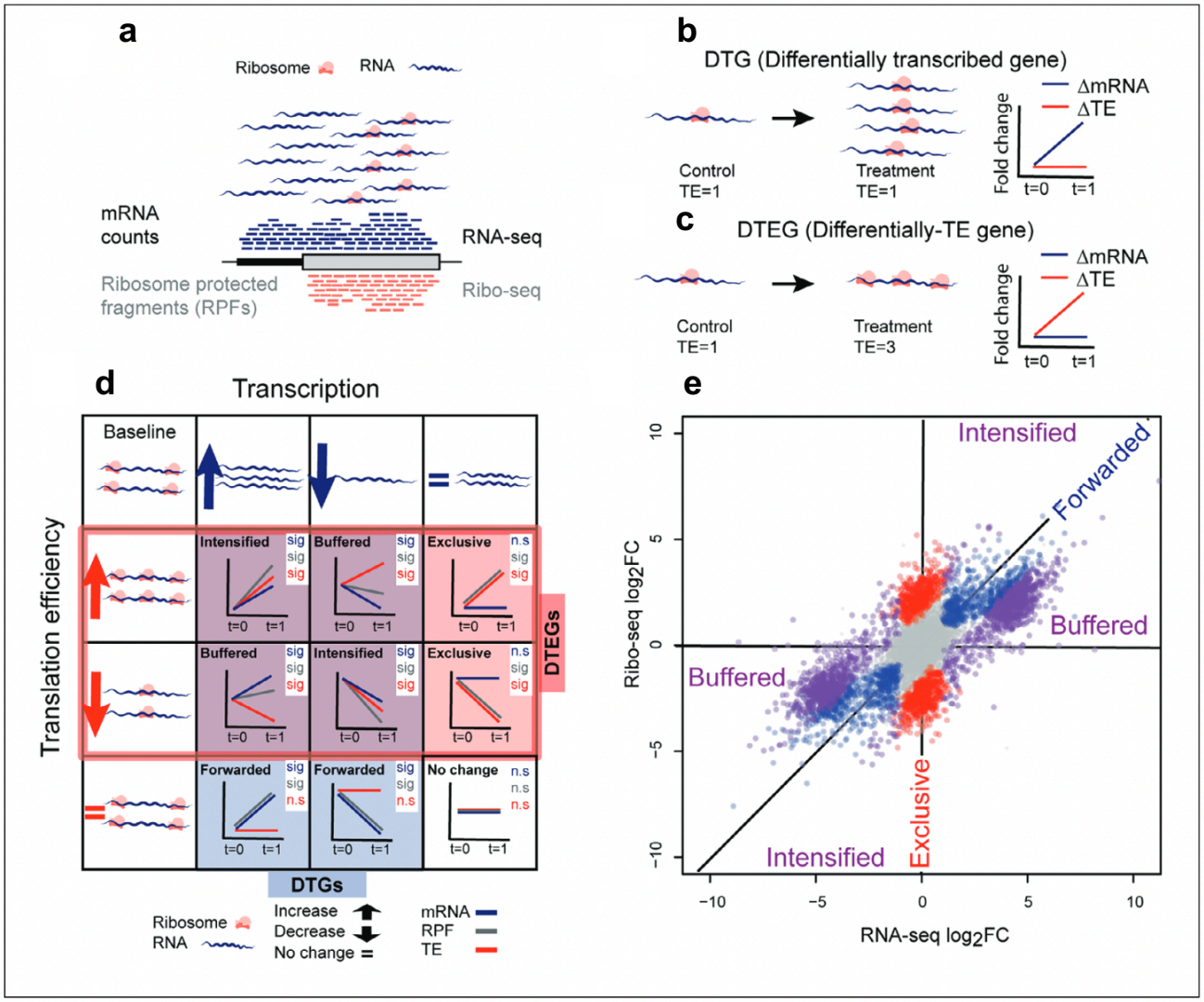


Figure 27: Transcriptional and translational regulation of gene expression. (**a**) Representation of the quantification of mRNA counts (using RNA-seq) and RPFs (using ribosome profiling). (**b**) Representation of differentially transcribed gene (DTG) with concomitant change in mRNA counts and RPFs and (**c**) differential translation efficiency gene (DTEG) with change in RPFs independent of change in mRNA counts, leading to a change in translation efficiency (TE). (**d-e**) Genes classification based on fold changes of RPF, mRNA, and TE. A gene can be either DTG, DTEG or both, based on the direction of change and can fall into one of the eight gene-regulatory possibilities. sig: significant, n.s.: not significant, RPF: ribosome protected fragment. From[442].

1.3 The endoplasmic reticulum (ER)

The endoplasmic reticulum is a continuous membrane system present in all eukaryotic cells and serves multiple functions including calcium storage, protein synthesis and lipid metabolism. The diversity of ER functions is performed by distinct domains; consisting of tubules, sheets and the nuclear envelope, each of which is associated with a specific function or functions. As the ER is an important site in protein synthesis and folding, it is equipped with a brunch of surveillance mechanisms to prevent misfolded proteins from transiting the secretory pathway and to ensure that persistently misfolded proteins are associated with a degradative pathway. Moreover, these processes are regulated by an adaptative intracellular signaling pathway known as the unfolded protein response (UPR^{ER}), which permits an appropriate adaptation of cells to proteins alterations. In addition to its role in protecting the ER, the UPR^{ER} has been presenting emerging critical physiological roles outside the realm of protein misfolding.

In this chapter, we will discuss several aspects of ER functions, structure and shaping and its contact sites with other organelles. Next, we will describe protein quality control systems associated with the ER, with a focus on the unfolded protein response signaling pathway. Finally, we will briefly cover some ER stress associated diseases, as well as UPR^{ER} modulating compounds.

1.3.1 ER functions

In 1902, Emilio Veratti made the most accurate description of a reticular structure in the sarcoplasm[445]. However, it took the scientific community another 50 years to rediscover the endoplasmic reticulum (ER). Since then, more advanced techniques, including three-dimensional electron tomography and confocal fluorescence microscopy, have revealed that the ER is a single continuous membrane-enclosed organelle, which is among the most architecturally striking of all eukaryotic organelles. It has a remarkable complex structure made up of the nuclear envelope (NE) and the peripheral ER. The latter consists of a network of tubules and sheets spread throughout the cytoplasm. This dynamic structure has a biological significance as it serves many roles in the cell, such as translation, protein synthesis and folding, calcium storage and lipid metabolism[446].

1.3.1.1 ER function in translation, protein synthesis and folding

There are two primary identified populations of ribosomes in eukaryotic cells: 1) free ribosomes dispersed through the cytoplasm, implicated in cytosolic protein synthesis, and 2) ER-bound ribosomes for protein synthesis of secreted and integral membrane proteins, but also of a subpopulation of cytosolic proteins[447]. Interestingly, considerable data now reveal multiple roles of the ER in mRNA translation, in addition of being a prominent site of mRNA localization (**Figure**

28). Indeed, half of all ribosomes and a similar fraction of the total mRNA are ER-associated in HeLa and HEK cells[381]. Moreover, ~75% of all cellular translation activity, in yeast, was found to be associated with the ER[448]. This indicates the magnitude of compartmentalization of mRNA translation to this structure. Therefore, the ER represents a biochemical environment (distinct from that of the cytosol) for translation, protein synthesis and folding, with numerous ER-localizing- regulatory factors and translational components, such as: translational regulators, ribonucleases, and RNA-binding proteins[447].

The translation of mRNAs encoding secretory, or membrane proteins initiates in the cytosol, then the mRNA-ribosome complexes are recruited to the endoplasmic reticulum membrane, with simultaneous emergence of a topogenic signal. Targeting of the ribosome-mRNA-nascent polypeptide chain complex is directed *via* the signal recognition particle (SRP) (**Figure 28a**) that binds to the hydrophobic core domain of topogenic signals, early in translation[449]. Then, the emerging polypeptide can enter the ER through the translocon. However, ribosomes engaged in the translation of cytosolic proteins, lacking topogenic signals, remain in the cytosol.
There is a confusing picture of ribosome binding to the endoplasmic reticulum. While SEC61 was accepted as the only ribosome receptor[450], substantial experimental evidence supports ribosome-binding activity for many ER proteins. One example of receptors is that of p180[451], an ER membrane protein, identified as an important protein for both ribosome binding and protein translocation (**Figure 28b**,**c**).

If the protein is destined to be an integral membrane protein, it will be anchored within the phospholipid bilayer where it remains, whereas proteins destined to enter the secretory pathway or the lumen of membrane-bound organelles, will be directed for transport[452]. In addition, following their protein synthesis and translocation into the ER lumen, secretory proteins undergo proper folding and modifications, including N-linked glycosylation, disulfide bond formation and oligomerization[453].

To protect themselves against the accumulation of misfolded and mistargeted proteins in the early secretory pathway, eukaryotic cells have developed extensive quality control processes that center on the endoplasmic reticulum. Canonical pathways include ER-associated degradation (ERAD) and ER-phagy[454], as will be further discussed below (**"1.3.5 Protein quality control systems of the ER"**).
Besides its role in protein synthesis, an emerging body of work demonstrated that the ER is also the central site for siRNA processing and silencing activity[455] (**Figure 28d**).

Overall, these studies reveal a fundamental role of the ER in mRNA translation for both secreted and non-secreted proteins, and that this membrane system is a critical hub of post-transcriptional gene regulation.

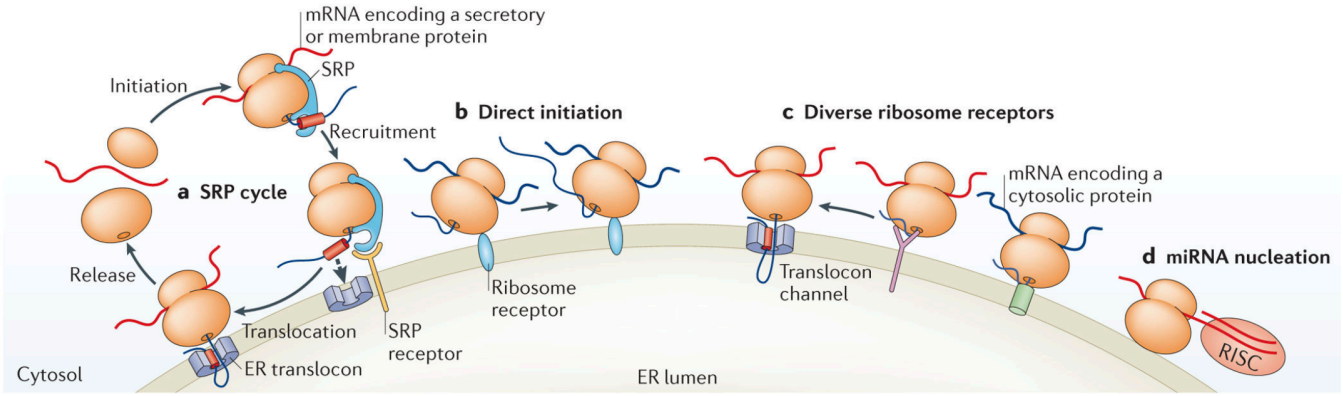


Figure 28: The implication of the ER in RNA translation. (**a**) Cytosolic ribosomes are targeted to the ER *via* the signal recognition particle (SRP) pathway. (**b**) Many mRNAs, like those encoding secretory and cytosolic proteins undergo direct initiation on stably ER-bound ribosomes. (**c**) Shown are diverse ribosome receptors. (**d**) The ER is involved in miRNA-mediated silencing. From[447] .

1.3.1.2 ER function in lipid synthesis

In addition to its role in protein synthesis, the ER is also a site for lipid biogenesis. It has key roles in membrane production, lipid droplet/vesicle formation and fat accumulation for energy storage. Lipid synthesis occurs at membrane interfaces and organelle contact sites (**see below; "1.3.4 Membrane contact sites between ER and organelles"**). Interestingly, the membrane structure of the ER changes regularly to adapt to the changing cellular lipid concentrations. The ER produces several types of lipids, such as phosphatidylcholine (PtdCho) and phosphatidylethanolamine (PtdEtn), as well as less abundant membrane lipids, such as phosphatidylinositol (PtdIn) and basic sphingolipid structures[456]. In addition, the ER ensures cholesterol homeostasis, as it contains the sterol regulatory element-binding protein family of cholesterol sensors[457].

1.3.1.3 ER function in calcium (Ca^{2+}) metabolism

Ca^{2+} binds to thousands of proteins, therefore it is involved in many intracellular and extracellular signaling networks, crucial for gene expression, protein synthesis, cell proliferation, differentiation, metabolism, and apoptosis[458]. The ER is the major store of intracellular Ca^{2+}, thus it regulates Ca^{2+} levels, and reciprocally many ER functions are controlled in a Ca^{2+}-dependent way[459] (**Figure 29a**). The ER contains several calcium channels/pumps[452] to regulate Ca^{2+} levels:

- Inositol 1,4,5-trisphosphate (IP3) receptors (IP3R), responsible for releasing Ca^{2+} from the ER into the cytosol, when intracellular levels are low.
- Ryanodine receptors (RyRs) that bind Ca^{2+} in response to increased cytoplasmic levels of Ca^{2+}.

- Dyhydropyridine receptors (DHPRs), which are voltage dependent Ca^{2+} channels that activate RyRs, leading to Ca^{2+} release.
- Sarcoendoplasmic reticular Ca^{2+} ATPases (SERCAs) that pump back Ca^{2+} into the ER.
- STIM1 proteins cluster in regions of ER apposition to the plasma membrane, after the depletion of luminal Ca^{2+}. At these regions, clustered STIM1 traps Orai1 subunits and assembles them into active Ca^{2+} release-activated channels (CRAC), allowing for uptake of extracellular Ca^{2+} into the ER lumen[452] (**Figure 29b**).

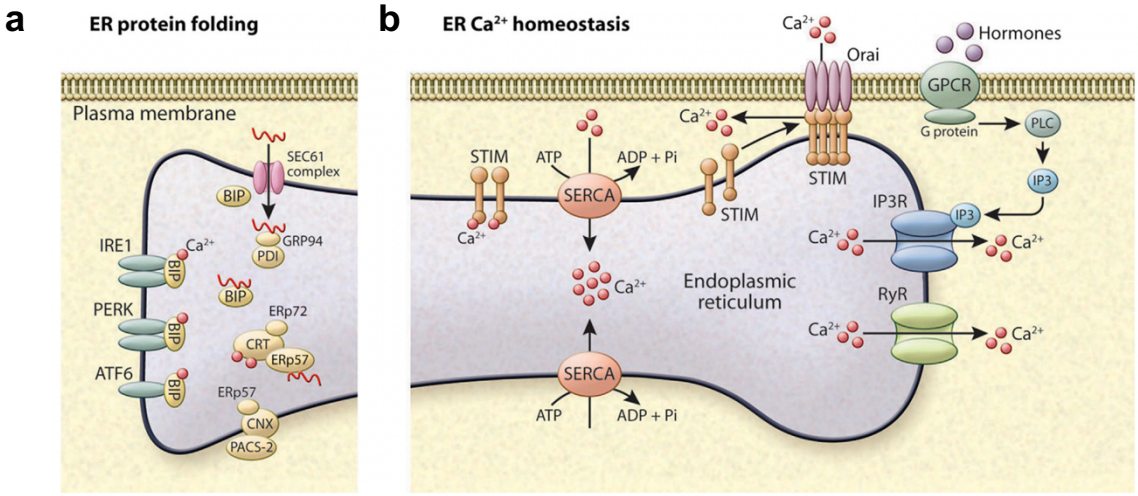


Figure 29: Homeostasis of Ca^{2+} in the ER. (**a**) The nascent polypeptide is translocated through the SEC61 pore. Chaperone assistance is provided by lectin chaperones such as calnexin (CNX) and calreticulin (CRT), and the ATP-driven chaperone BiP (GRP78). Calnexin and calreticulin bind Ca^{2+} with high capacity. (**b**) Toolkit implicated in the ER Ca^{2+} homeostasis. IP3R: inositol 1,4,5-trisphosphate (IP3) receptors, RyR: ryanodine receptor, Ca^{2+}: calcium. From[460].

1.3.2 ER structure

The two major domains of ER are 1) nuclear envelope (NE) and 2) cytoplasmic cisternae and tubules, which form the interconnected peripheral ER. The NE is a distinct domain of the ER, made up of two flat ER membrane bilayers, which stack to form the inner and outer nuclear membranes (INM and ONM), separated by the perinuclear space (PNS)[461]. Various molecules, including RNAs and proteins, are transported through hundreds of nuclear pores spanning the ONM and INM of the NE. Many mechanisms go into shaping the nuclear envelope (NE), including linker proteins between the INM and the ONM, nuclear pores, and the cytoskeleton[462]. On the other side, the peripheral ER branches out of the ONM into the cytosol, all the way to the plasma membrane (PM), as an extensive network of cisternae and tubules (**Figure 30**).

- **<u>Peripheral endoplasmic reticulum domains:</u>**

The peripheral ER is an interconnected network of two domains: flat sheet-like cisternae, and a polygonal array of tubules. An ER sheet presents a consistent luminal spacing of ~30 nm in

yeast and 50 nm in animal cells and exhibits low membrane curvature[463]. In contrast, ER tubules are long cylindrical units, spread throughout the cytoplasm, with high membrane curvature in cross-section. The morphology of the ER was originally classified by electron microscopy, which found that sheets tended to be “rough” (due to the presence of membrane bound ribosomes), thus they form the rough ER (RER), whereas tubules were largely devoid of ribosomes and therefore “smooth”[464], forming the smooth ER (SER). The SER is often more convoluted than RER that tends to be more granular in texture (**Figure 30a**). The RER performs functions associated to translation, translocation, post-translational modification, biosynthesis of membrane and secretory proteins and their folding[465]. However, the functions of SER are less understood, but they can be associated with several processes like lipid synthesis and signaling between the ER and other organelles[463] (**Figure 30b**). Both sheet and tubular domains are present in all eukaryotes, but their relative amounts and organizations vary depending on cell type, which reflects the different functions of these cells[463]. For example, cells with high capacity to secrete proteins, such as pancreatic secretory cells and B cells are made up of sheets, whereas cells involved in lipid synthesis, calcium signaling and sites of contact for other organelles, such as neurons, muscle cells, and epithelial cells, possess an ER composed of tubules[452].

An additional configuration of the peripheral ER includes cortical ER, which refers to regions of the peripheral ER that are closely apposed to the plasma membrane (PM). This displays an intermediate phenotype between sheets and tubules, with regions that are flat and others with high curvature[463]. Interestingly, the cortical ER presents a hybrid ribosome density: there is no bound ribosomes on the side that faces the PM, whereas the side facing the cytosol has ribosomes[466]. These contact sites are very important for exchanging small molecules such as lipids and signals. For example, in budding and fission yeast, the ER is closely apposed to ~40% of the PM[466]. Another example is that of muscle cells, having invaginations of the PM, called T-tubules that are closely opposed to the ER. At these contact sites, calcium signaling occurs during muscle contraction[467].
Overall, it is evident that the type and amount of ER structures, as well as the ratios of these structures to one another, reflect the function of a particular cell type. Further studies are needed to uncover the signaling pathways that determine the biogenesis of each ER type in a specific cell type.

Strikingly, recent super-resolution microscopy studies discovered the nanoscopic ER structures, characterized by 10- to 100-nm internal length scales. The four ER membrane nanostructures, imaged in considerable detail, include: 1) ER tubular matrices which are densely packed tubular arrays, previously thought of as flat sheets[468] (**Figure 31a**), 2) ER sheet nanoholes that are circular dynamic pores of about 100 nm diameter, coexisting with uniform sheet regions[469] (**Figure 31b**), 3) internal membranes of ER Exit Sites (ERES) (**Figure 31c**), and 4) ER transport intermediates[470]. Recently, a common physical mechanism of shaping of all these currently

known ER nanostructures was proposed, based on two factors: 1) membrane curvature mechanisms and 2) ultra-low membrane tensions[471].

Taken together, these studies challenge the dogma that the peripheral ER consists only of two distinct morphologies (flat sheets and curved tubules), but they also help to address other key issues regarding ER function in healthy cells and during disease pathogenesis.

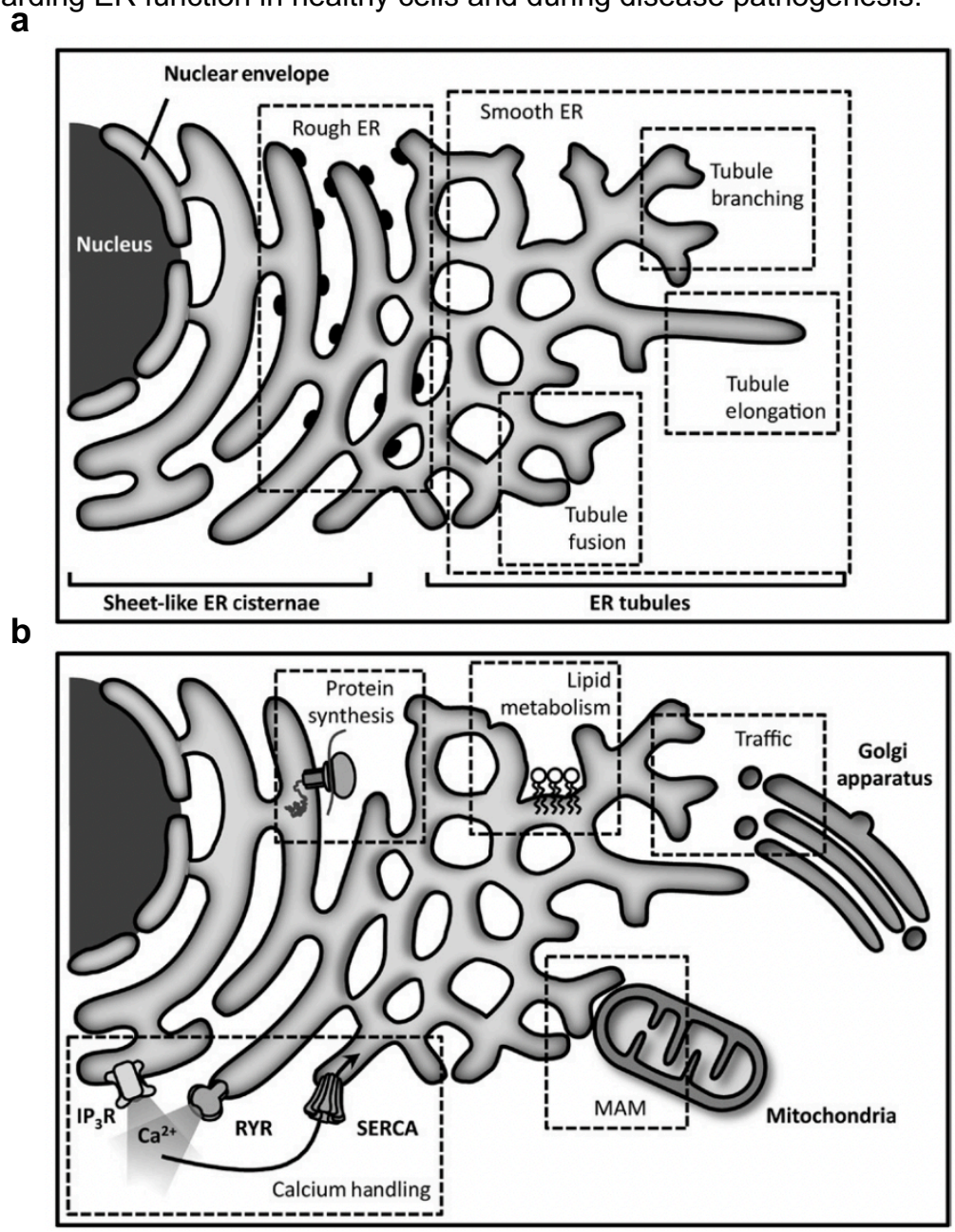


Figure 30: Structures and functions of the endoplasmic reticulum (ER). (**a**) Subdivision of the ER in three domains: the nuclear envelope, the sheet-like ER, the tubular ER. The sheet-like ER rich in ribosomes is called rough ER, whereas the tubular ER with fewer ribosomes is commonly called smooth ER. The latter can elongate, fuse and branch inside the cell. (**b**) The diverse functions of the ER in the cell: calcium homeostasis, lipid and protein synthesis, quality control and degradation of proteins, cell trafficking and regulating the functions of other organelles, such as mitochondria. MAM: Mitochondria-associated ER membrane, SERCA: sarcoendoplasmic reticular Ca^{2+} ATPase, RyR: ryanodine receptor, IP3R: inositol 1,4,5-trisphosphate (IP3) receptors. From[711].

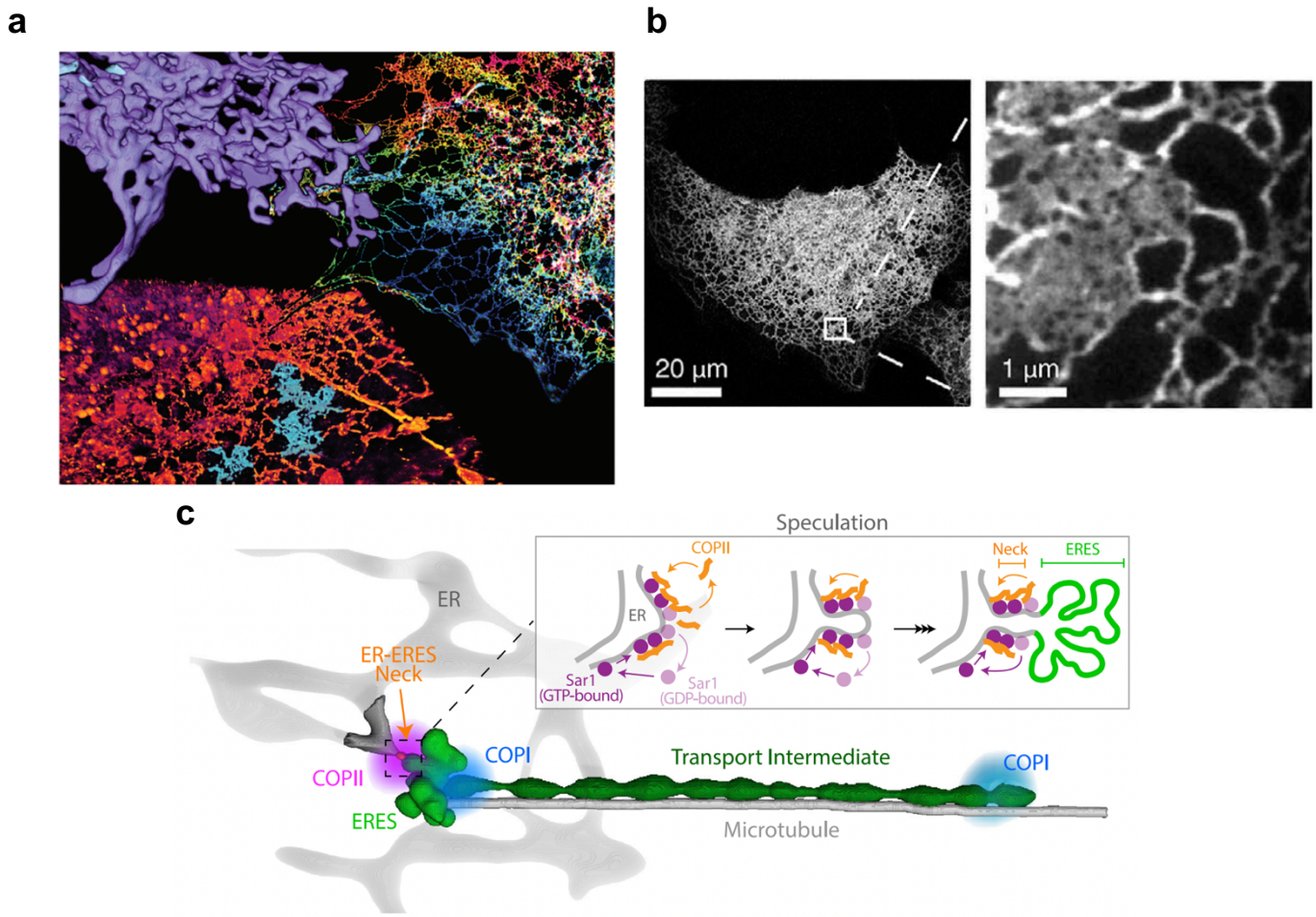


Figure 31: Nanoscopic level of organization of the ER. (**a**) Super-resolution imaging showing the peripheral ER, formed of densely clustered tubule-shaped structures. Upper right = an ER protein marker, lower left = internal cellular lipids, and upper left = electron microscopy reconstruction of tubular matrices. From[712] . (**b**) ER sheet nanohole image showing confocal (left) and Stimulated Emission Depletion Microscopy (STED) of magnified boxed region (right) images of ER sheets and tubules in mammalian cells. From[469]. (**c**) Model showing the formation of Endoplasmic Reticulum Exit Sites (ERESs), and the distributions of COPII and COPI and ERES associated transport carrier. ERES are presented as intertwined tubules in continuity with the ER by a narrow neck. This tubular structure has a diameter of ~350-390 nm, with COPII placed closer to the neck and COPI nearer to its rims, as these two proteins orchestrate the secretory pathway. When the secretory cargo is released into ERESs, their overall size is enlarged. From[470].

1.3.3 ER shaping and remodeling

1.3.3.1 *Formation of ER tubules*

Several proteins that promote ER tubules shape have been identified, perhaps the reticulon family of proteins (Rtns) is the most well-studied group of proteins[472]. The Rtns (four genes in mammals) contain a reticulon homology domain (RHD), composed of two tandem hydrophobic

segments. These proteins contribute to the bending of the membrane, which displaces lipids in the outer leaflet of the bilayer leading to curvature of the membranes. Overexpression of some reticulon isoforms leads to more ER tubules at the expense of sheets, whereas their deletion or depletion has the opposite effect[473]. Tubule promoting factors also include 1) proteins with tandem transmembrane hairpins (TMHs), such as ADP-ribosylation factor-like 6 interacting protein 1 (Arl6IP1) and family with sequence similarity 134, member B (FAM134B)[474], 2) the small GTPase Rab10[475], 3) members of the DP1/Yop1/REEP5/6 and 4) REEP1-4 family[476]. Interestingly, an amphipathic helix (APH) C-terminal to Yop1p RHD is also important for tubule formation and may insert into membrane as an additional wedge[477] (**Figure 32a**).

1.3.3.2 Formation of ER Sheets

Sheets can be formed through three mechanisms. First, membrane at the edge of ER sheets presents a curvature similar to that of a tubule, assuming that curvature of sheets and tubules is likely generated in the same way and with same proteins[478]. Second, Climp-63, localized exclusively in sheets, bridges the two apposed membranes using its luminal coiled coil (CC) domain, which leads to a fixed thickness of an ER cisterna[479]. Third, the sheet-enriched integral membrane proteins kinectin and p180 are likely important to maintain flattened the surface of ER sheets[478] (**Figure 32b**).

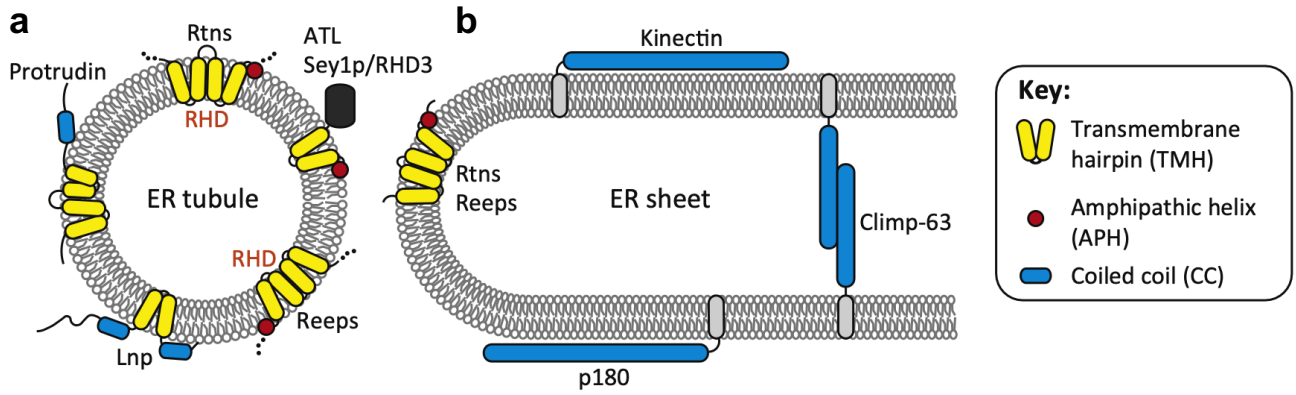


Figure 32: ER-shaping determinants shown in either tubules or sheets. Representation of molecules implicated in ER tubules (**a**) or ER sheets (**b**) formation. ER: endoplasmic reticulum, RHD: reticulon homology domain, Rtns: reticulon family of proteins. From[480].

1.3.4 Membrane contact sites between ER and organelles

A whole new field of investigating the molecular mechanisms, cell biology, physiological and pathological implications of close proximities between organelles defined as “membrane contact sites (MCSs)”, is now emerging. This field gained momentum when several examples of functional apposition between membranes became obvious[481]. One example of well-studied contacts involved the ER, the most abundant membrane compartment of the cell, which forms

several MSCs with multiple membrane systems: the ER-mitochondria, ER-PM (plasma membrane), ER-Golgi, ER-peroxisomes, and ER-lipid droplets (LDs) contacts[480] (**Figure 33**). These contacts involve protein-protein interactions and/or protein-phospholipid interactions and are dynamic in term of shapes and areas to correlate with the functional demands of the contact.

A typical contact site is formed of resident proteins (termed tethers), effectors, and regulators. Tethers are the physical basis for the contact sites, effectors fulfill the functional roles and regulators mediate the characterization of contact sites[482].

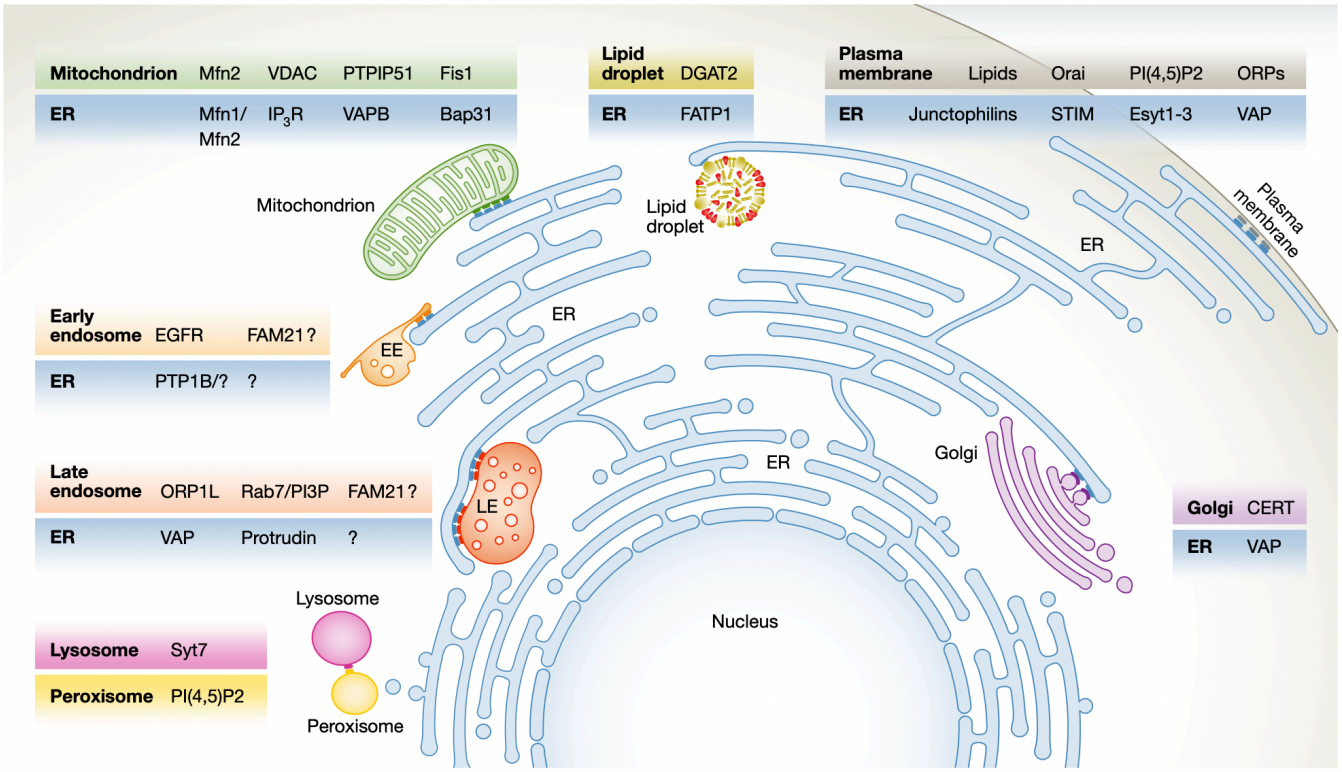


Figure 33: ER contact sites in mammalian cells. Several examples of proteins mediating these contacts are shown. From[483] .

1.3.4.1 ER-plasma membrane contact sites

The ER makes extensive contacts with the plasma membrane (PM), providing an ideal platform for non-vesicular transport of signaling molecules. Many pairs of interactions have been identified for ER-PM contacts. For instance, upon Ca^{2+} depletion in the ER, STIM1 oligomerizes and becomes activated, leading to its accumulation at ER-PM contacts to bind and trap Orai1. This binding is important to open Orai1 channels, allowing extracellular Ca^{2+} to flow into the cell. STIM1 can also bind with PI(4,5)P2 on the PM to enhance contact formation[480]. Additionally, numerous lipid transfer proteins can mediate ER-PM tethering, and facilitate exchange of lipids, including glycerolipids, phospho-inositides, and cholesterol. Examples of lipid transfer proteins include extended synaptotagmins (ESyts), TMEM24, oxysterol-binding protein (OSBP)-related

proteins (ORP) 5 and ORP8, and GRAM domain-containing proteins (GRAMDs). This high number of molecules mediating the ER-PM contacts have revealed the importance of the communication between the ER and the PM, in many physiological processes such as: the regulation of intracellular Ca^{2+} dynamics and signaling, refilling of intracellular Ca^{2+} stores *via* store-operated Ca^{2+} entry (SOCE), and the control of lipid traffic and signaling[484].

1.3.4.2 ER-mitochondria contact sites

In the early nineties, when purifying mitochondria, researchers found that mitochondrial fractions were consistently “contaminated” by ER-derived membranes. This was the first evidence for an intimate relationship between the ER and the mitochondria[485]. Subsequent studies showed that mitochondria-ER contact sites (MERCSs) are abundant in various tissues and cell types and are hubs for the exchange of metabolites. Indeed, they play critical roles in various biochemical and signaling functions like Ca^{2+} homeostasis, lipid transfer, and regulation of organelle dynamics[486]. Additionally, both ER and mitochondria harbor enzymes to synthesize lipids, thus apposition of ER and mitochondria is a logical strategy to facilitate lipid transport through the aqueous cytosol[487]. MERCSs have also been suggested to define sites of mitochondrial DNA replication and mitochondrial fusion[488] and constitute a platform for autophagosome biogenesis[489]. In order to regulate this wide variety of functions, several ER-mitochondrial tethers are recruited to these unique contact sites. Indeed, in eukaryotic cells, a large variety of proteins regulating MERCS formation have been identified, such as Mfn1/Mfn2, Bap31 and VDAC[486].

1.3.4.3 ER-Golgi contact sites

ER-Golgi contact sites are important to regulate lipid homeostasis and trafficking. Indeed, a lipid gradient across the ER-Golgi MCS is orchestrated by various families of lipid transfer proteins, including oxysterol-binding proteins (OSBPs) that deliver sterol to Golgi and transfer PI(4)P for degradation by Sac1[490]. Additionally, CERT factor transports ceramide from the ER to the Golgi for sphingomyelin synthesis, in a non-vesicular manner[491].

1.3.4.4 ER-endosomes contact sites

Recent studies have revealed the existence of numerous contact sites with numerous functions, between the endoplasmic reticulum and endosomes in mammalian cells. Interestingly, some contact sites share common features, such as containing 1) VAP-A, an ER protein, as a structural component or as a targeting factor, 2) cholesterol-binding proteins and 3) confirmed or putative phosphoinositide-binding domains[492]. ER-endosome contact sites are important for endosome positioning, cholesterol transfer, receptor dephosphorylation, endosome fission, negative control of endosome fusion and contact sites in Ca^{2+} transfer[493].

1.3.4.5 ER-lipid droplets contact sites

Numerous identified proteins are enriched or exclusively located at the interfaces between lipid droplets (LDs) and the ER. These proteins have fulfilled a broad range of functions, including LD biogenesis, and maintenance of lipidic connections between LDs and ER[494].

1.3.5 Protein quality control systems of the ER

One-third of the eukaryotic proteome (*e.g.,* secretory and membrane proteins) is synthesized in the endoplasmic proteome[495]. Therefore, to maintain ER proteostasis, several dedicated protein quality control systems working in concert are employed to monitor protein biogenesis. These systems are summarized in **Figure 34** and include 1) ER chaperones, 2) the unfolded protein response (UPR^{ER}), 3) ERAD, and 4) ER-phagy.

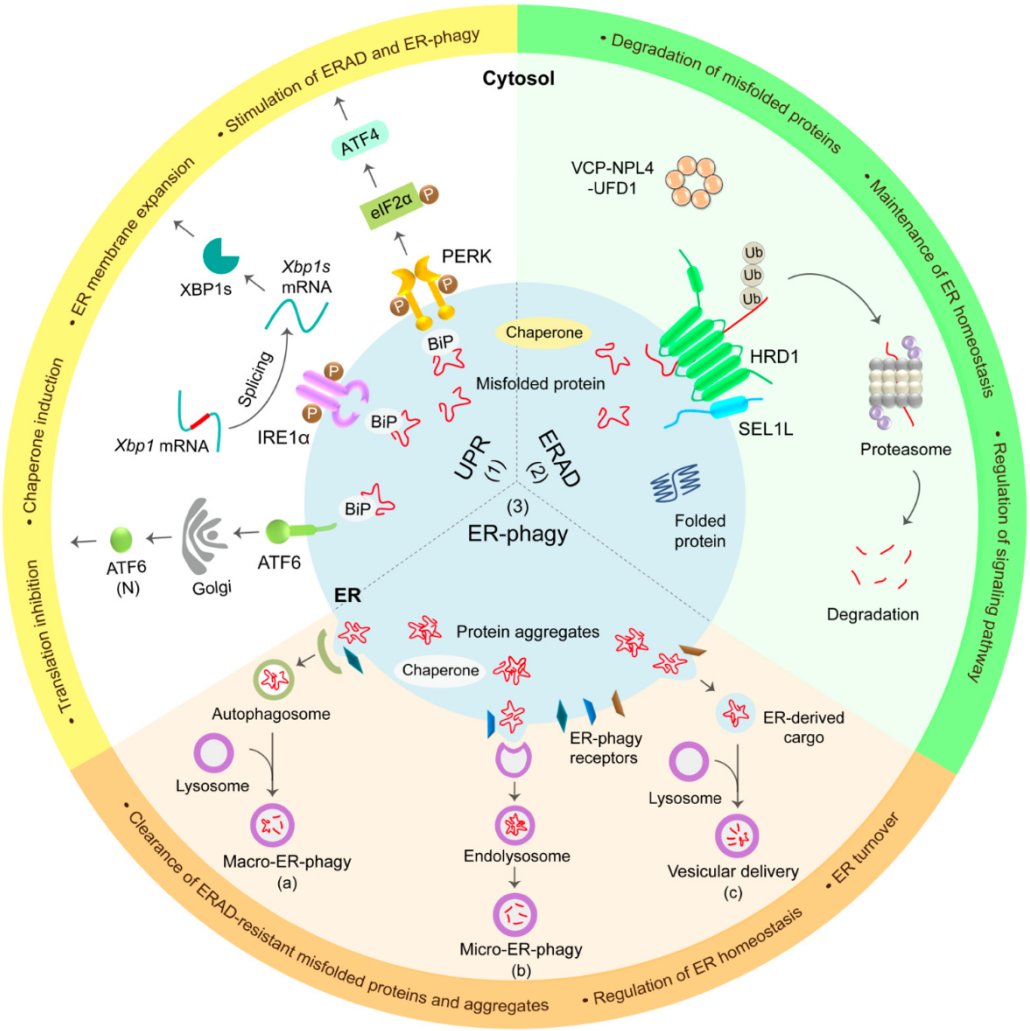


Figure 34: Overview of mammalian ER protein quality control mechanisms. Unfolded protein response (UPR^{ER}), ER-associated degradation (ERAD) and ER-phagy systems are depicted. From[495].

1.3.6 ER chaperones

1.3.6.1 ER chaperones implicated in protein folding

Newly synthesized proteins moving into the ER undergo post-translational modifications and folding to achieve their native conformation with the help of numerous chaperones and folding enzymes (**Figure 35**). They include members of both Hsp70 and Hsp90 families of molecular chaperones that are present in the ER[495]. For instance, binding immunoglobulin protein (BiP), a heat shock protein (HSP) 70 chaperone family member, is one of the most abundant ER chaperones playing a critical role in protein folding in the ER. BiP substrate binding domain (SBD) interacts with the exposed hydrophobic or aggregation-prone regions of client proteins, whereas its nucleotide binding domain (NBD) binds and hydrolyzes ATP to control the substrate-binding affinity. BiP assists folding *via* ATP-dependent cycles of binding and release, where ADP-bound BiP exhibits a high affinity for client proteins, thus allowing their efficient folding and shielding them from aggregation. BiP activity is further regulated by diverse co-chaperones of the HSP40 family of ER-localized DnaJ (called ERdjs) and HSP110 family of nucleotide exchange factors (NEFs). In addition to its role in protein folding, BiP has been implicated in other quality controls such as UPR^{ER}, ERAD, and ER-phagy[454]. Another abundant ER-resident Hsp90 that mediates ATP hydrolysis, is Grp94, required for the folding of a select pool of substrates[496] (**Figure 35a,b**).

1.3.6.2 ER chaperones implicated in glycosylation

Glycosylation is a ubiquitous modification of newly synthesized proteins, playing important roles in the regulation of protein folding and quality control[497]. The biosynthesis of the lipid-linked oligosaccharide (LLO) starts on the cytosolic side of the ER membrane by OST (oligosaccharyltransferase), as the nascent polypeptide exits the translocon. Glycoproteins then undergo trimming by ER luminal glucosidases I and II to form monoglucosylated glycan (Glc1Man9GlcNAc2), which is the substrate for the ER lectin chaperones Cnx (calnexin) and its soluble orthologue Crt (calreticulin). Calnexin and calreticulin, combine a lectin-like glycan-binding domain with a flexible arm, the P-domain that recruits other chaperones, to regulate quality control of glycoproteins in the endoplasmic reticulum[498]. Once the terminal glucose residue is trimmed, calreticulin and calnexin are released, and folded glycoproteins can be exported from the ER. However, if a protein is not properly folded, the UGGT (UDP-glucose:glycoprotein glucosyltransferase) catalyzes the re-addition of glucose to a nearby glycan, subjecting the monoglucosylated protein to additional calreticulin/calnexin-mediated folding cycles[499] (**Figure 35d,e**). Other ER lectin chaperones include osteosarcoma amplified 9 (OS9), XTP3-transactivated gene B protein (XTP3-B), and ER degradation-enhancing mannosidases (EDEM) that participate in eliminating misfolded glycoproteins *via* the ERAD mechanism[495] (**Figure 35f**).

1.3.6.3 ER chaperones implicated in formation of disulfide bonds

The formation of disulfide bonds between cysteine residues is also important for the proper folding and function of ER proteins and is catalyzed by the protein disulfide-isomerase (PDI) family. More than twenty mammalian oxidoreductases of the PDI family have been identified with multiple structural features and enzymatic activities, mediating them to function in several aspects of ER quality control mechanisms, such as UPR^{ER} signaling and ERAD. For the formation of disulfides, PDI can be oxidized by ER-resident peroxidases such as the thiol oxidase Ero1 (ER oxidase 1) and PrxIV (Peroxiredoxin IV)[454] (**Figure 35c**).

Overall, molecular chaperones are the primary mediators of ER quality control. They help polypeptides to fold and evaluate the conformations of their substrates.

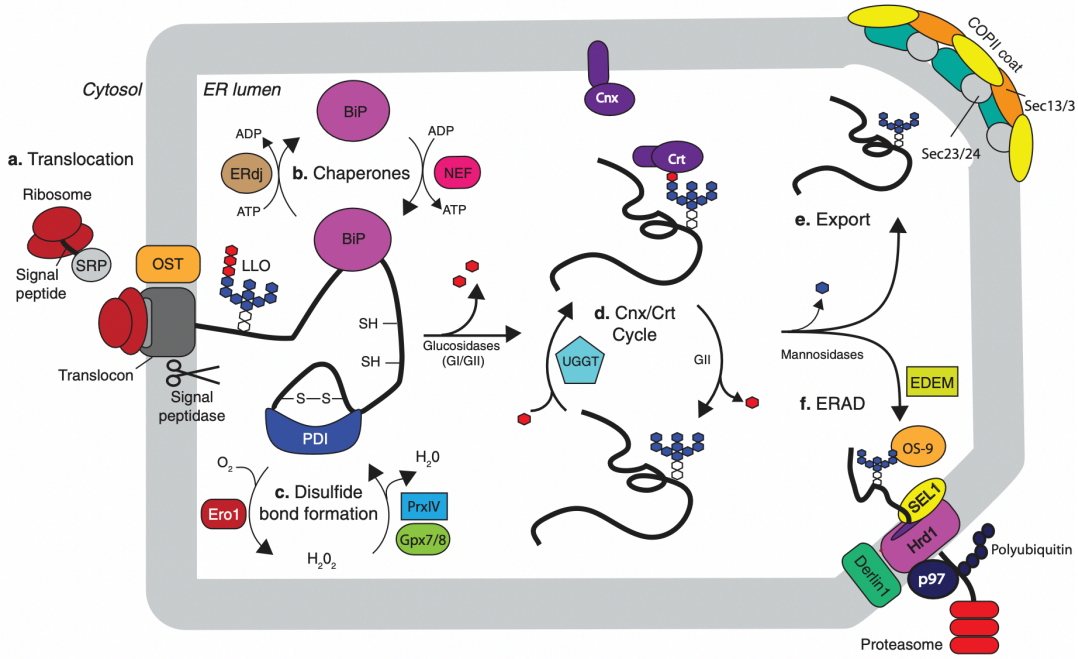


Figure 35: ER chaperones and foldases implicated in protein quality control. (**a**) Protein translocation. (**b**) Chaperones. (**c**) Disulfide-bond formation. (**d**) Calnexin (Cnx)/calreticulin (Crt) cycle. (**e**) Protein export. (**f**) ER-associated degradation (ERAD). From[454].

1.3.7 The unfolded protein response (UPR^{ER})

1.3.7.1 Key players of the UPR^{ER}

Perturbation of ER homeostasis caused by Ca^{2+} depletion, hypoxia, altered glycosylation or viral infection leads to the accumulation of unfolded proteins, which activates the unfolded protein response in the ER, UPR^{ER500}. In mammals, the UPR^{ER} has evolved into a complex network of signaling events that target various cellular responses. UPR^{ER} is mediated by the activation of

at least three classes of sensors of ER stress: inositol-requiring enzyme 1α (IRE1α) and (IRE1β), protein kinase RNA-like ER kinase (PERK) and activating transcription factor 6 (ATF6; both α and β isoforms) (**Figure 36**). Each unique sensor includes i) an ER luminal domain that senses aberrant conditions within the ER lumen and/or membrane compositions, ii) an ER-transmembrane domain, and iii) a cytosolic domain that transfers information to transcriptional and translational machineries. The most plausible explanation of such elaborate complexity of vertebrate UPR^{ER} pathways in comparison to yeast, which is defined by a single linear pathway initiated by Ire1p and the downstream transcription factor Hac1p[501], is that an expanded UPR^{ER} provides greater flexibility. Therefore, a wider range of inputs can be accommodated by the multiple overlapping and distinct pathways, leading to multiple cellular outputs that finely tune cellular needs[502].

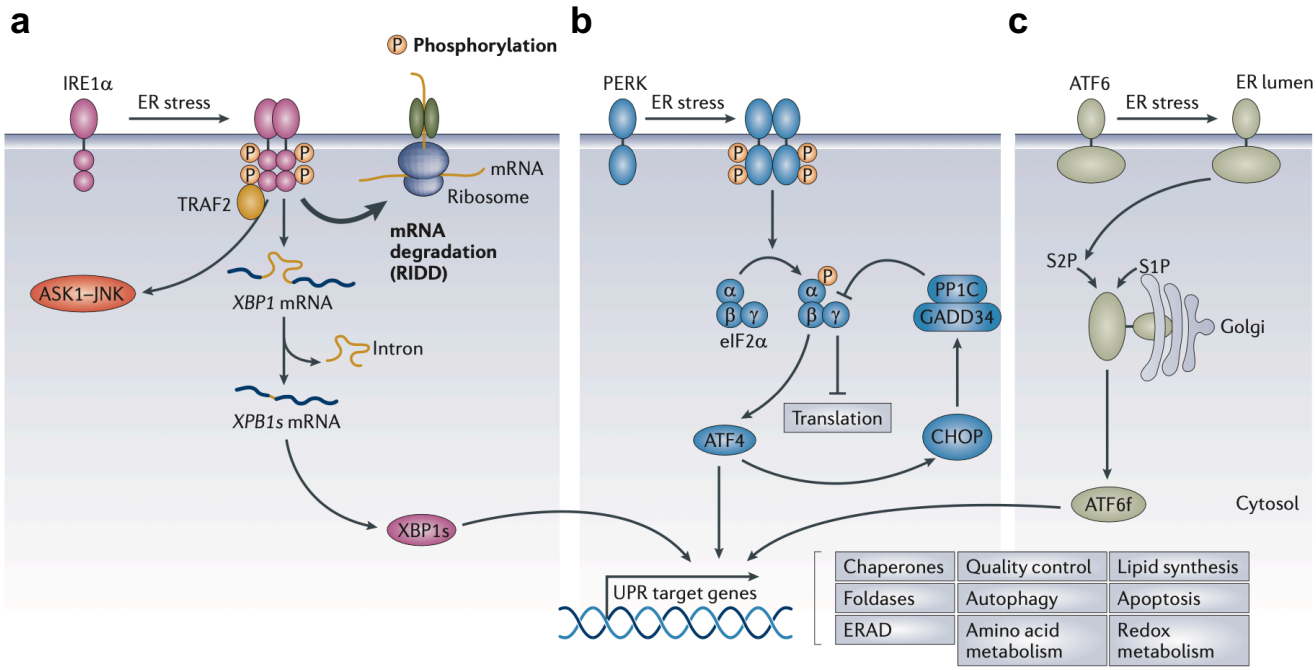


Figure 36: Schematic representation of the three pathways of unfolded protein response (UPR^{ER}) and their downstream effects. (**a**) IRE1, (**b**) PERK and (**c**) ATF6 pathways. From[510].

- **IRE1**:

IRE1 branch is the most evolutionarily conserved UPR^{ER} branch. Both human IRE1 isoforms (IRE1α and β) share a structural similarity consisting of an N-terminal ER luminal sensor domain and a C-terminal cytosolic effector region. The latter contains both kinase and endoribonuclease (RNase) domains. Although structural similarity, IRE1α and IRE1β display differential activities. Interestingly, it has been reported that human IRE1β mediates the site-specific cleavage of 28S rRNA and translational attenuation of protein synthesis[503]. Furthermore, IRE1α is expressed

ubiquitously whereas its paralog, IRE1β, is restricted mainly to the gastrointestinal tract and the pulmonary mucosal epithelium[504].

IRE1α has two distinct enzymatic activities, mediated by cytosolic kinase and RNase domains. Upon activation, under ER stress conditions, IRE1α dimerizes and trans-autophosphorylates (**Figure 36a**). The phosphorylation in the activation loop of IRE1α kinase domain, is not only necessary to activate its cytosolic RNase domain but is also important to initiate the recruitment of tumour necrosis factor receptor-associated factor 2 (TRAF2) and JNK pathway signaling. IRE1α's RNase domain induces a selective cleavage of dual stem loops within the X-box binding protein 1 (XBP1) mRNA. Therefore, it catalyzes the excision of 26 nucleotides intron from XBP1 mRNA and produces a frameshift that allows the translation of a longer isoform called spliced XBP1 (XBP1s). XBP1s is a basic leucine zipper (bZIP) transcription factor that translocates to the nucleus to direct the transcription of a wide range of targets including the expression of chaperones, foldases and components of the ER-associated protein degradation (ERAD) pathway, leading to relieve ER stress and restore its homeostasis[504]. Moreover, XBP1s modulates phospholipid synthesis, which is required for ER membrane expansion under ER stress[505].

The IRE1α nuclease can also act on a broader range of substrates through a mechanism called regulated IRE1-dependent decay (RIDD)[506]. RIDD is a conserved mechanism in eukaryotes by which IRE1α cleaves transcripts at a defined consensus sequence (CUGCAG) accompanied by a stem-loop structure[504]. It is required to maintain ER homeostasis by reducing the load of newly synthesized peptides entering the ER or by promoting apoptosis[507].

- **PERK:**

PERK is an ER-resident transmembrane protein with an ER luminal domain as well as a cytoplasmic kinase domain. Like IRE1α, PERK also oligomerizes and trans-autophosphorylates, when activated upon sensing ER stress. Once activated, PERK phosphorylates the ubiquitous translation initiation factor eIF2α on serine 51, which inhibits eukaryotic translation eukaryotic initiation factor 2B (eIF2B) and thereby downregulates global protein synthesis. This helps to reduce the flux of protein entering the ER in order to alleviate ER stress. However, eIF2α phosphorylation allows translation of a specific set of mRNAs containing short open reading frames in their 5'untranslated regions (5'UTRs)[508]. One of these encodes the transcription factor ATF4, playing key roles in autophagy, amino acid metabolism and the synthesis of stress-induced proteins[258] (**Figure 36b**).

The 5'UTR of ATF4 contains two upstream open reading frames (uORFs), the second of which overlaps the start codon of the protein-coding sequence. Under normal conditions, when eIF2-GTP is abundant in non-stressed cells, uORF1 is translated and initiation at uORF2 is efficient. Therefore, the peptide encoded by uORF2 will be translated with high efficiency, preventing the

translation of ATF4. However, during stress conditions, eIF2α is phosphorylated and the levels of ternary complex (TC) are reduced, resulting in leaky scanning of the 40S ribosome subunit, and therefore the ternary complex bypasses the uORF and allows ATF4 translation[509] (**Figure 37**). Notably, ATF4 regulates the expression of genes implicated in the apoptosis such as the pro-apoptotic factor CCAAT/enhancer-binding protein (C/EBP) homologous protein (CHOP) and growth arrest and DNA damage-inducible 34 (GADD34)[510].

- **ATF6:**

ATF6 is a type II transmembrane protein that belongs to an extensive family of leucine zipper proteins. In humans, it is encoded by two different genes: *ATF6A* for ATF6α (670 amino acids) and *ATF6B* for ATF6β (703 amino acids). Biochemical and physiological characteristics of ATF6α are significantly better documented than ATF6β. Although the two ATF6 paralogs share high homology, ATF6α has been described as a potent transcriptional activator, whereas ATF6β is a poor transcriptional activator that may inhibit activation by ATF6α[511].

ATF6 is cleaved, through regulated intermembrane proteolysis (RIP) process by releasing its cytoplasmic domain, which then regulate gene expression by entering the nucleus[500]. Indeed, after activation in the ER, ATF6 is transported to Golgi where it is cleaved by two Golgi-resident proteases membrane bound transcription factor peptidase: site-1 protease (S1P) and site-2 protease (S2P), leading to ATF6 cytosolic N-terminal portion (ATF6f) that enters the nucleus and induces UPR^{ER} gene expression[504] (**Figure 36c**). ATF6f comprises a transcriptional activation domain (TAD), a bZIP domain, a DNA-binding domain and nuclear localization signals[512]. Among others, ATF6α induces the expression of genes involved in ERAD machinery, ER and Golgi biogenesis, chaperones and UPR^{ER} mediators including BiP and X-box binding protein 1 (Xbp1), which contribute to maintain proteostasis and to increase the regulatory output of the IRE1 arm, respectively[511].

Additional ER-resident bZIP transcription factors with ill-defined functions are also involved in ER stress signaling, such as Luman, cAMP responsive element-binding protein 3 like 1 (OASIS), cAMP responsive element-binding protein 3 like 2 (BBF2H7), CREB3L3 and CREB[504].

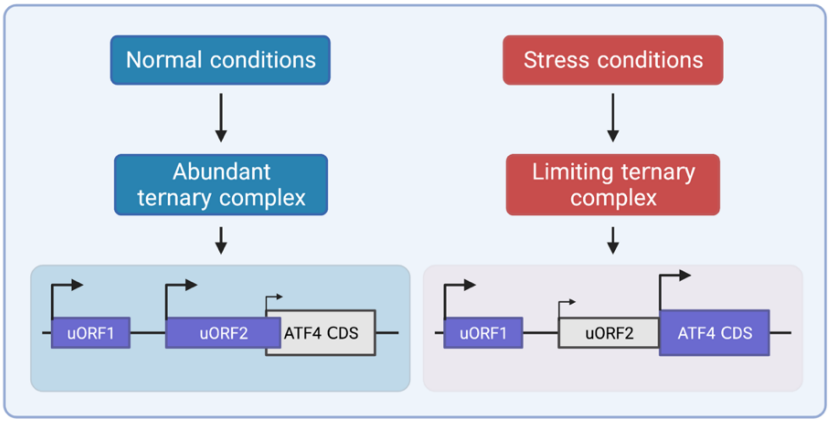


Figure 37: Upstream open reading frames (uORFs) translation regulation of ATF4 under normal and stressed conditions. CDS: coding sequence. Created with Biorender.com.

1.3.7.2 UPR^{ER} and apoptosis

As described in the previous section (**see above; "1.3.7.1 Key players of the UPR^{ER}"**), the UPR^{ER} alters the transcriptional and translational programs, to deal with the accumulation of aberrant or misfolded proteins. Indeed, two distinct waves of adaptative cellular responses are induced by UPR^{ER}. As an immediate reaction, i) inhibition of general translation by PERK, ii) selective degradation of mRNA encoding for certain ER-located proteins induced by IRE1 through the RIDD mechanism, and iii) ER-phagy reduce the influx of proteins into the ER to re-establish its homeostasis. The second wave triggers a massive gene expression response through the regulation of UPR^{ER} transcription factors, leading to the upregulation of UPR^{ER} target genes: i) IRE1α activates XBP1s, implicated in ER-associated degradation (ERAD) and ER membrane expansion under ER stress, ii) ATF6f directly controls genes encoding ERAD components and XBP1, and iii) ATF4 controls the levels of genes related to redox balance, amino acid metabolism and protein folding. If these numerous UPR^{ER}-induced mechanisms fail to alleviate ER stress, both intrinsic and extrinsic apoptosis pathways can become activated.

Physiological processes demanding high rate of protein synthesis and secretion sustain UPR^{ER}'s adaptive programs activation without leading to cell death pathways (**see below**; "**1.3.7.7 Roles of UPR^{ER} in physiological processes"**). However, above a certain threshold, conditions of irreversible ER damage results in apoptosis and involve a series of complementary pathways[513]. The B cell lymphoma 2 (BCL-2) protein family is important for the regulation of ER stress-induced apoptosis[514]. Once activated, BCL-2 homology 3 (BH3)-only proteins, such as BCL-2-interacting mediator of cell death (BIM) and p53, regulate the activation of BAX and/or BH antagonist or killer (BAK) to trigger apoptosis[515]. Sustained PERK signaling upregulates the pro-apoptotic transcription factor C/EBP-homologous protein (CHOP), leading to the downregulation of the anti-apoptotic protein BCL-2 and inducing the expression of some BH3-only proteins and DNA

damage-inducible 34 (GADD34)[513]. Therefore, reactive oxygen species (ROS) induced by GADD34, in addition to altered calcium homeostasis owing to inositol-1,4,5- trisphosphate receptor (IP3R) activation, may open the mitochondrial permeability transition pore (PTP), which promotes apoptosis[516]. Moreover, CHOP, ATF4, and p53 can regulate the expression of a subset of BH3-only proteins[515]. IRE1α may also lead to apoptosis by activating JNK and RIDD pathways[517]. Other complementary mechanisms have also been proposed to induce cell death under ER stress, including the participation of Caspase 2 in cleaving the BH3-only protein BH3-interacting domain death agonist (BID), which activates BAK and BAX[516] (**Figure 38**).

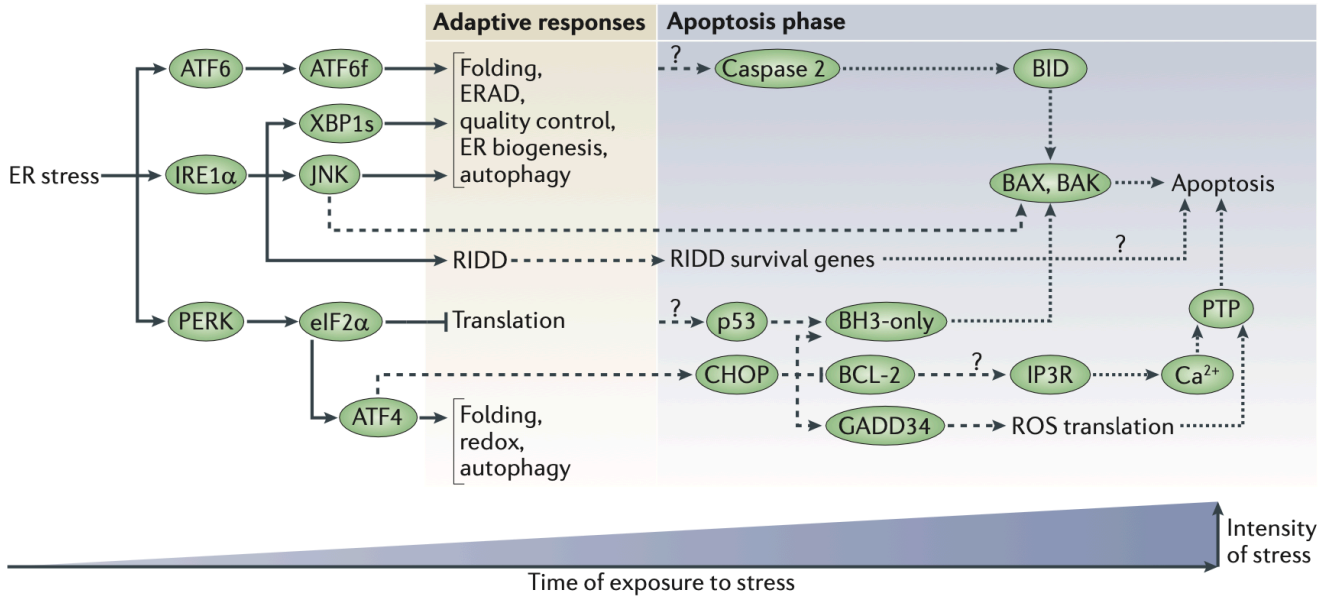


Figure 38: Distinct cell fate decisions observed under ER stress. Cells can either adapt to ER stress or undergo to apoptosis phase depending on duration and intensity of ER stress. Dashed arrows indicate steps of transition from adaptive responses to apoptosis. Dotted arrows indicate events that mediate apoptosis. Question marks indicate an unclear mechanism. From[516].

1.3.7.3 Sensing proteotoxic ER stress

Three established models have been reported to explain how misfolded proteins, referred to as proteotoxic stress, induce the UPR^{ER}: i) direct association, ii) BiP competition and iii) BiP allosteric models[518] (**Figure 39**). As IRE1 is the most conserved UPR^{ER} branch, it will be illustrated to describe these 3 models.

- **Direct association model:**

This model postulates that misfolded proteins bind directly to the luminal domain (LD) of IRE1, which mediates conformational changes that result in the oligomerization of IRE1 LD, thereby activating UPR^{ER} signaling[518] (**Figure 39a**). This model emerged from the crystal structure of the conserved core region of the LD that resembles the peptide-binding groove of major

histocompatibility complexes (MHCs)[519]. Additionally, mutation of residues within this groove leads to impairment of IRE1 signaling, in yeast [519] and a recent structural and biochemical analysis showed that human IRE1 LD can bind to both peptides and unfolded proteins *in vitro*[520].

The role of BiP in this model is to fine-tune the activity of IRE1 sensor. Indeed, a study suggested that mutation of BiP binding site displayed reaction kinetics, consistent with BiP acting as a buffer of IRE1 activity in yeast[521].

A recent crystal structure of PERK LD, bound to misfolded peptide suggests that its oligomerization is dependent on the direct binding of unfolded proteins to activate the UPR^{ER} [522]. Moreover, PERK LD had flexible binding regions that have been reported for several molecular chaperones. This suggests that the peptide binding groove of PERK can vary according to the size of its peptide ligand, in order to bind a wide range of unfolded proteins[523].

- **BiP competition model:**

In this model, BiP binds the luminal domain of IRE1, as a chaperone-substrate type interaction *via* its substrate binding domain (SBD) to form a repressive complex[524]. The formation of this complex, mediated by ERdj4, stimulates BiP ATPase activity, resulting in ERdj4 dissociation and causing IRE1 LD to form monomers, leading to the repression of UPR^{ER} signaling[525]. However, under high ER stress, BiP and ERdj4 are occupied with misfolded proteins, which impedes BiP association with IRE1 LD[525]. This allows IRE1 to form dimers, which in turn activates UPR^{ER} signaling. Therefore, BiP SBD is the same site of binding to misfolded proteins and IRE1 LD, resulting in a competition for this site, which is the central tenet of the “competition” model. In addition to its binding to IRE1, BiP also binds to PERK and ATF6 in unstressed cells but dissociates from them during acute ER stress[524].

Overall, according to this model regulated by nucleotide binding, BiP acts as a negative regulator of UPR^{ER} signaling, as its binding sequesters the UPR^{ER} sensors in an inactive state, whereas its dissociation upon ER stress licenses their activation (**Figure 39b**).

- **BiP allosteric model:**

The allosteric model indicates an interaction independent of nucleotides, between the nucleotide binding domain (NBD) of BiP and IRE1 LD. Under ER stress, misfolded proteins bind exclusively to the canonical BiP SBD, which leads to UPR signaling by a dissociation of BiP NBD from IRE1 LD[526]. As BiP presents two different sites of binding for misfolded proteins and IRE1 LD, this model is not competitive but rather tends to be allosteric (**Figure 39c**).

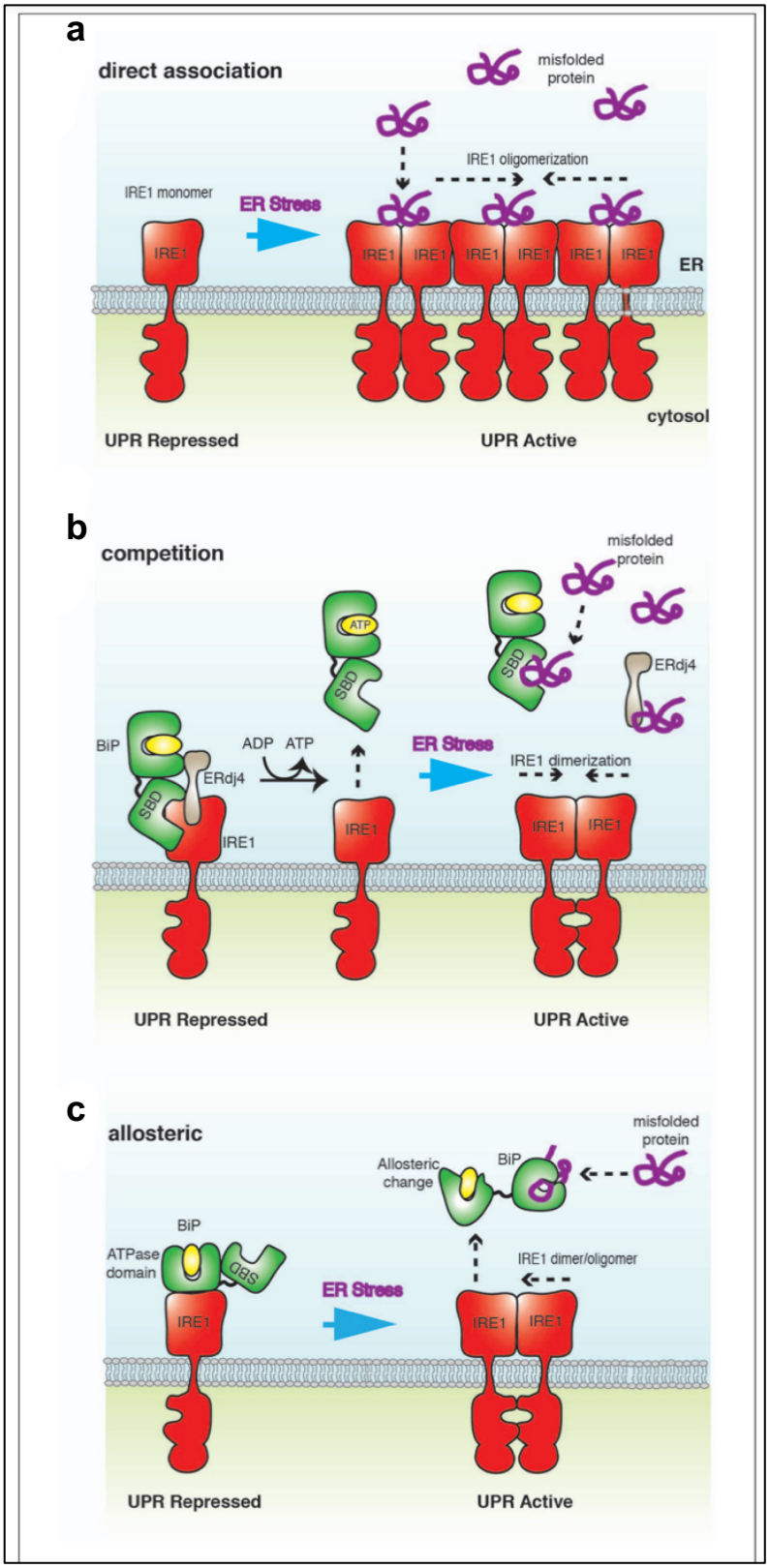


Figure 39: ER stress-sensing mechanisms. Overview of the (**a**) direct association, (**b**) BiP competition and (**c**) BiP allosteric models. SBD: substrate binding domain. From[518].

1.3.7.4 Lipid-dependent regulation of the UPR^{ER}

As described above, the three UPR^{ER} branches can be activated by various cellular stresses, such as glucose deprivation and disruption of calcium homeostasis, resulting in ER stress through the accumulation of misfolded proteins[504] (*i.e.,* proteotoxic stress).

Besides proteotoxic stress, firm links between lipids and UPR^{ER} signaling were provided by several studies in yeast. First, the UPR^{ER} transducer IRE1 and its downstream transcription factor HAC1/IRE2 (the yeast ortholog of the metazoan XBP1) were required for growth in medium deprived of inositol[527]. Second, UPR^{ER} signaling in yeast was strongly activated by deletion of genes regulating lipid metabolism[528]. Additionally, perturbation of cellular lipid composition also activates the UPR^{ER} in mammalian cells. Indeed, enhanced UPR^{ER} signaling has been observed in cholesterol-loaded macrophages[529], in pancreatic beta cells exposed to saturated fatty acids[530] after perturbation of sphingolipid metabolism[531] and in the liver of mice fed with a high-fat diet[532].

Lipid composition can affect protein folding in the ER and indirectly activate UPR^{ER} transducers. Changes in lipid composition can perturb ER calcium homeostasis, which inhibits the function of calcium-dependent enzymes and chaperones. For instance, ER stress signaling in the liver of obese mice[532] and in cholesterol-loaded macrophages[533] correlated with an inhibition of the SERCA pump and activation of the UPR^{ER}. However, several studies have showed that lipid changes may also affect UPR^{ER} signaling independently of their effect on protein folding in the ER. In yeast, depletion of inositol, the phospholipid building block, activated the UPR^{ER} without causing luminal unfolded protein stress[534]. Additionally, in *C. elegans*, deletion of the subunit of the transcriptional regulator complex Mediator, mdt-15, led to an increase in membrane lipid saturation and the activation IRE1 and PERK, without formation of protein aggregates in the ER. This may suggest that IRE1 and PERK activation is mediated *via* a component that is independent of unfolded protein stress[535]. Direct evidence of the direct activation of UPR^{ER} by lipid changes, independently of their effects on unfolded proteins in the lumen of the ER, was provided by the observation that IRE1 and PERK lacking their luminal sensing domains were activated in yeast deprived of inositol[536], or mammalian cells exposed to saturated fatty acid[537].

In recent years, it has become clear that atypical lipid compositions of the ER membrane, collectively referred to as lipid bilayer stress (LBS), can potently and directly activate the UPR^{ER}, independently of unfolded proteins, by inserting their transmembrane domain in the ER membrane. Indeed, UPR^{ER} sensors can be activated in response to ER membrane perturbations caused by changes in phospholipids (PL)[538] composition, in cholesterol[539], sterol[540] and inositol[536] levels and by changes in lipid accumulation[541] and saturation[537] (**Figure 40**). LBS activation of the UPR^{ER} is evolutionary conserved and has been described in worms, yeast, and mammals[542].

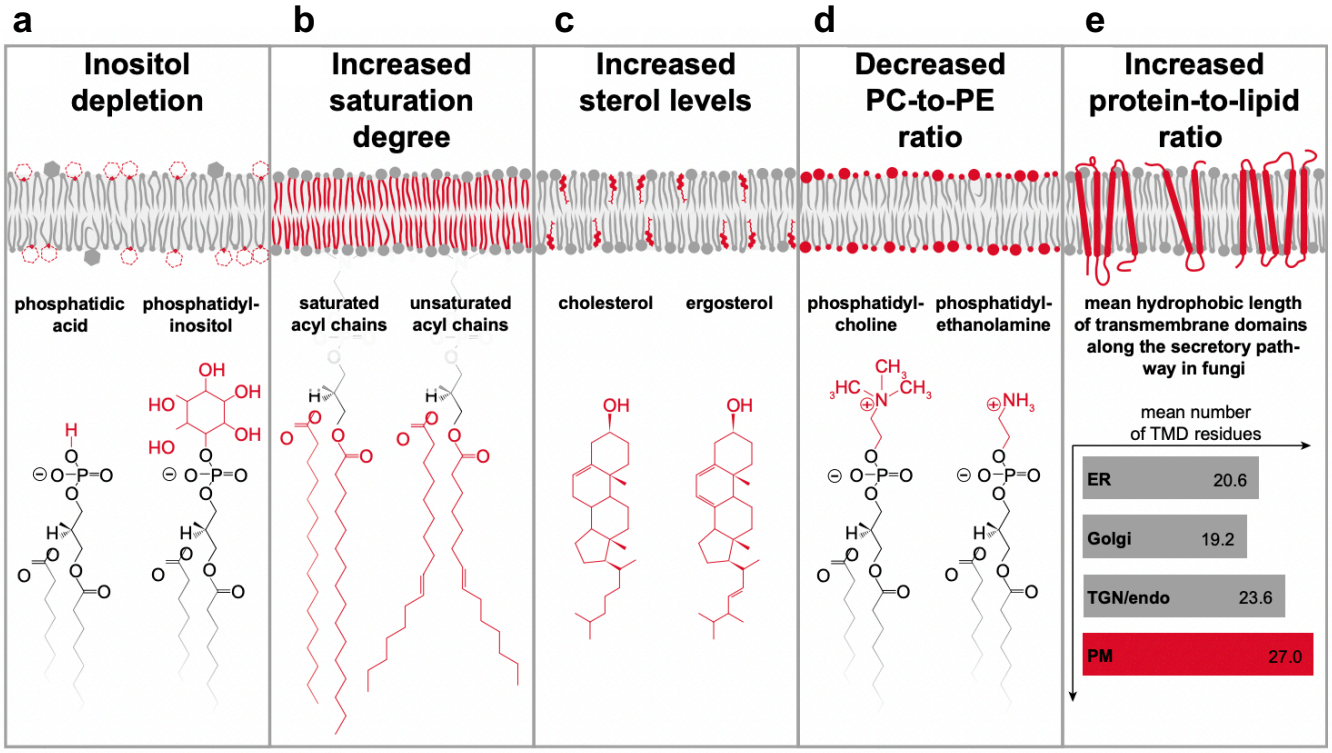


Figure 40: Illustration of conditions causing lipid bilayer stress (LBS) and unfolded protein response (UPR^{ER}) activation. Inositol depletion reducing phosphatidylinositol (PI) lipids level (**a**), increased level of saturated fatty acyl chains (**b**), increased sterol levels (**c**), decreased PC-to-PE ratio (**d**), and increased protein to lipid ratio (**e**) tend to activate the UPR^{ER}. From[542].

1.3.7.5 Sensing lipid bilayer stress

Lipid bilayer stress (LBS) induced by elevated levels of saturated fatty acids within the biological membrane can be sensed by IRE1α and PERK, lacking their luminal stress-sensing domain[537]. In this study, they suggested that IRE1α and PERK can sense LBS *via* their transmembrane domains. They also proposed that fatty acid saturation causes a decrease in membrane fluidity, which favors IRE1α and PERK oligomerization (**Figure 41a**). Additionally, the conserved amphipathic helix within the luminal domain of yeast IRE1, together with its transmembrane domain were proposed to drive local membrane compression and acyl chain disordering, contributing to IRE1 oligomerization[543]. Taken together, these findings suggest that both secondary structures of IRE1 and PERK at the membrane, and the biophysical properties of the surrounding membrane can drive the LBS sensing mechanism.

ATF6 was also shown to respond not only to protetoxic stressors, but also to sphingolipids: dihydrosphingosine (DHS) and dihydroceramide (DHC), while being non-responsive to ceramides[544]. Importantly, a conserved VXXFIXXNY sequence motif within ATF6 transmembrane domain was shown to be important for the ability of ATF6 to respond to DHS or DHC. However, this binding motif was not found in IRE1 or PERK (**Figure 41b**), suggesting that

IRE1 and PERK are tasked with sensing more general membrane characteristics, whereas ATF6 is activated by specific sphingolipid inducers.

Finally, other ER-resident proteins can sense specific lipid species to maintain lipid homeostasis. For instance, the membrane-bound transcription factor sterol regulatory element binding element (SREBP) can regulate cholesterol synthesis at the ER[545,546].
Together, these lipid sensors maintain normal ER lipid homeostasis independently of the UPR^{ER}.

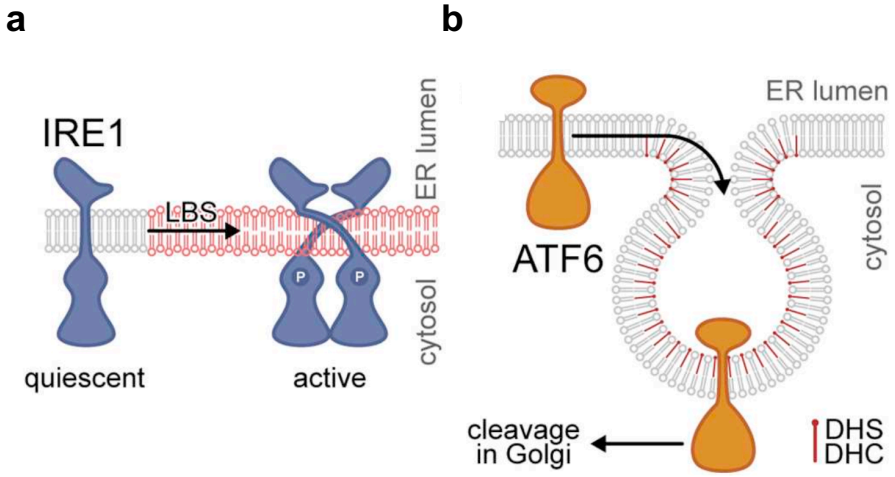


Figure 41: Proposed models for sensing lipid ER stress. (**a**) Lipid bilayer stress (LBS) favors the bending of IRE1 transmembrane domain, which promotes IRE1 activation. (**b**) ATF6 can be activated by the increase in dihydrosphingosine (DHS) or dihydroceramide (DHC). This favors membrane curvature and ATF6 packing within COPII-mediated vesicle, which results in the release of cleaved cytosolic ATF6f. Adapted from[523].

1.3.7.6 The "UPRosome": multiple regulatory checkpoints

Evidence is accumulating for possible mechanisms that control the temporal behavior of UPR^{ER} signaling and determine cell fate in switching from pro-survival responses to death programs under ER stress. Initially, UPR^{ER} has been viewed as a direct transduction of ER stress levels. However, emerging findings indicate that the three UPR^{ER} branches are regulated through post-translational modifications and association of positive and negative regulators that specifically affect their activation (**Figure 42**). Many laboratories have identified binding modulators of UPR^{ER} sensors. Most of these studies described binding partners with IRE1α leading to the definition of a complex IRE1α signaling platform that has been referred to as "UPRosome"[547]. This platform initiates multiple highly regulated signaling responses to regulate the activation status of IRE1α in terms of signaling intensity and kinetics of activation/inactivation.

Several proteins associate physically to IRE1α to regulate the amplitude of its signaling, such as the pro-apoptotic proteins BAX and BAK[548], the cytosolic chaperone heat shock protein 72

(HSP72)[549], Protein Tyr Phosphatase 1B (PTP1B), the MAPK-related proteins ASK1-interacting protein 1 (AIP1), JNK-inhibitory kinase (JIK), and JUN activation domain-binding protein 1 (JAB1)[516]. Most of these regulators enhance IRE1α signaling, whereas other proteins, such as Fortilin inhibits IRE1α signaling, increasing cell death resistant under ER stress[550] (**Figure 42a**). Many of IRE1α-binding partners have been described. available data about these interactors are summarized in[551].

The UPRosome may also serve as a platform to allow the crosstalk between the UPR^{ER} and other signaling pathways. For instance, IRE1 binding to TRAF2 activates the JNK and autophagy pathways. However, its binding with the adaptor proteins Nck engages nuclear factor kB (NF-kB)[552].

Furthermore, several post-translational modifications can affect the signaling of IRE1 and PERK. Protein kinase A (PKA) phosphorylates IRE1α and engages its RNase domain in the absence of ER stress, whereas PP2A and the ER phosphatase PP2Ce lead to its dephosphorylation[552]. PARP16 can ADP-rybosylates IRE1 and PERK, thus enhancing their activities. S-nitrosylation of a specific cysteine of IRE1 inhibits its RNase activity, while it enhances PERK signaling[552]. Controlling IRE1α stability is also important to adjust the UPR^{ER}. Indeed, IRE1α levels have been shown to be regulated by 1) DDRGK1, a critical component of the ubiquitin-fold modifier 1 (Ufm1) system, 2) an interaction with the selective autophagy receptor optineurin and 3) the ERAD pathway[552].

Although less explored, PERK (**Figure 42b**) and ATF6 can be individually modulated through the binding of specific factors. PERK signaling can be enhanced by binding to N-myc downstream-regulated gene-2 (NDRG2), canopy homolog 2 (CNPY2), and the small GTPase Rheb[552]. However, the cold-inducible RNA-binding protein 3 (RBM3) inhibits PERK signaling through a binding with Nuclear Factor 90 (NF90)[553]. A few direct regulators of ATF6α have been reported, such as Wolfram syndrome 1 (WFS1) that mediates the degradation of ATF6 by the proteasome[554], NF-Y (also known as CBF), YY1, TATA-binding protein (TBP) and XBP1s[516].

Overall, these examples reflect the highly dynamic and complex nature of UPR^{ER} signaling, which may explain the diversity of the cellular responses controlled by this signaling pathway. However, it is still very important to perform systematic interactome screens to define the composition of the UPRosome complexes at the level of the three major UPR^{ER} stress sensors and their assembling dynamic under ER stress conditions and in different cell types.

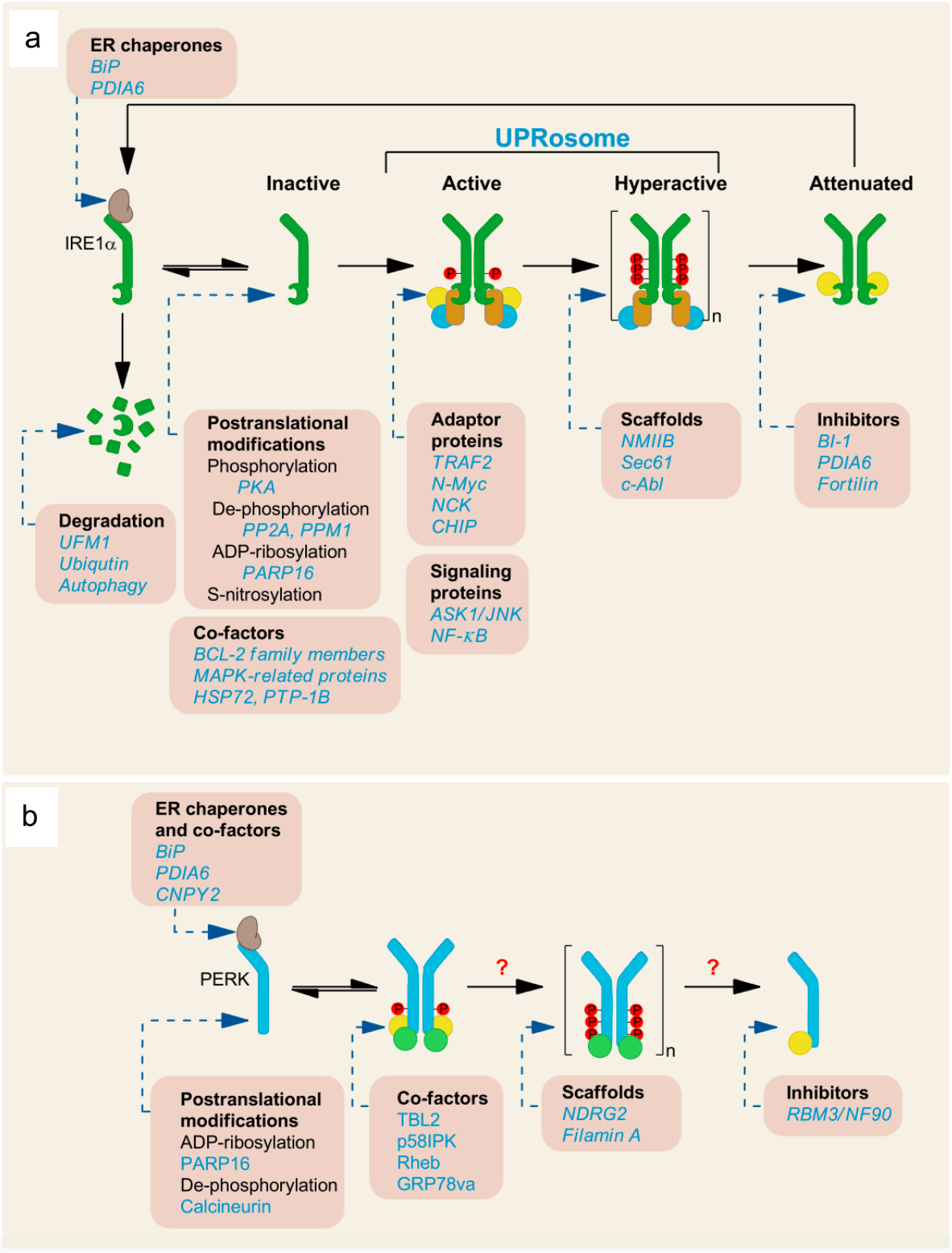


Figure 42: Multiple regulatory checkpoints of the UPR^{ER}. Regulation of IRE1 (**a**) and PERK (**b**) signaling. UPR^{ER} stress sensors activity can be regulated by the binding of a diversity of co-factors and post-translational modifications. This binding affects their downstream signaling, protein stability, kinetics of activation and attenuation and the crosstalk of the UPR^{ER} with other signaling pathways. From[713].

1.3.7.7 Roles of UPR^{ER} in physiological processes

As we move forward in our understanding of the UPR^{ER}, we are recognizing that UPR^{ER} signaling has fundamental roles in multiple physiological processes, beyond its role in protecting ER physiology[555]. It is now becoming clear that the UPR^{ER} has been drafted into the role of a regulator of basal homeostasis, through its activation by different physiological stimuli, far outside the realm of protein misfolding[502]. Indeed, it is increasingly obvious that what began as an organelle-specific stress response has expanded into a complex signaling network playing a central homeostatic role in normal vertebrate physiology. In this view, the inputs are physiological stimuli that activate part or all of the UPR^{ER}, leading to physiological outputs from innate immunity, energy, and lipid metabolism to cell differentiation, mediated by both transcriptional and non-transcriptional mechanisms (**Figure 43**).

- **UPR^{ER} in immunity and inflammation:**

Several reports showed the crucial functions of UPR^{ER} in immunity and inflammation[556]. XBP1 was found to be upregulated after the exposure of cells to the pro-inflammatory cytokine interleukin-6 (IL-6)[557]. Moreover, XBP1 deficiency in mice and *Caenorhabditis elegans* ablates their ability to eliminate bacterial pathogens[557]. Toll-like receptors (TLRs) stimulation specifically triggers XBP1 mRNA splicing but represses ATF6 and PERK signaling to enhance the transcription of pro-inflammatory cytokines. In fact, some evidence suggested that XBP1 mRNA splicing by TLRs is independent of protein misfolding, IRE1α-dependent and is controlled through a specific signaling branch involving the adaptor proteins myeloid differentiation primary response 88 (MYD88), TIR domain-containing adaptor protein (TIRAP), TRAF6 and NADPH oxidase 2 (NOX2)[558,559] (**Figure 43**).

- **UPR^{ER} in glucose and lipid metabolisms:**

The ER is now viewed as an essential apparatus in the coordination of metabolic programs linked to it, by its ability to modulate synthetic and catabolic pathways of various nutrients. Strikingly, the three UPR^{ER} branches intersect with a variety of inflammatory and stress signaling systems including the NF-κB-IκB kinase and JNK-AP1 pathways, and with networks activated by oxidative stress, all of which can influence metabolism[560].

UPR^{ER} branches are implicated in glucose metabolism. For instance, PERK-deficient and eIF2α mutant mice showed defective gluconeogenesis in the livers. This effect has been linked with islet cell dysfunction or death triggered by ER failure[561]. Further studies demonstrated the role of IRE1α activation in the regulation of insulin biosynthesis, in secretory pancreatic β-cells[562]. Particularly, RACK1 mediates the assembly of the complex containing IRE1α, RACK1, and protein phosphatase 2A (PP2A), leading to dephosphorylation of IRE1α by PP2A and the inhibition of glucose stimulated IRE1α activation[563]. UPR^{ER} is also implicated in the modulation

of lipid metabolism[507]. For instance, in liver cells, XBP1s regulates the transcription of many genes implicated in fatty acid synthesis, such as Scd-1 (stearoyl-CoA desaturase-1), Acc2 (acetyl-CoA carboxylases 2), and Dgat2 (diacyl glycerol acyl transferase 2)[564].

It is becoming clear that intersection of the UPR^{ER} with different inflammatory and stress signaling pathways (lipid metabolism and energy control pathways) is critical in chronic metabolic diseases such as obesity, insulin resistance, and type 2 diabetes[565]. For instance, IRE1α activation is involved in insulin resistance in the liver due to signaling crosstalk between the IRE1α–JNK pathway and the subsequent phosphorylation of insulin receptor substrate 1 (IRS1), impairing insulin action[566].

Taken together, these findings strongly suggest that UPR^{ER} branches can adjust energy metabolism by monitoring the metabolic state of the cell (**Figure 43**).

- **<u>UPR^{ER} in cell differentiation programs:</u>**

Beside its physiological roles in highly secretory cells, UPR^{ER} also functions in cell differentiation. The roles of UPR^{ER} in stem cells, such as embryonic stem cells (ESCs), tissue stem cells, cancer stem cells and induced pluripotent cells, have been widely reported. These functions are dependent on the stem cell type. Indeed, the UPR^{ER} was implicated in the embryonic development, differentiation, and pluripotency of ESCs. In addition, it has been involved in the regulation of the self-renewal capacities of hematopoietic stem cells (HSCs), and recent studies are showing UPR^{ER} vital roles in neural differentiation during brain development[567].

Some examples of the implication of UPR^{ER} in cell differentiation programs include XBP1 and IRE1α. For instance, XBP1 is activated in B cells when they differentiate to plasma cells, possibly by inhibiting transcriptional repressors of plasma cell differentiation, like interferon regulatory factor 4 (IRF4) and B lymphocyte-induced maturation protein 1 (BLIMP1)[568]. IRE1α was also required in the early and late stages of B lymphopoiesis[569]. Furthermore, XBP1 is implicated in the transcriptional regulation of key differentiation genes, such as muscle, intestine and stomach expression 1 (MIST1). XBP1-MIST1 axis was shown to be required for the maturation of gastric zymogenic cells[570]. Finally, the brain-derived neurotrophic factor (BDNF) activates *XBP1* splicing in the neurites, which contributes to neurite outgrowth. BDNF can also regulate IRE1α as it signals through TRKB or p75[571] (**Figure 43**).

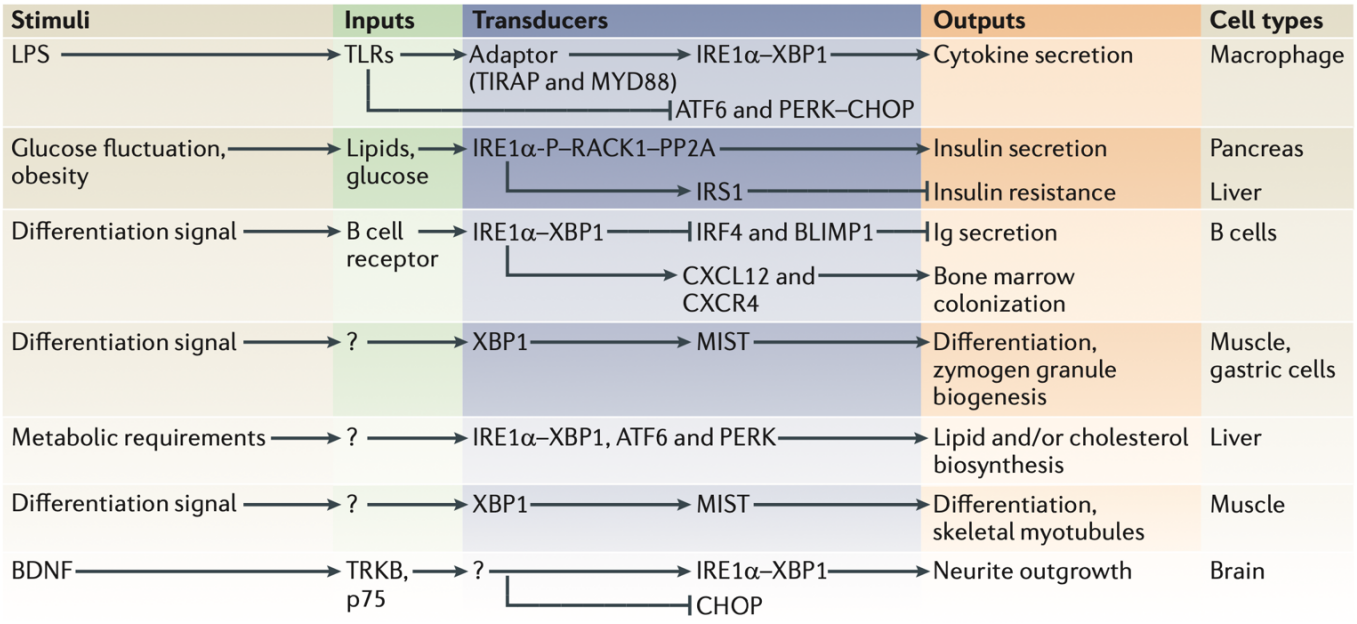


Figure 43: Novel physiological roles of the UPR^{ER}. From[516]

1.3.8 ERAD

Although correctly folded proteins are transported to their final destination, misfolded or unassembled proteins are retained in the ER and subsequently degraded by the "ER-associated degradation" (ERAD), a conserved pathway from yeast to mammals. ERAD is essential for maintenance of ER homeostasis by degrading misfolded or aberrant proteins but can also influence the organization and functions of the ER, and its interactions with adjacent organelles[572]. Briefly, the ERAD process begins by selecting the substrate protein, based on either glycosylation tags, mannose trimming status and/or conformational change. This can be aided by chaperones, such as 78-kDa glucose-regulated protein (Grp78; known as HSPA5), ER degradation-enhancing α-mannosidase-like protein (Edem) family proteins and osteosarcoma amplified 9 (Os9). Then, the substrate is retro-translocated into the cytosol through the dislocon proteins, such as the polytopic dislocon Hrd1 and degradation in endoplasmic reticulum protein (Derlin) family members (Derlin-1, Derlin-2 or Derlin-3). After retro-translocation, Hrd1 ubiquitilates the substrate, which will be targeted for proteasomal degradation by the ATPase valosin-containing protein Vcp (also known as p97) and other ubiquitin-modifying enzymes[573].

1.3.9 ER-Phagy

When misfolded ER proteins are not degraded by ERAD, due to the limitation of the size of the retro-translocation pore, ER-phagy is applied as a pathway for ERAD-resistant misfolded proteins or aggregates. ER-phagy is defined as a group of fundamental biological processes,

leading to lysosomal turnover of selected subdomains of the endoplasmic reticulum (ER). It can be classified into three types, termed 1) macro-ER-phagy (in which ER fragments are sequestered by double-membrane autophagosomes that eventually fuse with lysosomes/vacuoles), 2) micro-ER-phagy (in which ER fragments are directly engulfed by endosomes/lysosomes/vacuoles) and 3) LC3-dependent vesicular delivery mediated by the fusion of ER vesicles into lysosomes[574]. ER-phagy is mediated by ER-phagy receptors that may be present on distinct ER subdomains. So far, six ER membrane-integrated (RTN3, FAM134B, CCPG1, SEC62, TEX264, and ATL3) and three soluble ER-phagy receptors (C53, CALCOCO1, and p62) have been identified in mammals[495].

1.3.10 ER stress and diseases

Under normal physiological conditions, the ER protein quality control systems are able to maintain and sustain ER proteostasis by promoting protein folding, eliminating misfolded proteins or aggregates, and expanding the ER capacity when needed. However, these mechanisms are perturbated under pathological conditions.

- **Neurodegenerative diseases:**

The UPR^{ER} was observed to be activated in several neurodegenerative diseases such as: amyotrophic lateral sclerosis, Parkinson's disease, Huntington's disease, prion-related disorders, and Alzheimer's disease, and demyelinating neurodegenerative autoimmune diseases such as multiple sclerosis, Pelizaeus-Merzbacher disease, and transverse myelitis[575].

- **Inflammatory diseases:**

ER stress is an important pathway implicated in the development of intestinal inflammation associated with Inflammatory Bowel Disease (IBD) and likely other intestinal inflammatory disorders[576]. For example, IRE1β, XBP1 or PERK were activated in Crohn's disease and ulcerative colitis, and IRE1α and the subsequent splicing of *XBP1* mRNA were activated by Toll-Like Receptor 2 (TLR2) and TLR4[510].

- **Metabolic disorders:**

ER stress has been implicated in a variety of metabolic disorders, including obesity, type 2 diabetes, atherosclerosis, heart and liver diseases[577]. Several studies showed that IRE1 and XBP1 are implicated in metabolic regulation of lipid and glucose homeostasis. Indeed, RNAi-mediated silencing of IRE1α activity reduced hepatic gluconeogenesis[578], and XBP1 deletion in the liver of adult mice at early high fat diet feeding leads to glucose intolerance and insulin resistance[579]. Moreover, deletion of IRE1α in liver prompted hepatosteatosis, possibly due to decreased TG secretion[580].

- **Cancer:**

Emerging evidence are supporting the role of ER stress and UPR^{ER} in the establishment and progression of several cancers, namely breast cancer, prostate cancer and glioblastoma multiforme. The UPR^{ER} was reported to be implicated in each stage of tumor progression, including transformation, unrestricted cell division, angiogenesis, invasion, and metastatic spread. Moreover, the effectiveness of chemotherapy can be limited by UPR^{ER} activation, as this pathway can contribute to the development of chemoresistance. Although sustained UPR^{ER} activation leads to apoptotic signaling, cancer cells bypass this apoptotic switch to promote proliferation and metastasis[581].

1.3.11 UPR^{ER} modulating compounds

The UPR^{ER} has been implicated in the pathogenesis of several human diseases, which has led to significant interest in establishing compounds that modulate (activate or inhibit) one or more arms of the UPR^{ER}. This will provide new opportunities to define the therapeutic potential for targeting the UPR^{ER} in human diseases. Several reports described the discovery of these compounds, their mechanism of action, and their applicability for studying the importance of UPR^{ER} pathway in cellular and *in vivo* models[582]. **Figure 44** shows some small molecules (inhibitors and activators) targeting specific UPR^{ER} signaling components.

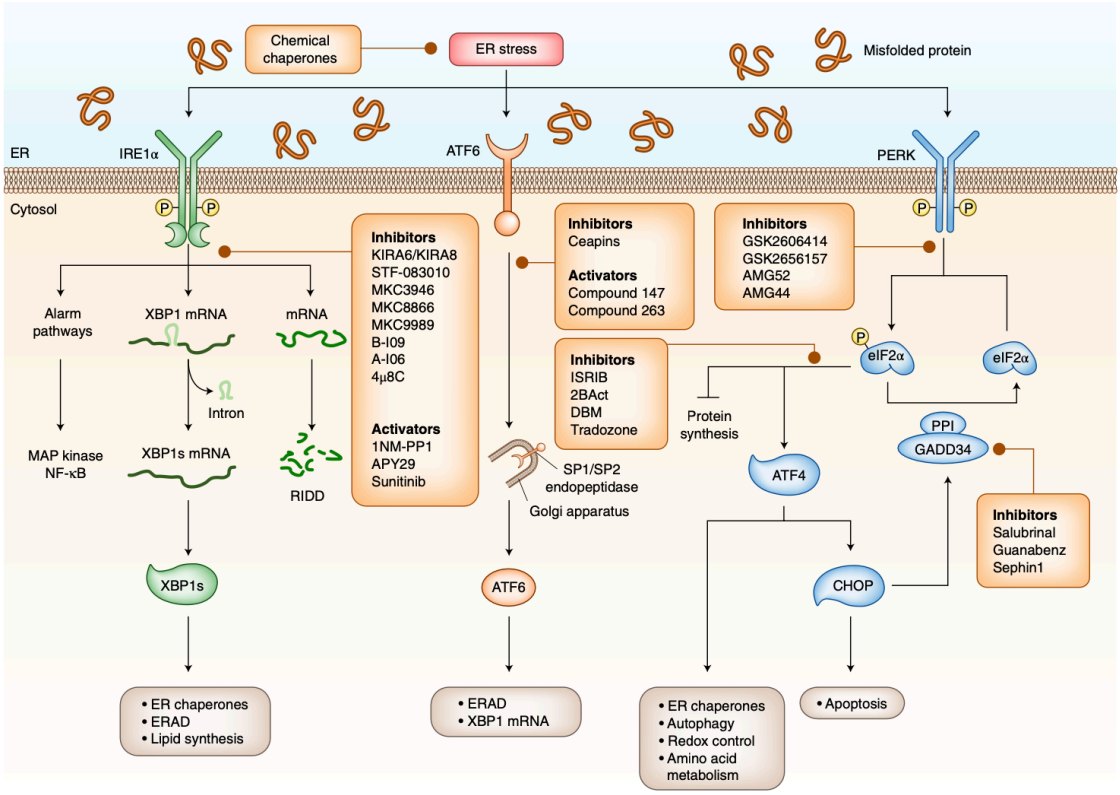


Figure 44: Small molecules targeting specific UPR^{ER} signaling components. From[583].

2 AIM

mRNA metabolism is tightly orchestrated by highly-regulated RNA binding proteins (RBPs), which allows cells to rapidly adapt to changing environmental conditions. The large repertoire of RBPs, with diverse affinity and specificity towards target RNAs, is likely to explicit the tremendous complexity of post-transcriptional regulation. Therefore, many studies were conducted to investigate the binding properties, RNA targets, and functional roles of these proteins.

The FET family is an interesting family of DNA and RNA binding proteins, which consists of FUS, EWSR1 and TAF15. These proteins are structurally similar and contain a number of evolutionary conserved motifs, such as the SYGQ-rich motif, RRM, ZnF, and RGG domains. In addition, FET proteins have both unique and overlapping functions. For instance, they have been associated with transcription, RNA splicing, miRNA processing, RNA transport and translation, signaling and maintenance of genomic integrity. Particularly, EWSR1 have attracted broad attention because of its involvement in tumor-related chromosomal translocations that associate the NTD of EWSR1 with various genes encoding transcription factors. Intriguingly, a growing body of evidence has described EWSR1 as a multifunctional protein, implicated in transcriptional and post-transcriptional regulation. Therefore, EWSR1 concerts a multitude of cellular pathways by itself or *via* multiple interactions with other molecules in a gene-context dependent manner. Due to its implication in post-transcriptional regulation, we hypothesized that EWSR1 protein might be particularly involved in the control of mRNA translation. Accordingly, in this project, we viewed EWSR1 through the lens of translation, which would thereby add another layer of EWSR1-mediated post-transcriptional regulation and represent an important finding for two main reasons. First, it will provide a better understanding of the molecular function of EWSR1, which can lead to novel therapeutic opportunities for EWSR1-related diseases. Second, it might underscore the importance of translational dysregulation in the emergence of several disorders.

On this basis, the first aim of this work was to examine whether EWSR1 was able to control mRNA translation at the genome-wide level. To address this, we performed ribosome profiling (RIBO-seq) experiment in HeLa cell line in the presence/absence of EWSR1. This experiment strengthened our assumption that EWSR1 might present a new role in the repression of mRNA translation. Next, we aimed at characterizing the molecular determinants of this new function. Therefore, we conducted MS2-tethering assays, mutants design and binding experiments (*e.g.,* co-IPs, GST-pulldown) to determine the region of EWSR1 responsible for the translational activity, but also to assess EWSR1 interaction with the translational machinery. Finally, we pursued to explore the biological relevance of our findings. Therefore, we evaluated the effect of the translational regulation by EWSR1, on the regulation of lipid homeostasis, ER shape and function and UPR^{ER} activation, using structural and metabolomics analysis (*e.g.*, transmission electron microscopy (TEM), lipidomic analysis).

3 RESULTS

3.1 EWSR1 is a repressor of mRNA translation

To decipher a potential role of EWSR1 in the regulation of mRNA translation, we investigated genome-wide translational changes in response to EWSR1 depletion using RIBO-seq (**Figure 45**).

To this aim, we prepared two independent series of RNA samples from *siCTL* or *siEWSR1#1* transfected HeLa cells, pre-treated with 100 mg/ml cycloheximide.

We first conducted a series of quality control analyses of our RIBO-Seq experiments. EWSR1 knockdown (KD) was confirmed by western blotting at the protein level (~ 85%) for the two replicates (**Figures 46a,b**)**.** Principal Component Analysis (PCA) revealed distinct clustering between the *siEWSR1#1* and *siCTL* conditions, both at the RIBO-seq and RNA-seq levels (**Figures 47a,b**). Translating ribosomes create a 3-nucleotide (nt) periodicity, known as "phasing". This triplet periodicity becomes visible when the read alignments are mapped to their P-site offsets[309]. Our data showed that ribosomal phasing against the ribosome protected fragment (RPF) length presents a predominant phase 0 enrichment, which is expected in high quality ribosome profiling experiments. In addition, distribution of RPF length in our RIBO-seq data set is centered around 29-30 nts, which corresponds to the expected fragment size protected by 80S ribosomes[309] (**Figure 47c**).

The average percentage content by gene biotype for both RIBO-seq and RNA-seq reads is shown in **Figure 47d.** The largest number of reads mapped to protein-coding mRNA (51.23%), while other alignments mapped to other biotypes including lncRNA (24.37%) and sncRNA, such as miRNA (0.88%), snRNA (0.65%), snoRNA (0.5%). Although the high proportion of reads mapping to lncRNA in our dataset was surprising, it has already been reported that lncRNA can be bound by ribosomes even though they do not code for proteins[584]. Interestingly, the mRNA reads were almost exclusively mapped to the CDS region (97.11%), while the reads mapping to untranslated regions or introns, each represented less than 1% of RPFs (0.31% to 3'UTRs, 0.58% to 5'UTRs and 0.95% to introns) (**Figure 47e**).

These analyses demonstrated that we generated a high-quality RIBO-seq library suitable for subsequent analysis.

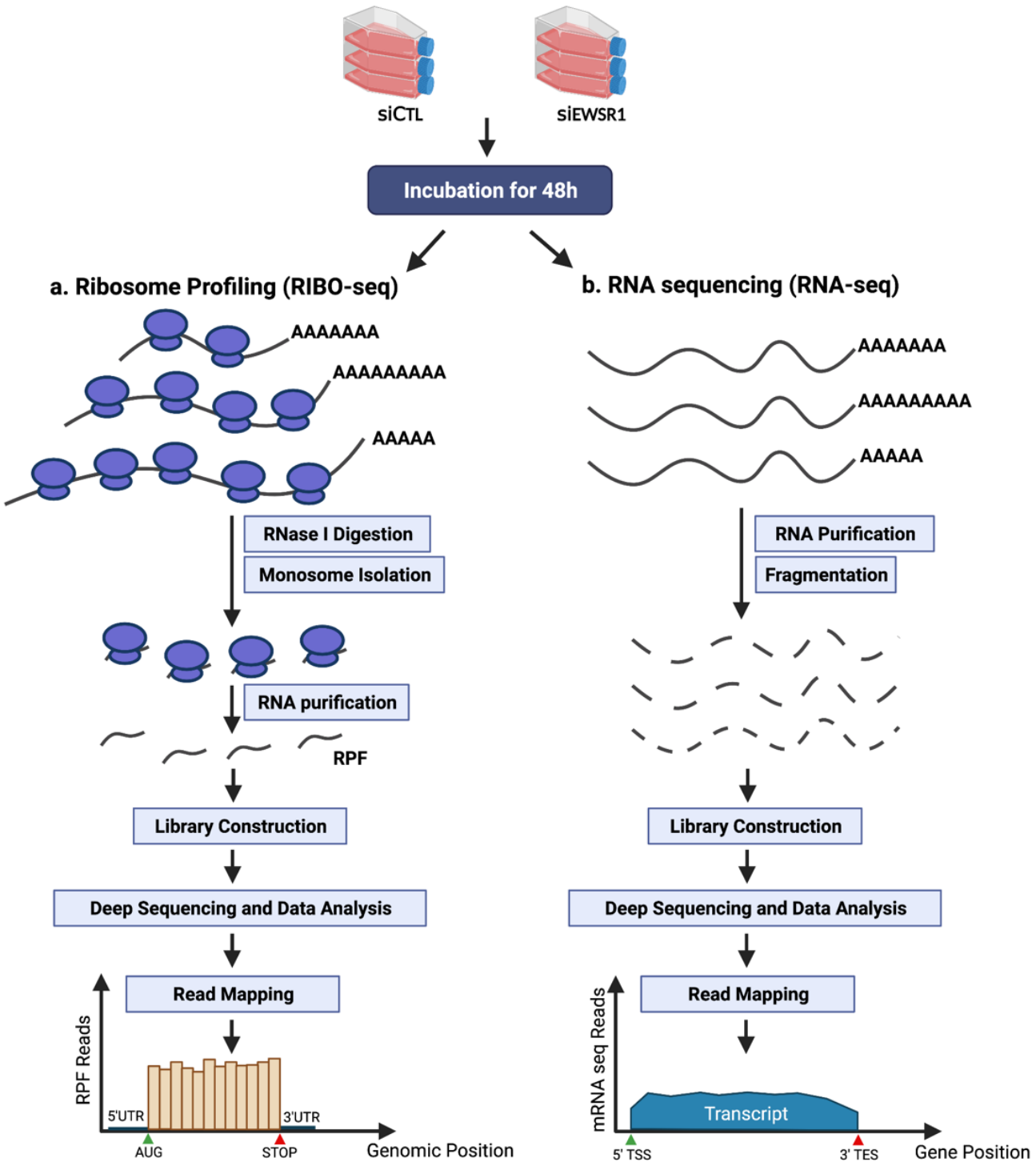


Figure 45: Overview of the RIBO-seq analysis workflow. Created with BioRender.com. RPF: ribosome protected fragments, TSS: transcription start site, TES: transcription end site.

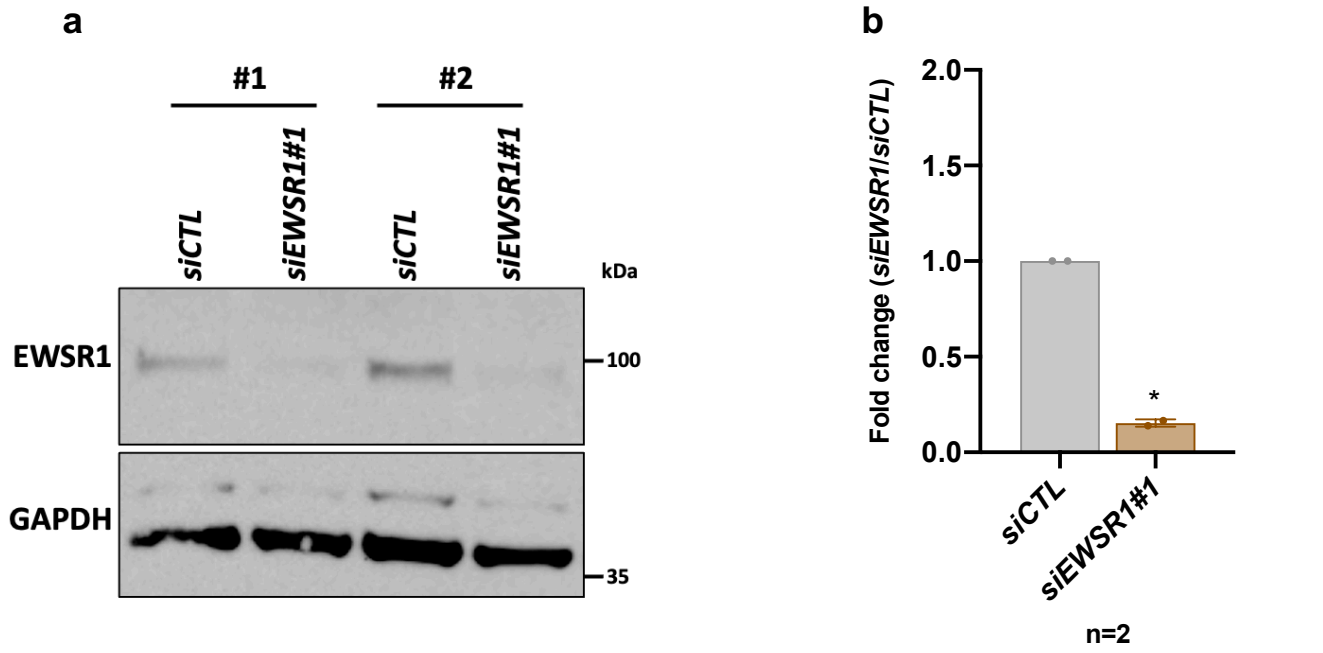
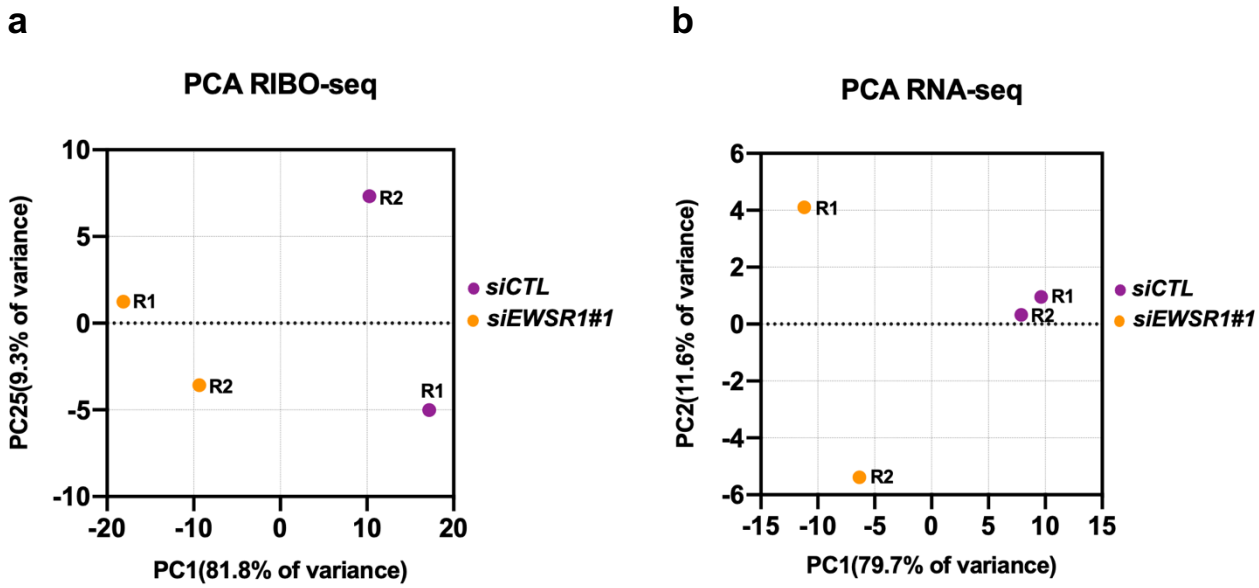


Figure 46: Verification of EWSR1 KD of the ribosome profiling experiment. (**a**) Western blotting of EWSR1 and GAPDH (loading control). Samples are total cell lysates from HeLa cells transfected with *siEWSR1#1* or *siCTL*. (**b**) The level of EWSR1 was quantified by band densitometry analysis using ImageJ. Results are shown as means ± sd (n = 2 independent experiments) after normalization to GAPDH. *P < 0.05 compared to the *siCTL* condition by one sample *t* test.



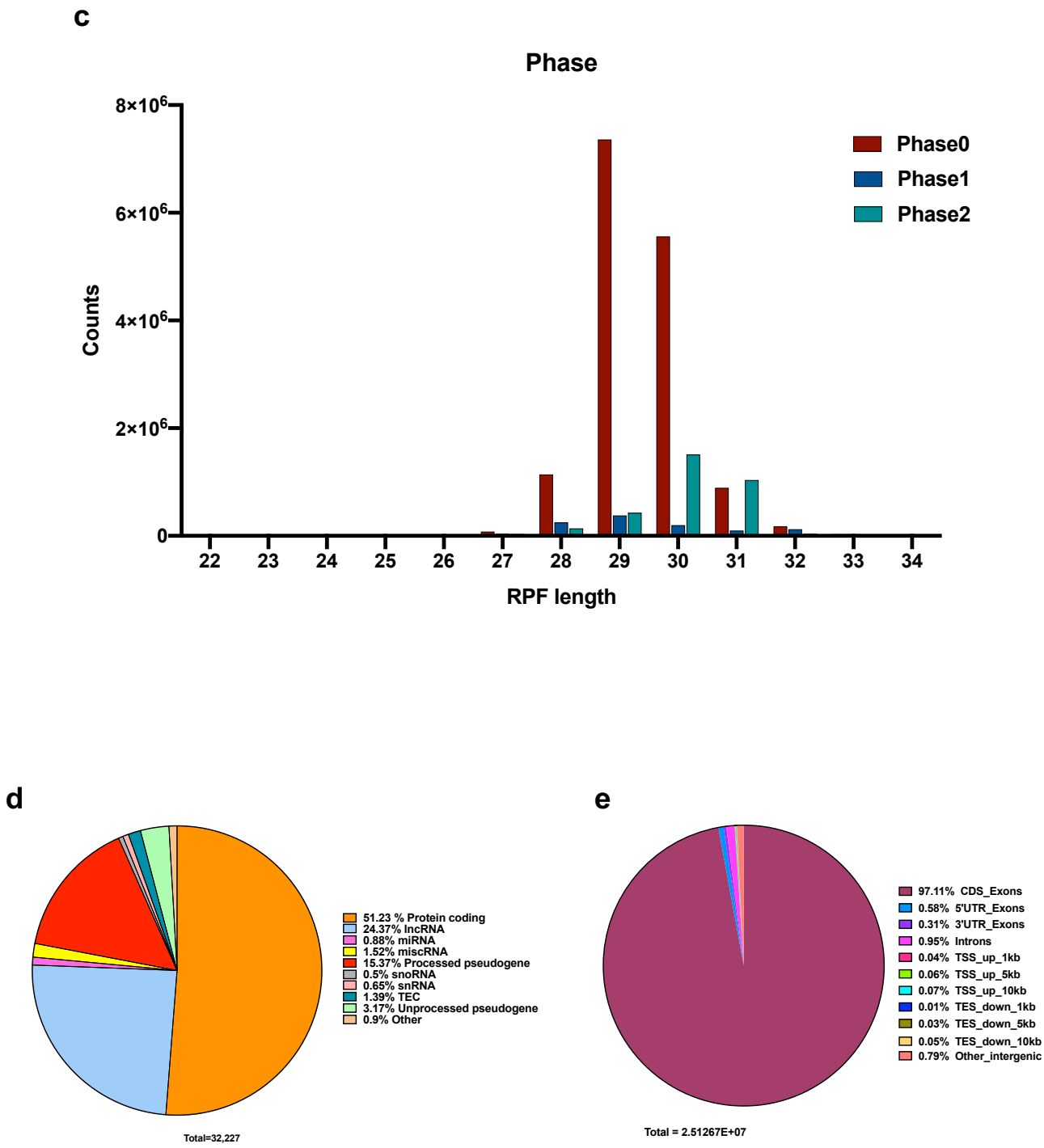


Figure 47: Quality controls of the ribosome profiling experiment. (**a, b**) Principal Components Analysis (PCA) of RIBO-seq (**a**) and RNA-seq (**b**) libraries from *siCTL* and *siEWSR1#1* HeLa cells (n = 2 independent experiments). (**c**) Plot of ribosomal phasing (phase 0, 1 and 2) against the ribosome protected fragment (RPF) length. (**d, e**) Pie charts describing read mapping distribution by gene biotype of both RNA-seq and RIBO-seq datasets (**d**) and to different genomic regions detected by RIBO-seq (**e**). Percentage of mapping for each gene biotype and each genomic region are indicated in the graphs.

Based on the RNA-seq analysis, we identified a total of 2,282 differentially expressed genes (DEGs), which were almost equally distributed between upregulated (n = 1,189) and downregulated (n = 1,093) upon EWSR1 knockdown (FDR < 0.05) (**Figure 48**).

We used Gene Set Enrichment Analysis (GSEA) (**Figure 49a**) to identify significantly deregulated pathways associated with EWSR1 knockdown, based on the KEGG pathway and Gene Ontology databases. This revealed that upregulated DEGs were associated with cancer-related pathways, including the TGF-β signaling pathway. Interestingly, the downregulated DEGs were associated with cell cycle, cell division and G2/M checkpoint, pointing to a pivotal role for EWSR1 in the progression through the cell cycle (**Figures 49b-d)**.

Based on these observations, we analyzed cell cycle progression in EWSR1 knocked down (EWSR1 KD) Hela cells by flow cytometry. We found that 56.93% and 79.23% of cells were in the G1 phase of the cell cycle in *siCTL* and *siEWSR1#*1 conditions, respectively. These observations suggested a blockage at the G1 phase when EWSR1 is knocked down, which was also highlighted by a significant reduction of the proportion of cells in the S (31.63% in *siCTL vs* 11.58% in *siEWSR1#1*) and G2 (13.1% in *siCTL vs* 7.35% in *siEWSR1#1*) phases (**Figure 50a**). Blockage of EWSR1 KD cells in G1 was also evidenced by upregulation of the cyclin-dependent kinase inhibitor $p21^{Cip1\ 585}$ (**Figure 50b**).These observations indicate that EWSR1 controls cell progression through G1 and correlate with a previous study showing that it activates genes involved in the G1/S transition in HEK293 cells[586].

Defects in cell cycle often lead to alterations in cell proliferation[587]. Indeed, cell proliferation analysis showed that EWSR1 knockdown using 2 different siRNAs (*siEWSR1#1* and *siEWSR1#2*) (**Figure 50c**) resulted in reduced proliferation of HeLa cells (**Figures 50d,e**). Altogether, our results are in agreement with the GSEA analyses and confirm a role for EWSR1 in the regulation of cell cycle progression.

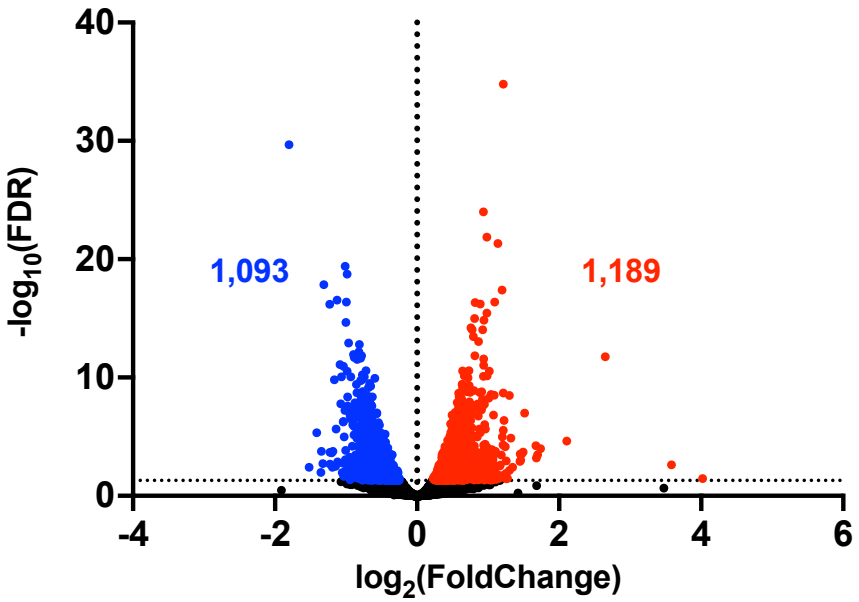


Figure 48: DEG upon EWSR1 KD. Volcano plot of significantly downregulated (blue) and upregulated (red) differential expressed genes (DEGs) upon EWSR1 knockdown in HeLa cells, as detected by RNA-seq analysis (n = 2 independent experiments, FDR = 0.05). Genes with unchanged expression are in black. Numbers refer to numbers of genes either up or downregulated.

a

Pathway	Gene ranks	NES	pval	padj
KEGG_CYTOKINE_CYTOKINE_RECEPTOR_INTERACTION		2.09	3.3e-03	1.4e-02
KEGG_N_GLYCAN_BIOSYNTHESIS		1.96	7.1e-03	2.6e-02
KEGG_TGF_BETA_SIGNALING_PATHWAY		1.70	2.0e-02	5.9e-02
KEGG_SPLICEOSOME		-1.72	1.2e-02	4.2e-02
KEGG_REGULATION_OF_ACTIN_CYTOSKELETON		-1.94	2.9e-03	1.3e-02
KEGG_PROGESTERONE_MEDIATED_OOCYTE_MATURATION		-2.04	1.8e-03	9.1e-03
KEGG_TIGHT_JUNCTION		-2.20	5.5e-04	3.2e-03
KEGG_CELL_CYCLE		-2.12	4.6e-04	3.1e-03
KEGG_PYRUVATE_METABOLISM		-2.36	1.5e-04	1.2e-03
KEGG_OOCYTE_MEIOSIS		-2.44	3.6e-05	3.7e-04
KEGG_GLYCOLYSIS_GLUCONEOGENESIS		-2.63	7.5e-06	1.0e-04
KEGG_PROTEASOME		-2.63	3.6e-06	7.3e-05
KEGG_VALINE_LEUCINE_AND_ISOLEUCINE_DEGRADATION		-2.70	6.6e-07	2.7e-05

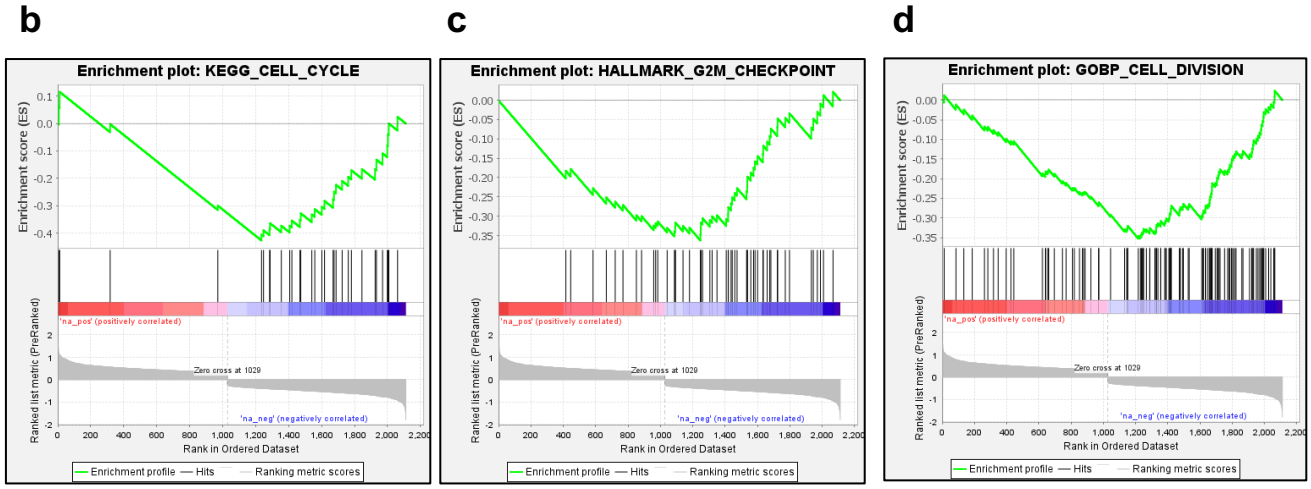
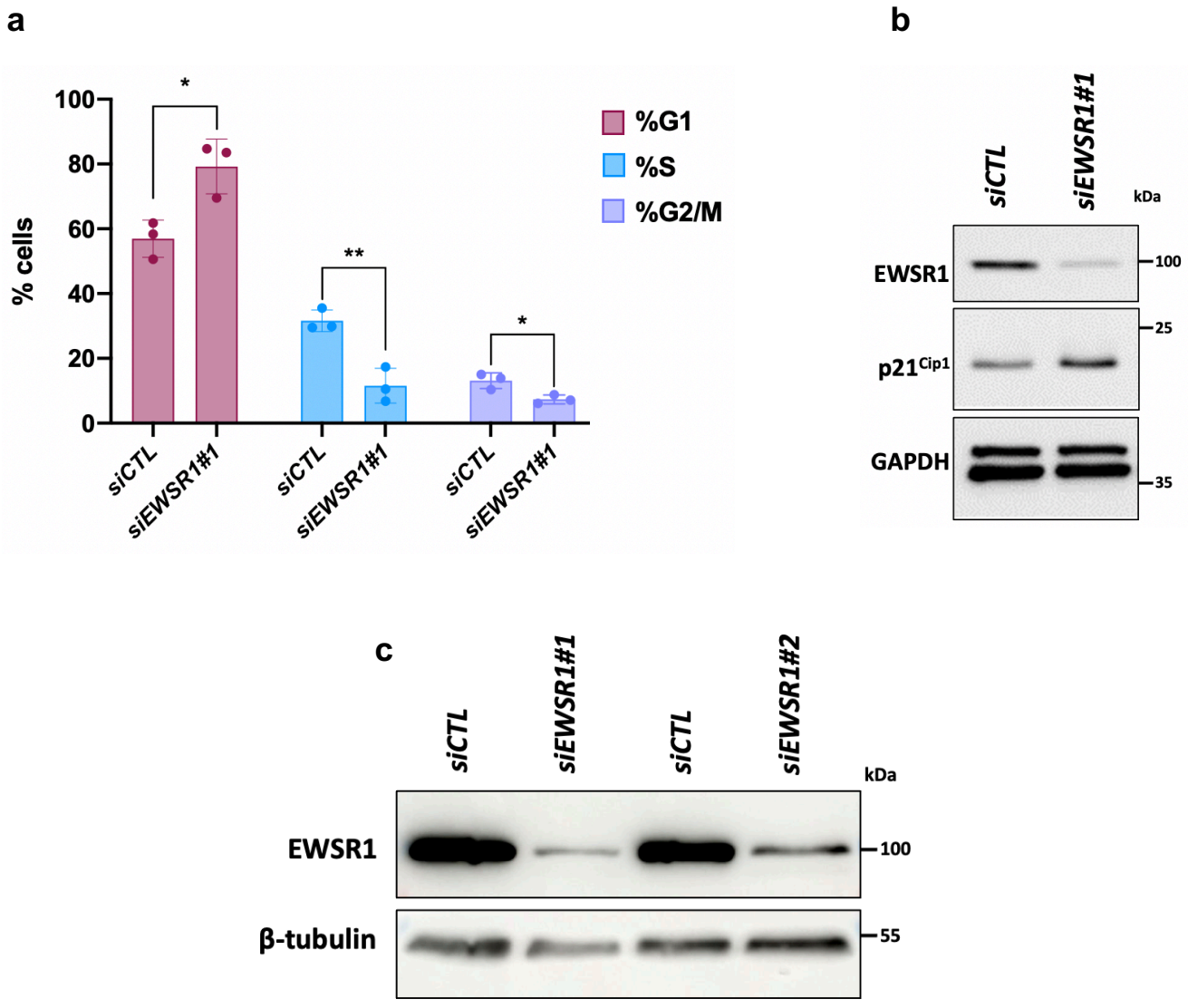


Figure 49: GSEA. (**a**) Gene Set Enrichment Analysis (GSEA) of DEG in EWSR1 KD cells. The third column represents genes ordered in a ranked list according to their differential expression between *siCTL* and *siEWSR1#1* conditions. Each vertical bar represents a gene in the ranked list that is included in the pathway ontology of each row. (**b, c, d**) Enrichment plots from GSEA: cell cycle (**b**), G2M checkpoint (**c**) and cell division (**d**).



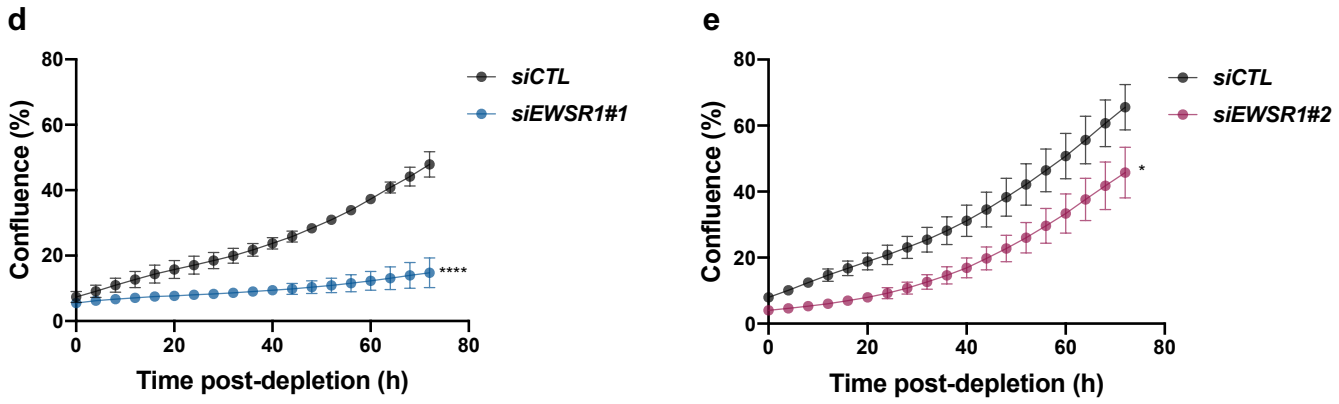


Figure 50: EWSR1 KD leads to cell cycle arrest and decreased proliferation. (**a**) Percentage of HeLa cells in *siCTL* and *siEWSR1#1* conditions in G1, S and G2/M phases of cell cycle. Results are shown as means ± sd (n = 3 independent experiments). *P < 0.05; **P < 0.01 compared to the *siCTL* condition by two-tailed unpaired Student's *t* test. (**b**) Western blotting of EWSR1, P21 and GAPDH (loading control). Samples are total cell lysates from HeLa cells transfected with *siEWSR1#1* or *siCTL*. (**c**) Western blotting of EWSR1 and β-Tubulin (loading control). Samples are total cell lysates from HeLa cells transfected with *siCTL, siEWSR1#1* or *siEWSR1#2*. (**d, e**) Graphs showing the time-dependent (h) evolution of confluence (% of confluence) of HeLa cells post-knockdown of EWSR1 with *siEWSR1#1* (**d**) or *siEWSR1#2* (**e**). Results are shown as means ± sd (n = 3 independent experiments). *P < 0.05; ****P < 0.0001 compared to the *siCTL* condition by two-tailed unpaired Student's *t* test.

Integration of RIBO-seq data with corresponding RNA-seq data allows assessment of mRNA translatability, on a transcriptome-wide basis. Translation efficiency (TE), which is defined as the number of RPF mapped to a particular gene, normalized to the transcript abundance (= the ratio of the RPF counts over mRNA counts for all reads mapped to the gene) is considered as a good proxy for mRNA translatability[442].

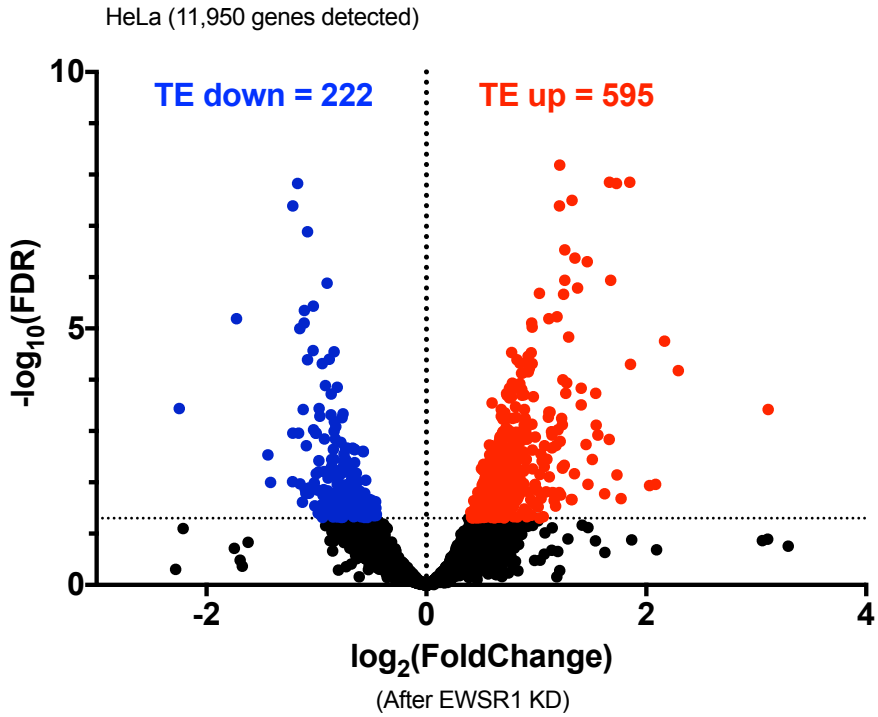
Genome wide analysis of TE revealed significant changes in translation of individual mRNA between control and EWSR1 KD HeLa cells: from 11,950 detected mRNA, we identified a list of 595 mRNA showing a significant increase in TE in EWSR1-depleted conditions (*i.e.*, TE up genes). In contrast, only 222 genes had a decreased TE upon knockdown of EWSR1 (*i.e.*, TE down genes) (**Figure 51a**). This raised the exciting possibility that EWSR1 could carry roles in the regulation of translation. Particularly, the observation that almost 3 times more mRNA showed increased TE following EWSR1 KD, pointed towards a repressive role for EWSR1 in mRNA translation.

Recently, *Li et al.* developed a novel tool to analyse RIBO-seq data, called RiboDiPA[588]. RiboDiPA allows identification of transcripts showing statistically significant differences in ribosome occupancy patterns between two conditions. By applying this algorithm to our RIBO-seq data, we identified 1,213 mRNA with significantly differential pattern (DP) of RPF distribution between *siCTL* and *siEWSR1*#1 conditions. Interestingly, around 40% (38.5%; 229/595) of the above TE up mRNAs were found among these (**Figure 51b**). Complementary to TE analysis, RiboDiPA confirms that depletion of EWSR1 correlates with changes of the translational landscape.

To further investigate the effects of EWSR1 on global mRNA translation, we used a SUrface SEnsing of Translation (SUnSET) assay. This assay allows monitoring and quantification of global protein synthesis by detecting puromycin-labeled polypeptides upon translation elongation blockage[589] (**Figure 52a**). Indeed, puromycin is a structural analogue of aminoacyl-transfer RNA (aminoacyl-tRNA; specifically, tyrosyl-tRNA) that can be incorporated into elongating peptide chains, *via* the formation of a peptide bond. Because it has a non-hydrolysable amide bond between the tRNA ribose moiety and the attached amino acid molecule, incorporation of puromycin leads to elongation termination and therefore release of truncated puromycin-bound peptides[590]. Western blotting with a monoclonal antibody against puromycin did not show an increase in protein synthesis after knocking down EWSR1 using two different siRNA (*siEWSR1#1* and *siEWSR1#2*) (**Figures 52b-e**). This suggests that EWSR1 is not a general inhibitor of translation. Next, we performed polysome profiling by sucrose-gradient separation technique of translating mRNAs[591] (**Figure 53a**). We observed no significant difference between the polysome profiles from *siCTL*- and *siEWSR1#1*-transfected HeLa cells, indicating that knocking down EWSR1 has no effect on global translation (**Figure 53b**).

Taking together, our results suggest that EWSR1 is implicated in the repression of translation of a specific subset of mRNAs.

a



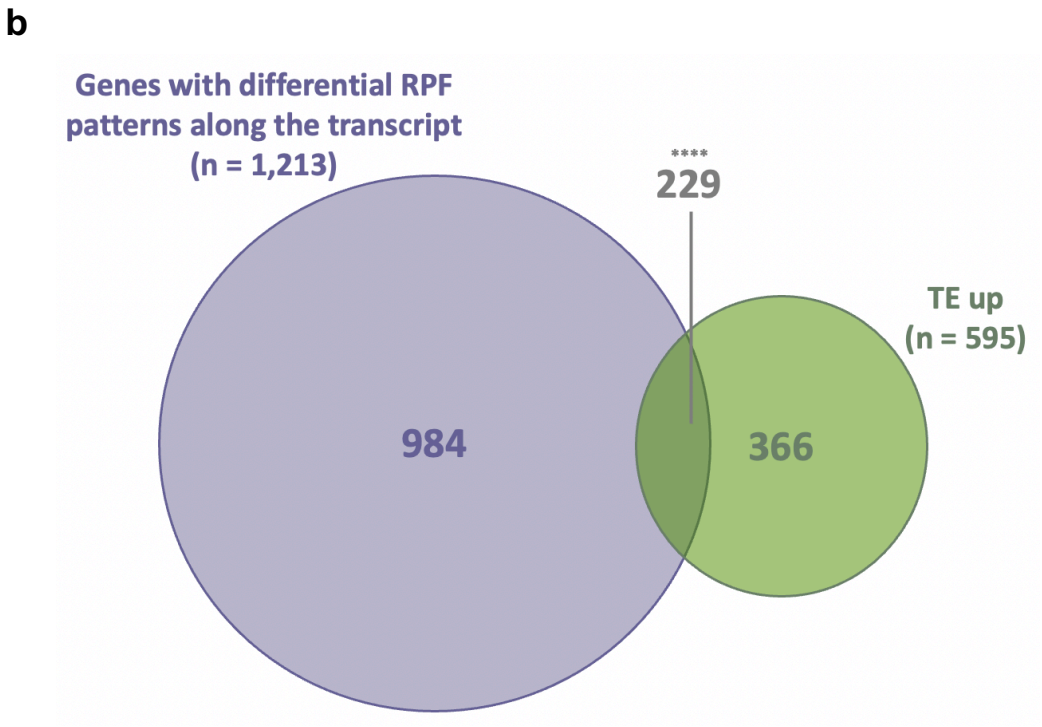
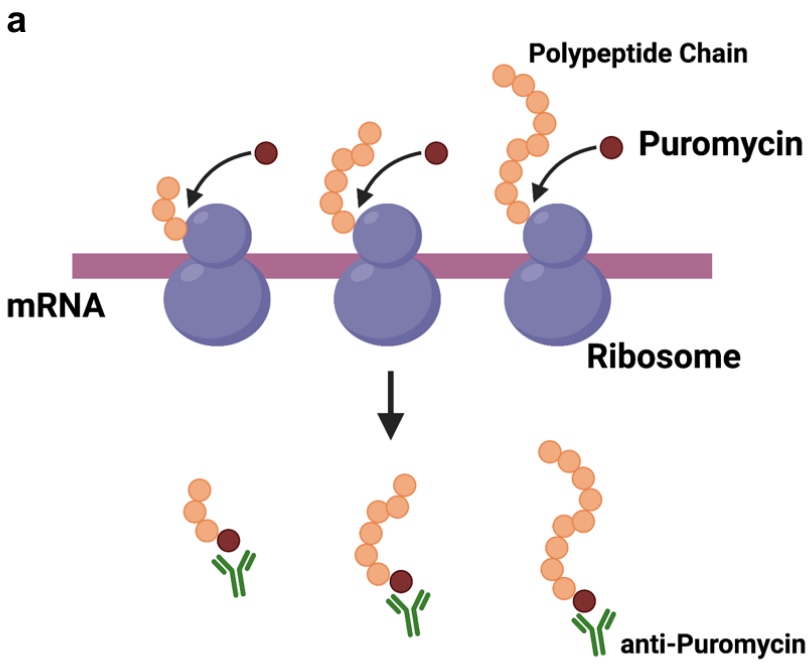


Figure 51: EWSR1 represses the translation of a specific subset of mRNA. (**a**) Volcano plot of significantly downregulated (blue) and upregulated (red) differential translation efficiency genes (DTEGs) upon EWSR1 KD with *siEWSR1#1* in HeLa cells (n = 2 independent experiments, FDR = 0.05). The numbers of DTEGs are indicated in the graph. Black dots are genes that did not show statistically significant differences. (**b**) Venn diagram showing the overlap between genes with differential RPF patterns along the transcript, as established using RiboDiPA and TE up genes. ****$P < 0.0001$ using Fisher's Exact Test.



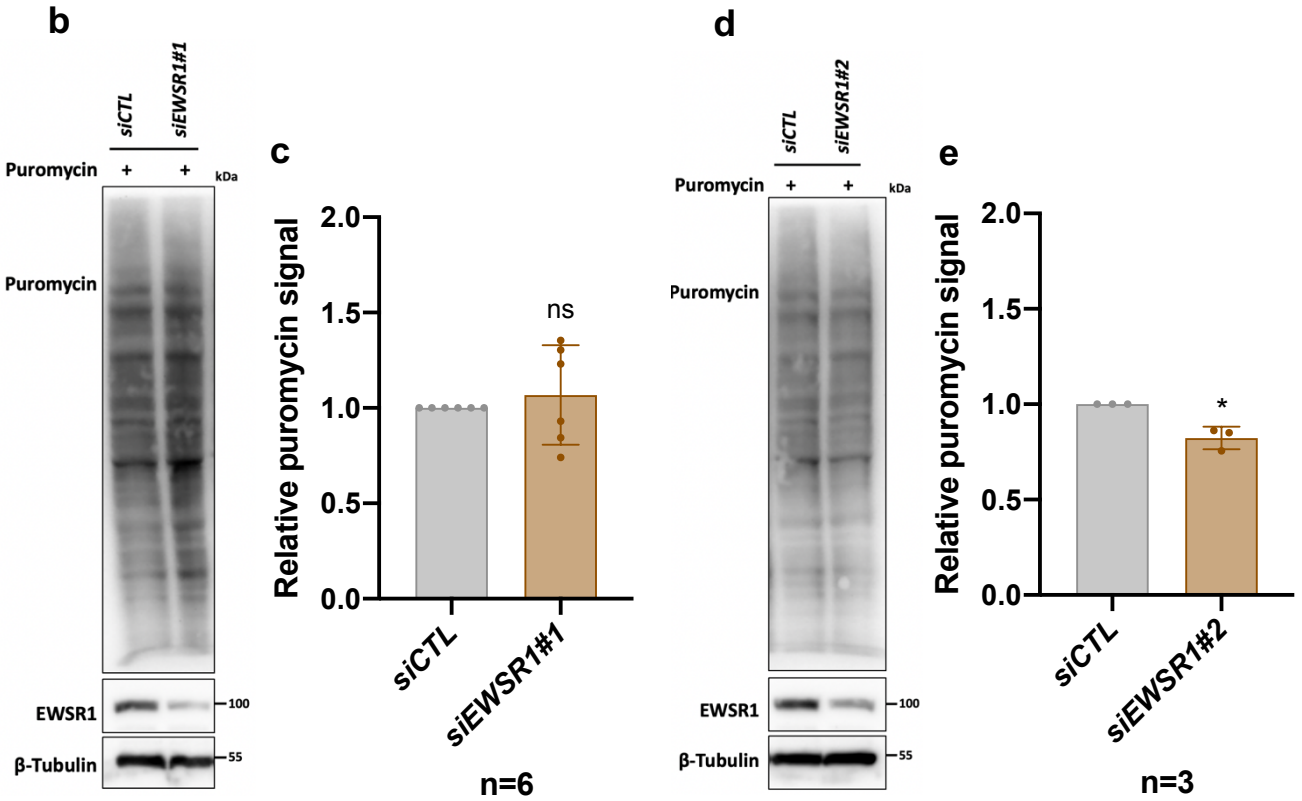


Figure 52: EWSR1 has no effect on global protein synthesis. (**a**) Schematic representation of the SUrface SEnsing of Translation (SUnSET) assay principle. Created with BioRender.com. Puromycin incorporation is revealed by western blotting. (**b, d**) Representative western blot of puromycin, EWSR1 and β-Tubulin (loading control). Samples are total cell lysates from HeLa cells transfected with *siCTL, siEWSR1#1* (**b**) or *siEWSR1#2* (**d**). (**c, e**) Levels of these proteins were quantified by band densitometry analysis using ImageJ. Results are shown as means ± sd (n = 6 independent experiments) (**c**) or (n = 3 independent experiments) (**e**) after normalization to β-Tubulin. **P* < 0.01, ns: not significant compared to the *siCTL* condition by one sample *t* test.

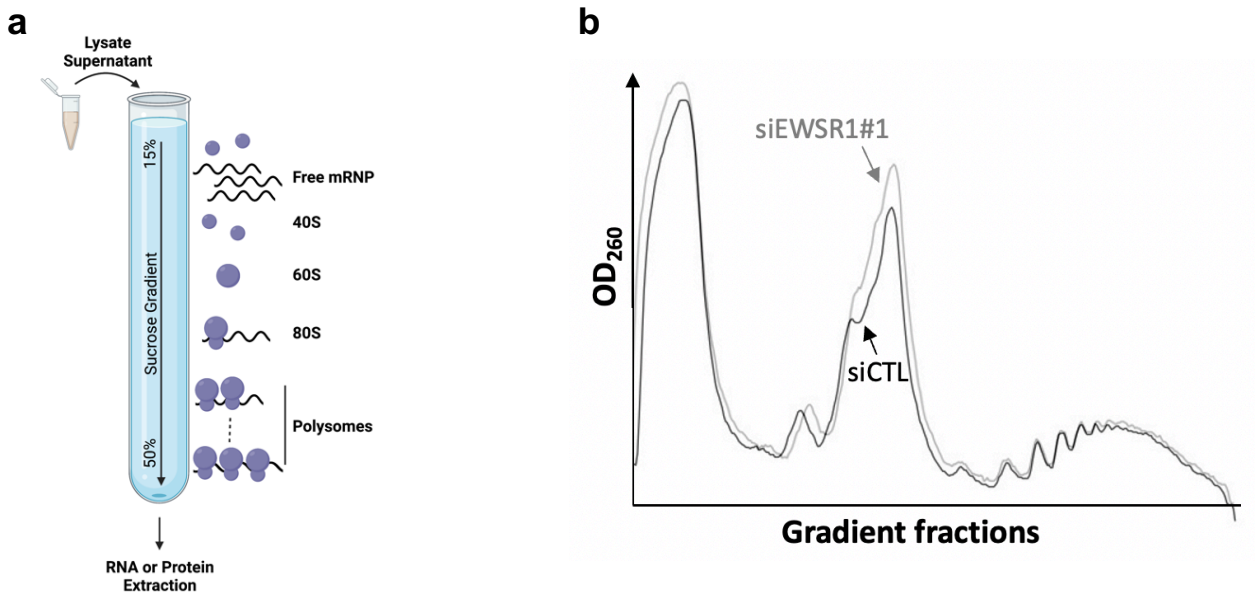


Figure 53: EWSR1 KD has no effect on polysomes profiles. (**a**) Schematic representation of the polysome profiling analysis. Created with BioRender.com. Ribosome-free and polysome-bound mRNAs are separated on a sucrose gradient (15-50%). (**b**) Polysome profiles of HeLa cells showing no difference between *siCTL* and *siEWSR1#1* conditions.

3.2 EWSR1 translational target mRNAs exhibit specific features

To gain molecular insights on how EWSR1 represses translation of specific transcripts, we investigated the sequence features of the TE up mRNA. We observed that on average, 5'UTRs of the TE up mRNA had a significantly higher G- and GC-contents compared to a control set (unchanged) (**Figures 54a,b**). Furthermore, the EWSR1 TE up targets also showed higher occurrence of G-quadruplex within their sequence (**Figure 54c**). In light of previous reports showing that the RGG domains of EWSR1 specifically binds to G-quadruplex RNA and DNA, these observations were highly interesting to us and suggested that EWSR1 might repress translation by directly binding to its target mRNAs[49,592].

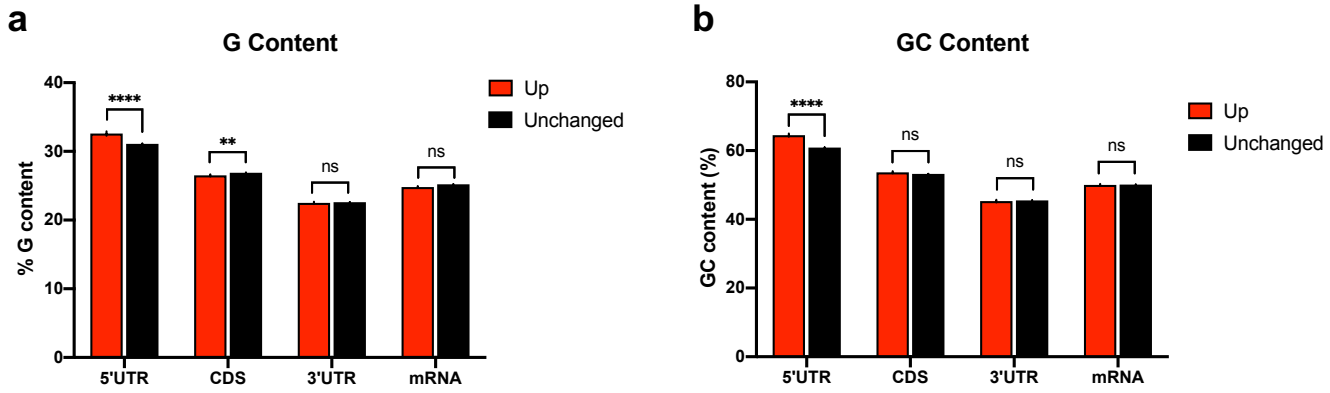
A number of studies highlighted the roles of the 5' and 3'UTRs as modulators of translation[593,594]. To support the model of EWSR1 being a direct translation inhibitor, we performed an RBP enrichment analysis in mRNAs from the TE up dataset using AME, based on the motifs highlighted from *Ray et al.*[595]. We found that 5'UTRs of TE up transcripts were enriched in motifs associated with PPRC1, RBM4, RBM8A, ZC3H10, LIN28A and the EWSR1 paralog FUS (**Figure 55a**). The 3'UTRs of TE up mRNAs were enriched in different motifs, including those of CPEB4, HuR, TIA1 and PCBP1 (**Figure 55b**). Remarkably, some identified RBPs, notably RBM4, FUS, LIN28A and TIA1, have been associated with repression of translation (**Figure 55c**). In addition, database curation of the BioGRID[4.4] repository revealed that some of the identified RBPs are known partners of EWSR1, such as RBM8A, FUS, RALY, HuR, TIA1 and U2AF2. This suggests that EWSR1 might repress the translation of its target mRNA by

collaborating with known translation inhibitors, for instance by enhancing their recruitment onto EWSR1 translational targets. An alternative, although not antagonistic model, could be that some EWSR1-partner RBP may increase EWSR1 binding to its target mRNAs, as collaborative interplay between RBPs is now well-documented[596].

Subsequently, we specifically searched for potential EWSR1 binding motifs in TE up mRNAs. As the EWSR1 binding motif remains poorly-defined, we turned towards FUS, another highly related FET protein family member. Recently, *Loughlin et al.*[597] showed that FUS uses a bipartite binding mode with its ZnF binding to GGU and its RRM binding to an RNA stem loop. Interestingly, we found such defined FUS bipartite motif to be significantly enriched in TE up mRNAs. The enrichement was found to be significant for the 3' and 5'UTRs, but also for the coding sequence (CDS) of EWSR1 translational targets (**Figure 55d**).

Another study conducted by *Van Nostrand et al.*[598] combined RNA Bind-N-Seq (RBNS) *in vitro* analysis, with enhanced CLIP (eCLIP) to characterize RNA elements associated with a large collection of human RBPs, generated as part of the Encyclopedia of DNA Elements (ENCODE) project phase III. They identified GGG(G|T|A)G as being the most likely EWSR1-associated motif. This motif was also significantly enriched in the TE up mRNAs, specifically in the 3' and 5'UTRs (**Figure 55e**).

Although correlative, these analyses all point towards a model in which mRNAs whose translation is inhibited by EWSR1 harbor specific structural and sequence features such as high GC- and G-contents, high incidence of rG4 structures, FUS bipartite motifs and EWSR1 RBNS sequences, all of which are compatible with direct binding of EWSR1 to its translational targets.



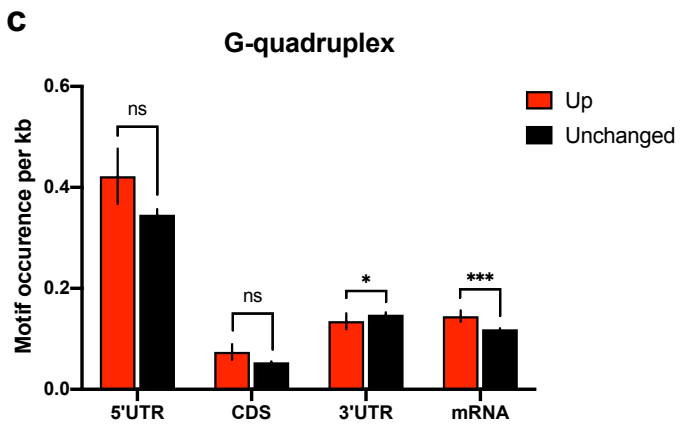
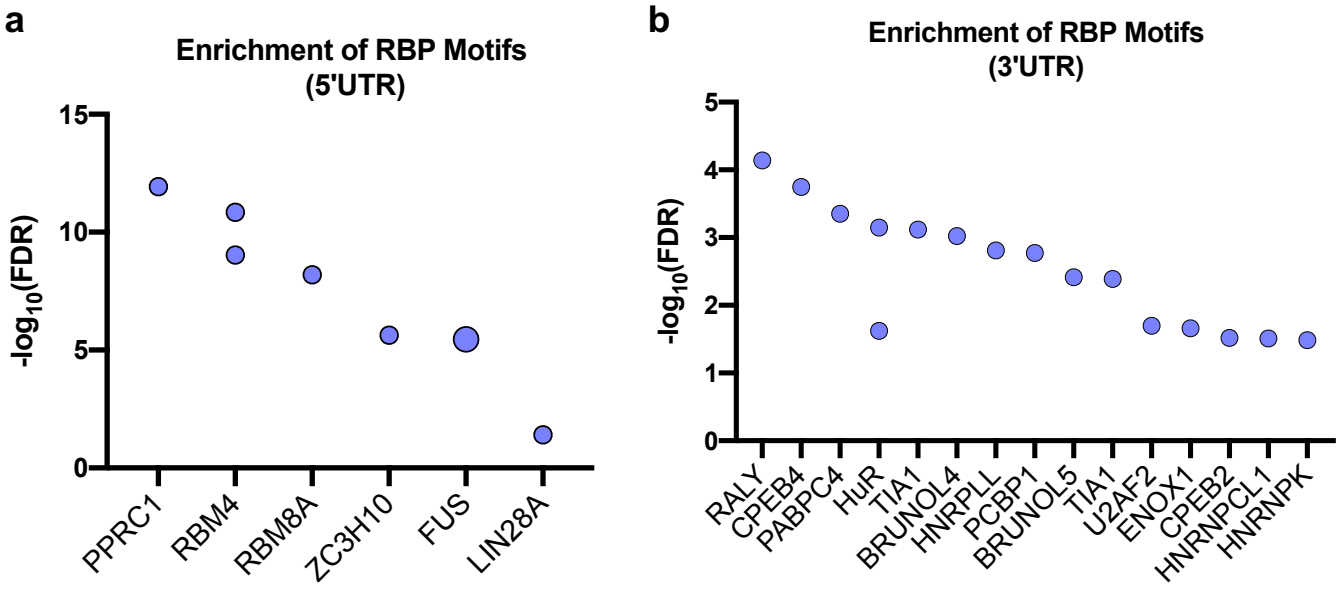


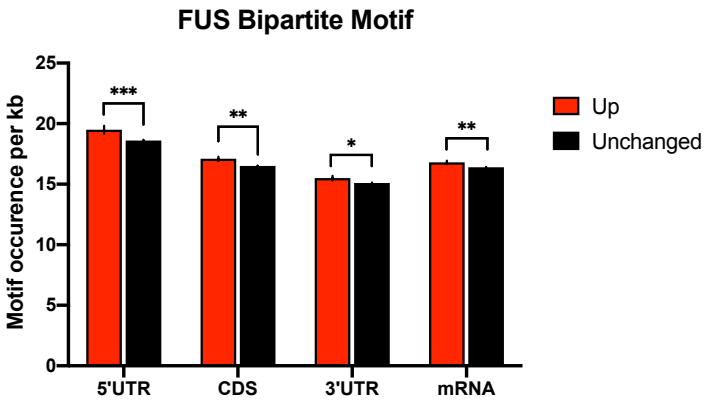
Figure 54: Features of EWSR1 TE up targets. (**a, b, c**) Comparison of the percentage of G content (**a**), GC content (**b**) and G-quadruplex (**c**) between unchanged and TE up genes in the 3' and 5'UTRs, CDS and the entirety of the mRNA. **P* < 0.05, ***P* < 0.01; ****P* < 0.001, *****P* < 0.0001, ns = not significant using Wilcoxon-Mann-Whitney test. Results are shown as ± SEM (n = 2 independent experiments).



c

RBPs	References
RBM4	599
ZC3H10	600
FUS	29
LIN28A	601
CPEB4	602
HuR	603
TIA1	604
hnRNP K	605

d



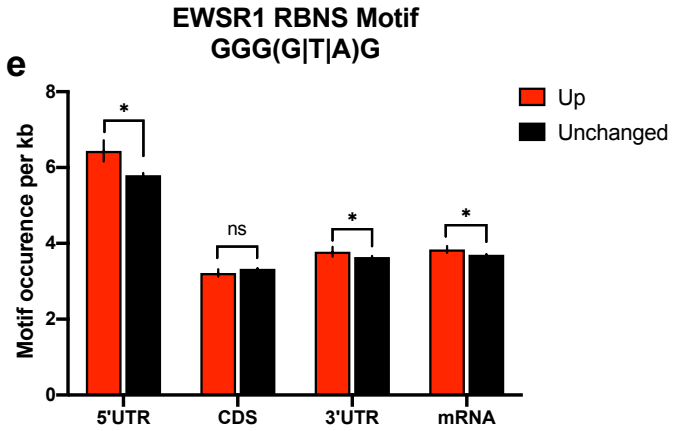


Figure 55: Sequence features of EWSR1 TE up targets are compatible with a model of direct binding of EWSR1. (**a, b**) RNA-binding protein motif enrichment analysis of 5' (**a**) and 3'UTRs (**b**) of EWSR1 TE up targets. These sequences were screened for RNA-binding motifs using AME from the MEME suite. Enriched motif(s) of RBPs with an adjusted *P*-value (padj) < 0.05 are shown. (**c**) Table showing RBPs implicated in the repression of translation with their corresponding references. (**d, e**) Comparison of FUS bipartite motif (**d**) and EWSR1 RBNS motif (**e**) occurrences per kb between unchanged and TE up genes in the 3' and 5'UTRs, CDS and the entirety of the mRNA. **P* < 0.05, ***P* < 0.01; ****P* < 0.001, ns = not significant using Wilcoxon-Mann-Whitney test. Results are shown as ± SEM (n = 2 independent experiments).

3.3 EWSR1 binds its translational targets and represses the translation of a tethered reporter

To support the model whereby EWSR1 binds to specific mRNA to repress their translation, we sought to test the ability of EWSR1 to bind to TE up transcripts. To this aim, we identified EWSR1 RNA targets based on previously published CLIP-seq data from *Hoel at al.*[51], *Paronetto et al.*[52] and *Van Nostrand et al.*[598]. We defined a "high-confidence dataset" of EWSR1-bound mRNAs that consisted of the 2,731 common mRNAs found in the three CLIP-seq datasets. Interestingly, more than 20% (22.5%; 134/595) of the TE up mRNAs were found in the set of high-confidence EWSR1-bound mRNAs, an overlap that was highly significant (**Figure 56a**). Because these mRNAs can be bound by EWSR1 and their translation is increased in the absence of EWSR1, we defined these 134 overlapping transcripts as “direct EWSR1 translational targets”. To cross-validate the binding of EWSR1 to these mRNA in HeLa cells, we performed RNA immunoprecipitation (RIP)-qPCR experiments. We immunoprecipitated EWSR1 with an anti-EWSR1 antibody (**Figure 56b**) and assessed the presence of 4 randomly chosen mRNAs from the “direct EWSR1 translational targets” list (*i.e., SEC61G, KDELR1, ZNF664* and *RPL32)* in the IP by RT-qPCR. Results were normalized to an unrelated *18S* rRNA. In contrast to GAPDH, a non-target mRNA, we found that 3 of the 4 tested mRNAs (*i.e., SEC61G, KDELR1* and *ZNF664)* were indeed enriched in the EWSR1 IP, compared to the IgG control IP (**Figure 56c**). This indicates that EWSR1 has the ability to associate with mRNAs whose it represses translation.

To validate this model, we used a previously described mRNA tethering assay[606]. This assay uses a *Renilla luciferase* mRNA (*RLuc*) reporter harboring in its 3’UTR, either 0 or 4 repeats of the binding sequence for the bacteriophage MS2 coat protein (MS2-CP). These reporters are designated as *RLuc-0* and *RLuc-4*, respectively. Another *Firefly* luciferase reporter gene (*FLuc*) is transcribed from the same bidirectional CMV promoter, but does not bear any MS2 binding site. This allows *FLuc* to be used as an internal control for transfection and recovery efficiencies. EWSR1 was fused to the MS2-CP peptide, allowing its specific recruitment to the 3’UTR of the *RLuc-4* reporter transcript (**Figure 57a**). In contrast, due to the absence of MS2 binding site in its 3'UTR, the *RLuc-0* mRNA is not tethered by MS2 fusion proteins.

To test the effect of EWSR1 on the expression of the *RLuc* reporter mRNA, we transfected HeLa cells with FLAG-tagged constructs encoding either MS2-CP alone (as control) or MS2-CP fused to the full-length EWSR1 construct (MS2-EWSR1), together with the *RLuc-0* (as the control reporter) or *RLuc-4* (as the tethered reporter). For each condition (*i.e.*, *RLuc-0* or *RLuc-4*), the levels of MS2-EWSR1 were comparable (**Figure 57b**). The results were expressed as ratios of normalized *RLuc*/*FLuc* activities. As shown in **Figure 57c**, recruitment of the MS2-EWSR1 protein to the *RLuc* transcript specifically reduced the expression of the *RLuc* protein by more than 50% (54%) compared to the MS2-CP control condition. Quantitative mRNA analysis by RT-qPCR showed no effect on the abundance of the *RLuc* transcript (**Figure 57d**), suggesting that

EWSR1 protein reduces the production of the *Renilla* protein when tethered to the coding mRNA, without affecting the abundance of the transcript. It was recently suggested that FUS might inhibit translation by recruiting the core miRISC component AGO2 to specific transcript[607]. However, our bidirectional reporter has no previously identified miRNAs sites, and knockdown of AGO2 (**Figure 58a**) did not affect the ability of MS2-EWSR1 to inhibit translation of the *Renilla* luciferase (**Figures 58b,c**), suggesting that EWSR1 translational inhibition is not mediated by the miRNA machinery. Importantly, we observed no effect when we performed the same assay with a non-tethered FLAG-EWSR1 construct, indicating that the reduction of *RLuc* activity by EWSR1 requires its tethering to the reporter mRNA (**Figures 59a,b**). Altogether, these observations support a model whereby EWSR1 represses mRNA translation through its binding onto its RNA targets.

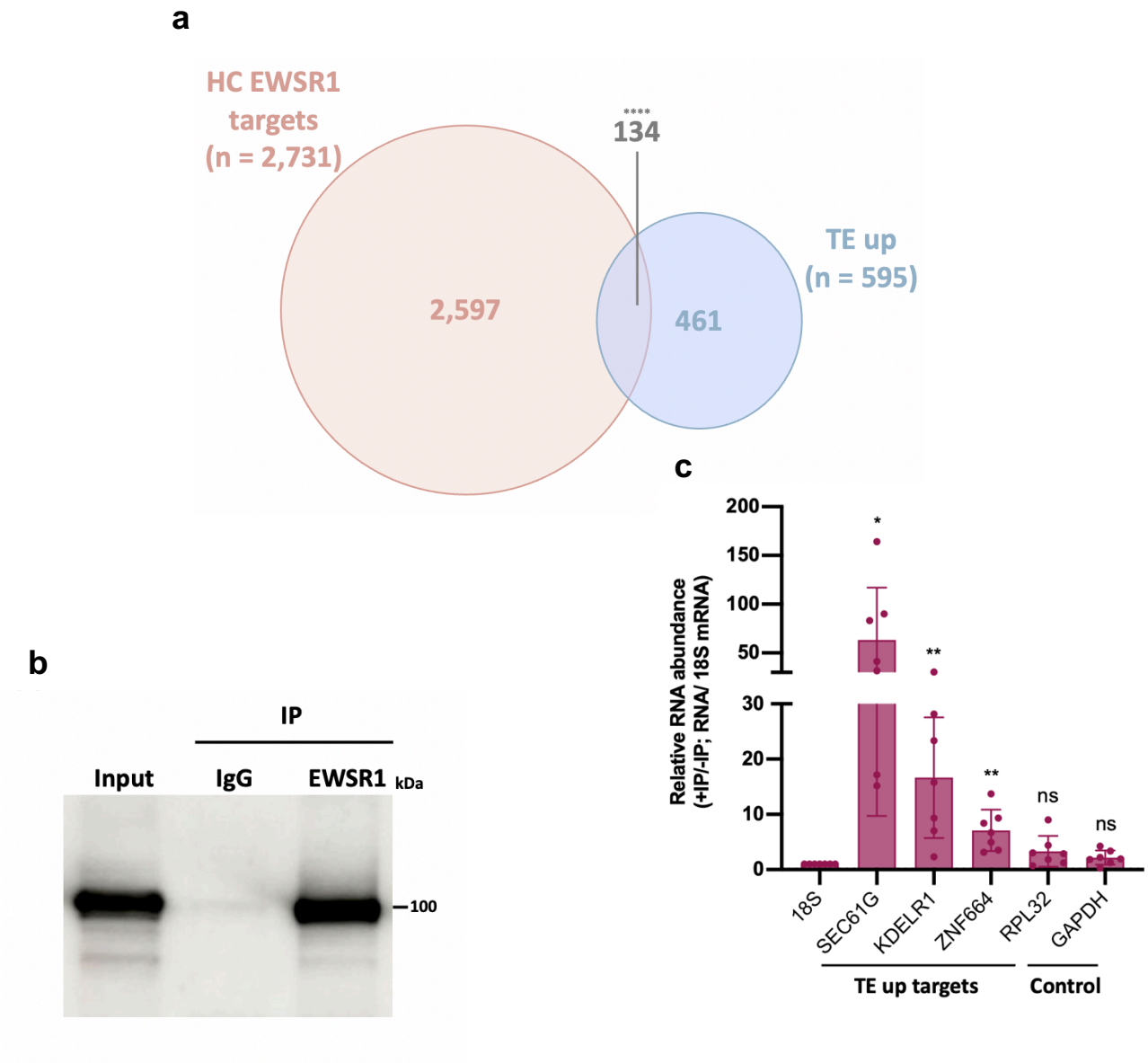
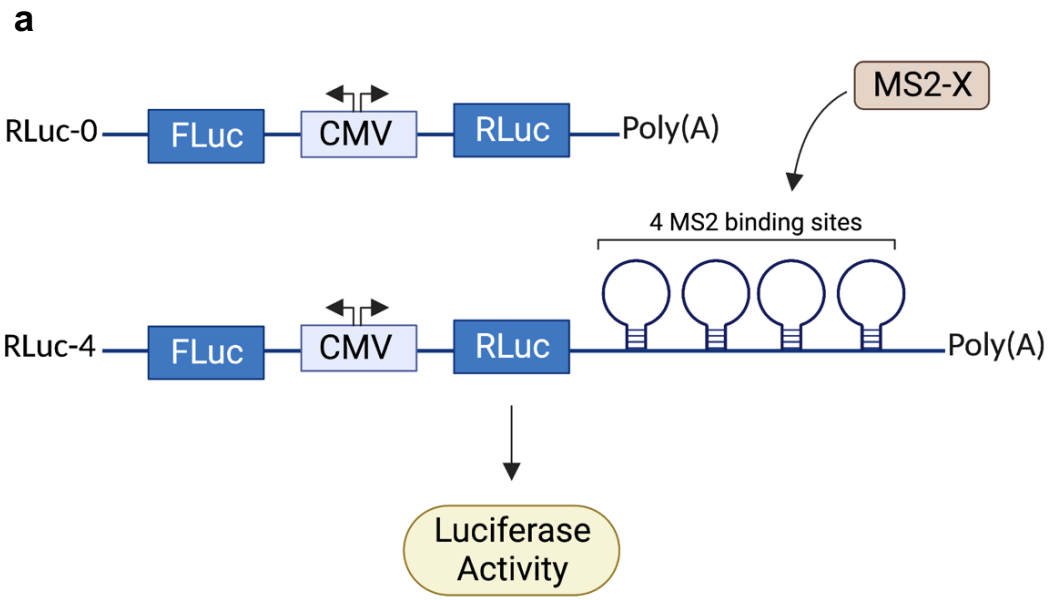
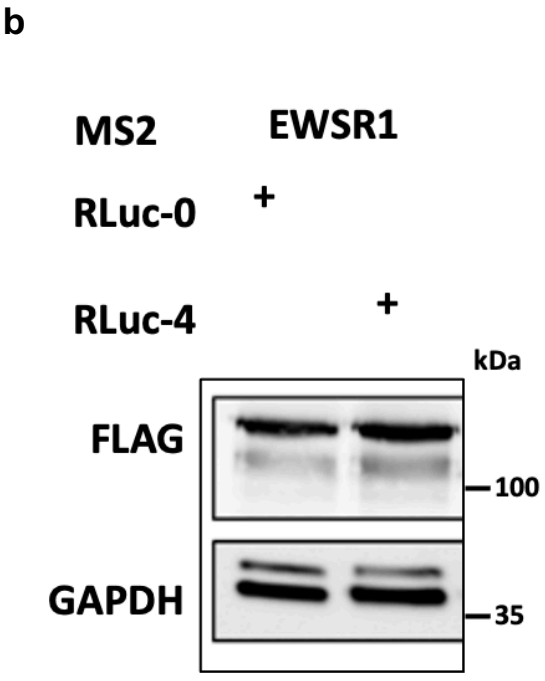


Figure 56: EWSR1 binds its translational targets. (**a**) Venn diagram showing the overlap between High Confidence (HC) EWSR1 targets curated from literature and TE up genes. ****$P < 0.0001$ using Fisher's Exact Test. (**b**) Western blot analysis of EWSR1 after RNA immunoprecipitation (RIP) in HeLa cells. (**c**) RT-qPCR detection of various EWSR1 TE up mRNA targets and control gene (GAPDH) in RIP from HeLa cells (n = 7 independent experiments). *$P < 0.05$, **$P < 0.01$, ns = not significant using one sample *t* test.





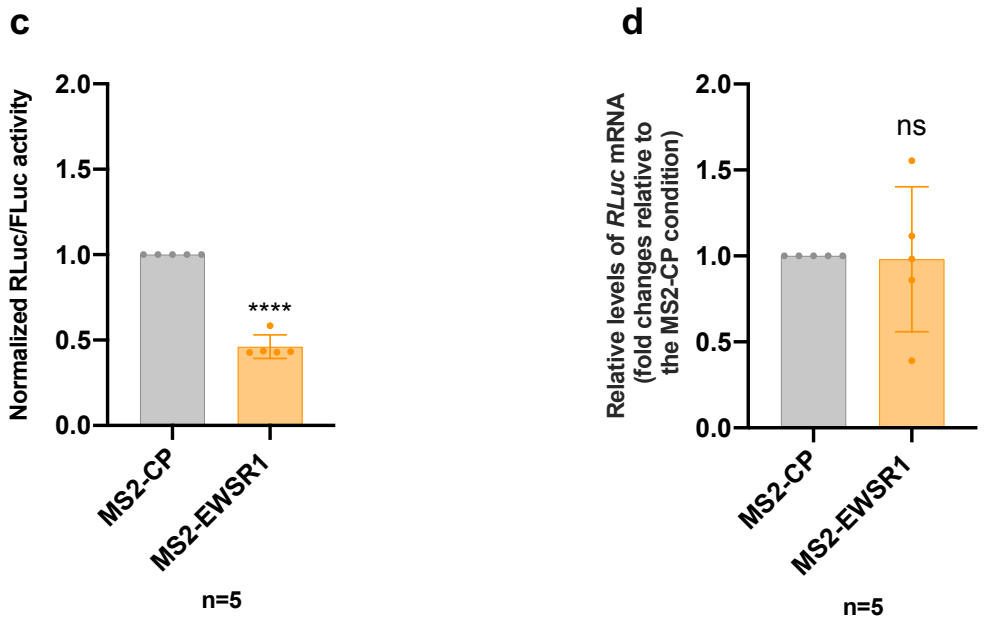
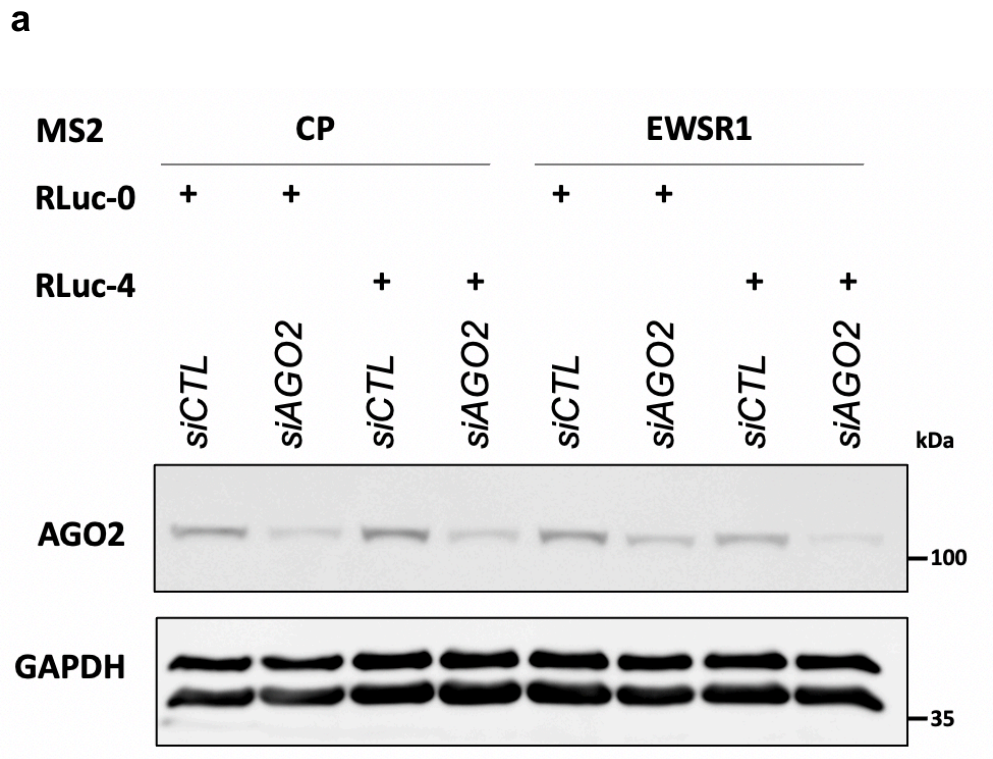


Figure 57: Tethering of EWSR1 represses translation of a *Renilla luciferase* mRNA reporter. (**a**) Schematic representation of tethering assay principle. Created with BioRender.com. Protein (X, *e.g.*, EWSR1) fused to the MS2-Coat Protein (MS2-CP) recognizes MS2-binding sites in the 3'UTR of the *Renilla luciferase* mRNA reporter with high specificity and affinity. A CMV bidirectional promoter controls the transcription of both *Firefly* and *Renilla* luciferases. The *FLuc* and *RLuc* activities are assayed using the Dual-Luciferase Reporter (DLR) Assay system. (**b**) Western blot analysis of the levels of MS2-EWSR1 with anti-FLAG antibody. GAPDH is used as a loading control. Samples are total cell lysates from HeLa cells transfected with MS2-EWSR1 together with the *RLuc-0* or *RLuc-4* reporters constructs. (**c, d**) Luciferase tethering assays in HeLa cells transfected with MS2-CP or MS2-CP fused EWSR1 constructs together with *RLuc-4* or *RLuc-0* reporters. Results are means of normalized *RLuc/FLuc* activities [(*RLuc-4/FLuc*)/(*RLuc-0/FLuc*)] (**c**) and *RLuc* mRNA levels, relative to the MS2-CP condition (**d**) ± s.d. (n = 5 independent experiments). ****P < 0.0001, ns = not significant compared to the MS2-CP condition by one sample *t* test.



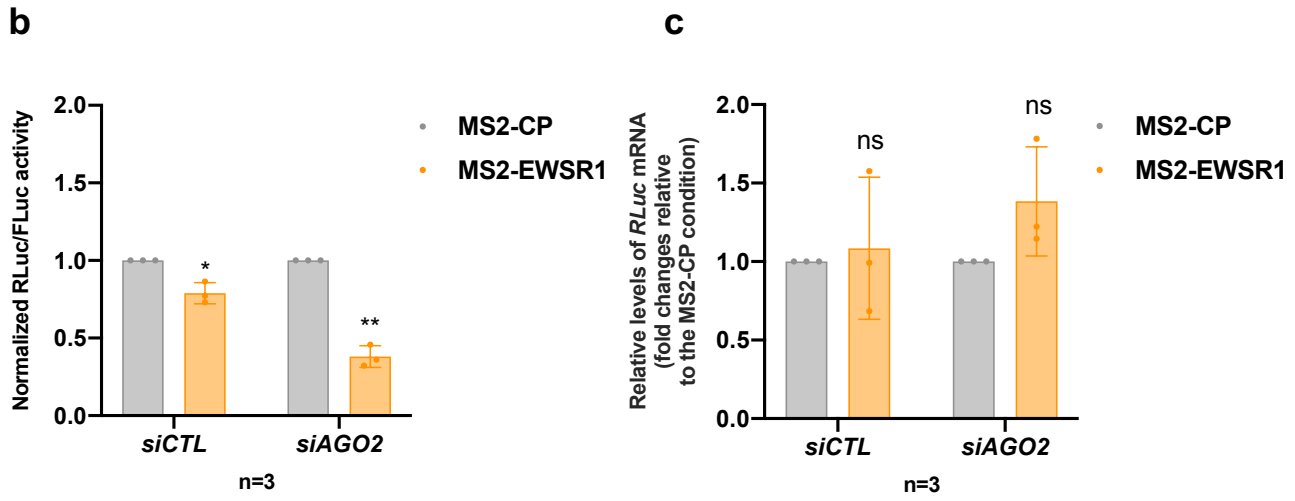


Figure 58: Repression of translation by EWSR1 does not rely on the miRNA machinery. (**a**) Western blot analysis of the levels of AGO2. GAPDH is used as a loading control. Samples are total cell lysates from HeLa cells transfected with MS2-CP or MS2-EWSR1 together with the *RLuc-0* or *RLuc-4* reporters constructs, in *siCTL* or *siAGO2* conditions. (**b, c**) Luciferase tethering assays in HeLa cells transfected with MS2-CP or MS2-CP fused EWSR1 constructs together with *RLuc-4* or *RLuc-0* reporters in *siCTL* or *siAGO2* conditions. Results are means of normalized *RLuc/FLuc* activities [(*RLuc-4/FLuc*)/(*RLuc-0/FLuc*)] (**b**) and *RLuc* mRNA levels, relative to the MS2-CP condition (**c**) ± s.d. (n = 3 independent experiments). *P < 0.05, **P < 0.01, ns = not significant compared to the MS2-CP condition by one sample *t* test.

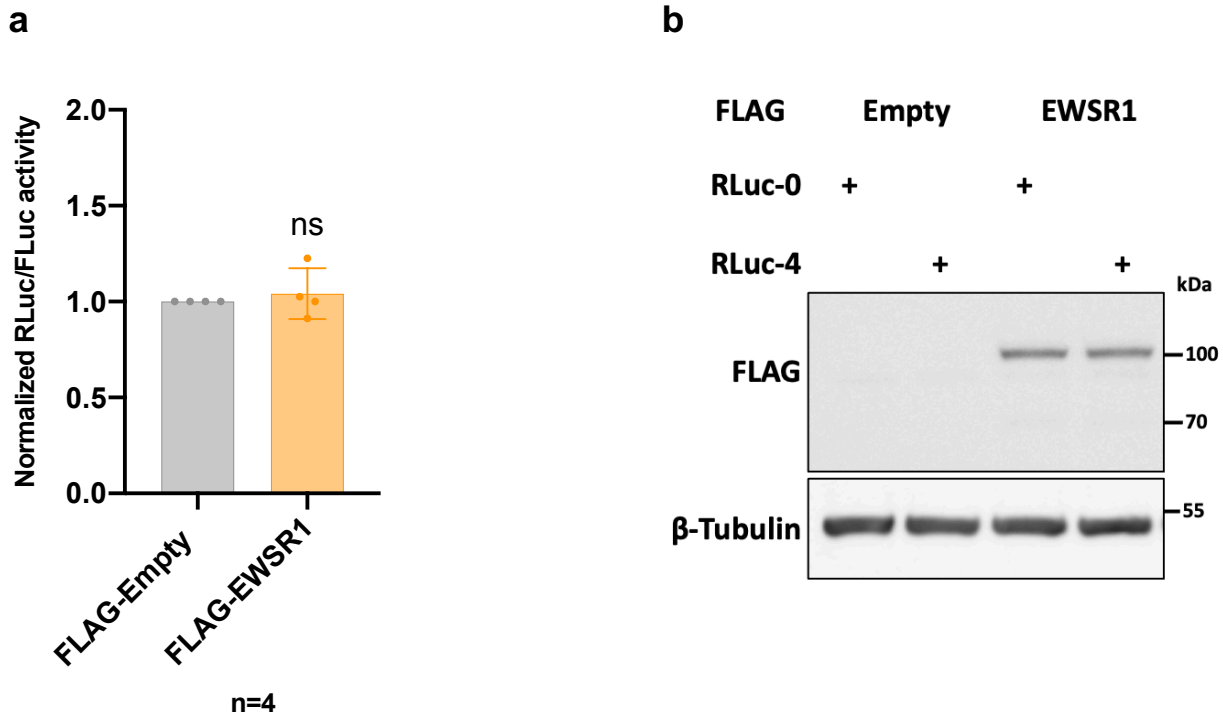
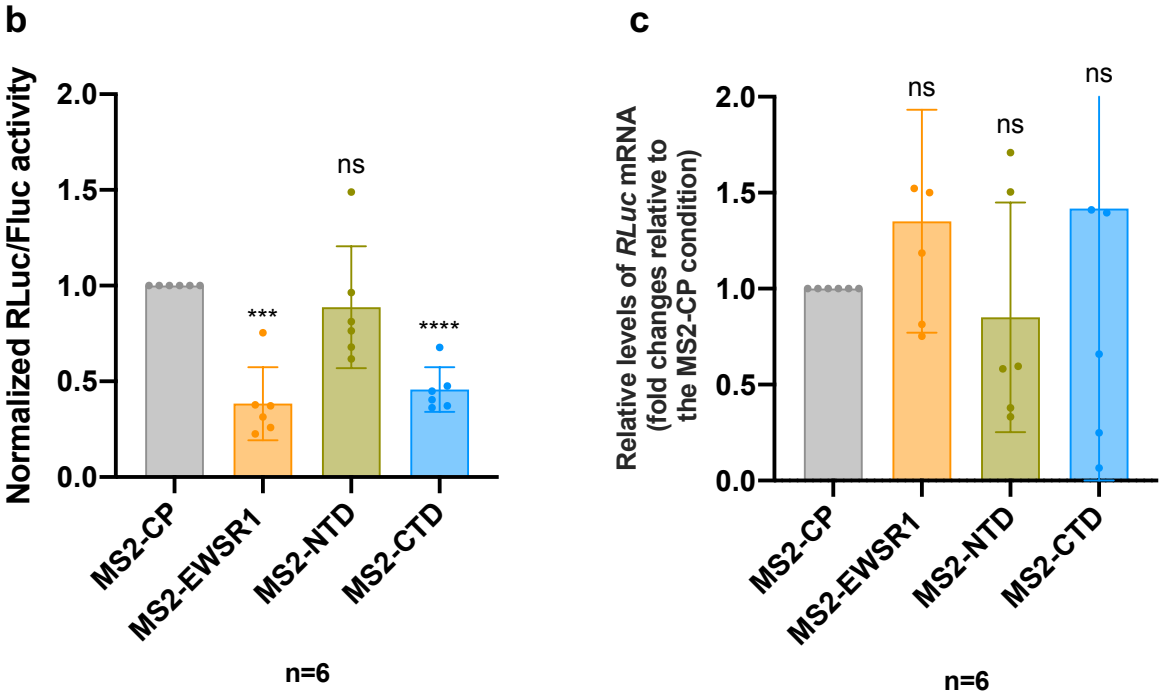


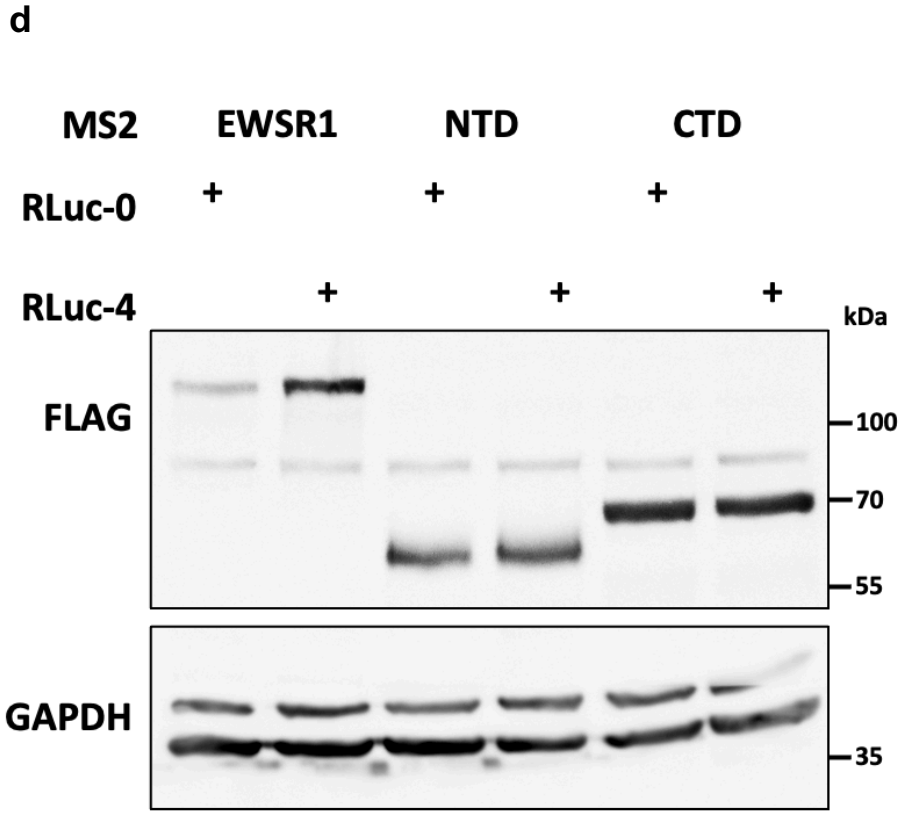
Figure 59: Non-tethered FLAG-EWSR1 does not repress the translation. (**a**) Luciferase tethering assays in HeLa cells transfected with a control empty vector (FLAG-Empty) or a vector expressing FLAG-tagged EWSR1 together with *RLuc-4* or *RLuc-0* reporters. Results are means of normalized *RLuc/FLuc* activities [(*RLuc-4/FLuc*)/(*RLuc-0/FLuc*)] ± s.d. (n = 4 independent experiments). ns = not significant compared to the FLAG-empty condition by one sample *t* test. (**b**) Western blot analysis of the levels of FLAG-Empty and FLAG-EWSR1 using an anti-FLAG antibody. β-Tubulin is used as a loading control. Samples are total cell lysates from HeLa cells transfected with FLAG-Empty or FLAG-EWSR1 together with *RLuc-0* or *RLuc-4* reporters constructs.

3.4 The RGG2-ZnF-RGG3 region is responsible for the translational function of EWSR1

To gain some molecular insights into its translational function, we thought to determine which region of EWSR1 was implicated in the repression of translation (**Figure 60a**). To this aim, we first independently tested the translational repressive activity of the N- and C-terminal moieties (CTD and NTD respectively) of EWSR1 in the MS2-tethering reporter assay. We transfected HeLa cells with FLAG-tagged constructs encoding either MS2-CP alone (as control) or MS2-CP fused to full-length EWSR1, its NTD or CTD regions. As previously shown (**see Figures 57c,d**), tethering of MS2-EWSR1 reduced the expression of the *RLuc* protein by more than 50%, with no effect on the abundance of the *RLuc* transcript, as assessed by RT-qPCR analysis. This translational inhibition of the reporter by full-length EWSR1 could be recapitulated by specifically tethering of its CTD region, but not its NTD region (**Figures 60b,c**), although both EWSR1 regions were expressed at comparable levels (**Figure 60d**). As control, inhibition of *RLuc* translation by EWSR1 CTD was not observed when transfecting an untethered version of the protein, *i.e.,* not fused to MS2-CP (**Figures 60e,f**), thus confirming that EWSR1-mediated translational repression requires its presence on the target transcript. To further support these observations, we used another previously described reporter, *RLuc*-6xMS2-polyA, containing 6 MS2 binding sites[608] (**Figure 60g**). Using this alternative reporter, the CTD exhibited an even stronger effect, reducing *RLuc* expression by more than 80% (83%) (**Figures 60h,i**). Altogether, these results identify the CTD as the region of EWSR1 responsible for its translational repressive function.

a


b
Normalized RLuc/Fluc activity
2.0
1.5
1.0
0.5
0.0
***
ns
****
MS2-CP
MS2-EWSR1
MS2-NTD
MS2-CTD
n=6
c
Relative levels of RLuc mRNA (fold changes relative to the MS2-CP condition)
2.0
1.5
1.0
0.5
0.0
ns
ns
ns
MS2-CP
MS2-EWSR1
MS2-NTD
MS2-CTD
n=6


d
MS2
EWSR1
NTD
CTD
RLuc-0
+
+
+
RLuc-4
+
+
+
kDa
FLAG
100
70
55
GAPDH
35

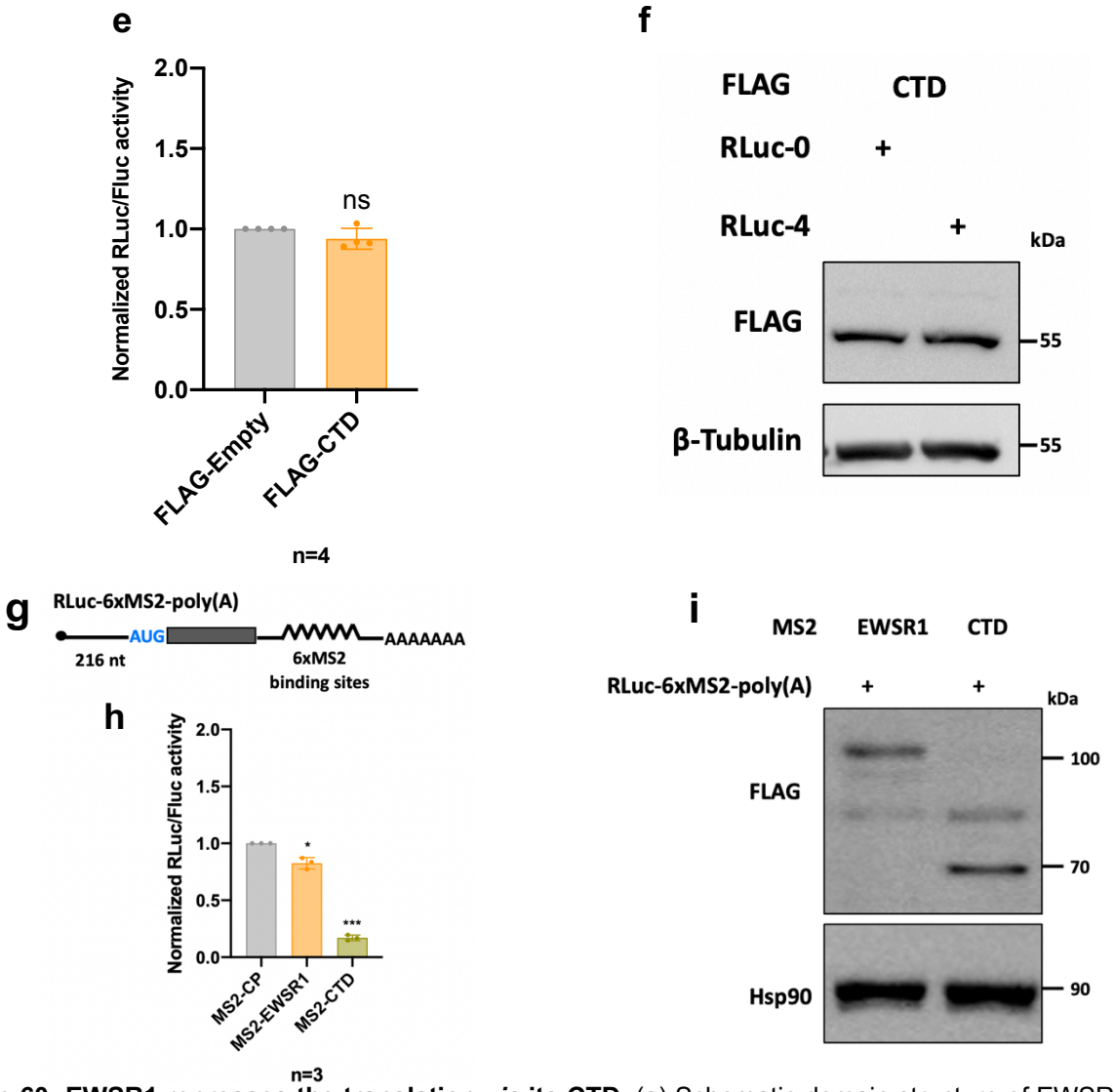
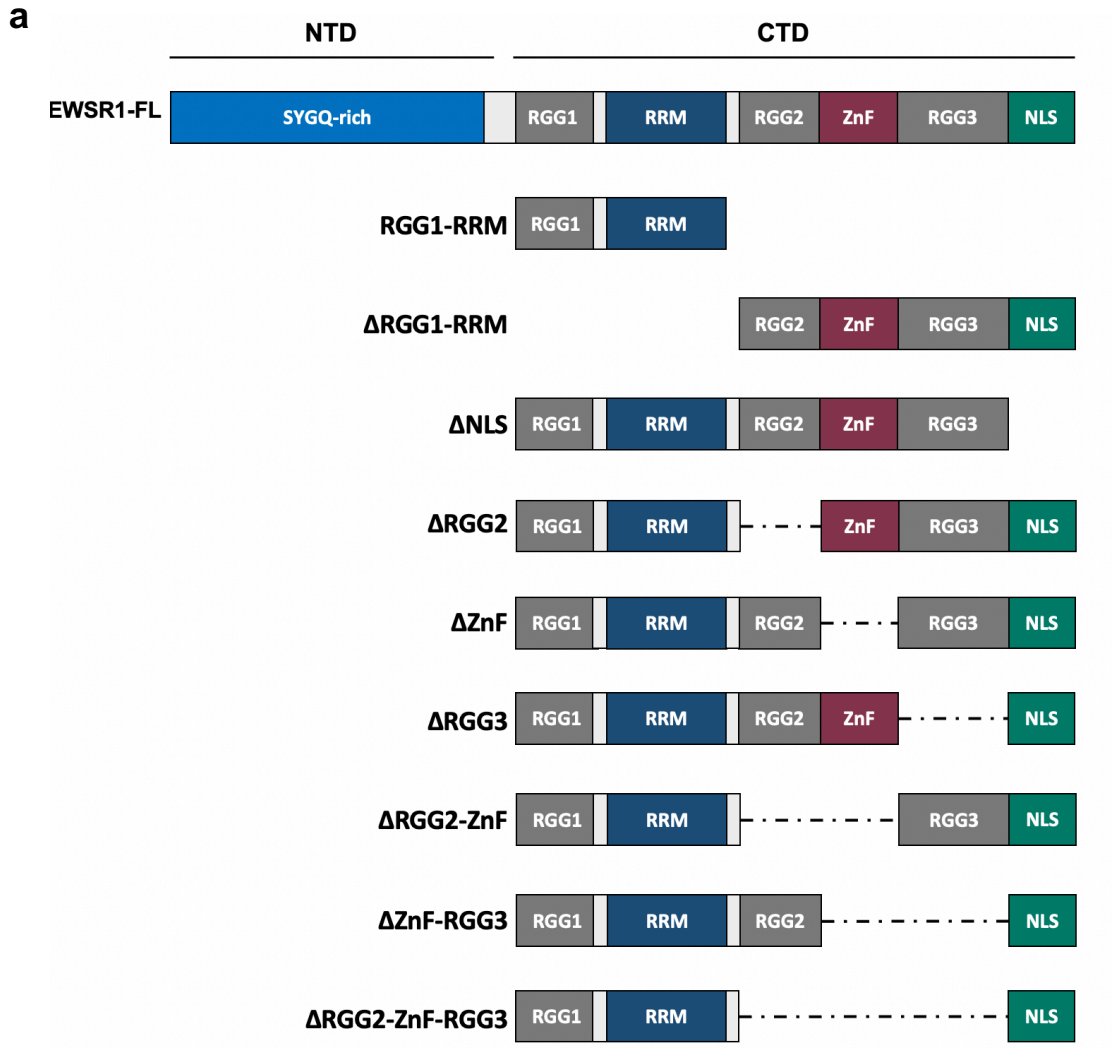


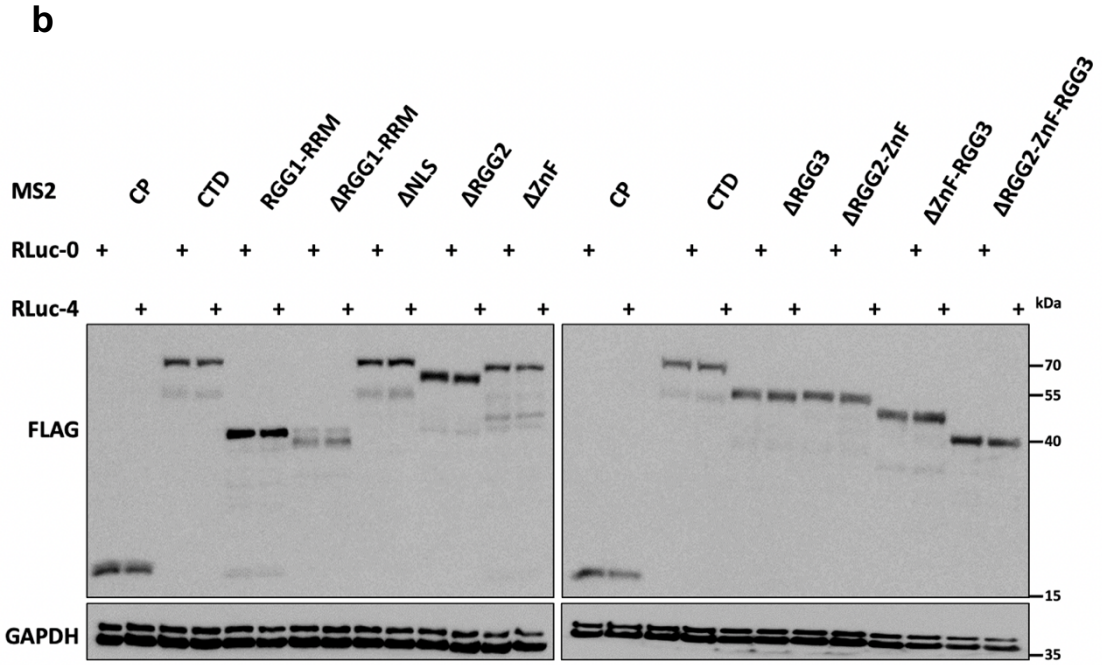
Figure 60: EWSR1 represses the translation *via* its CTD. (**a**) Schematic domain structure of EWSR1. NTD: N-terminal domain, CTD: C-terminal domain. (**b, c**) Luciferase tethering assays in HeLa cells transfected with MS2-CP or MS2-CP fused EWSR1, NTD or CTD constructs together with *RLuc-4* or *RLuc-0* reporters. Results are means of normalized *RLuc/FLuc* activities [(*RLuc-4/FLuc*)/(*RLuc-0/FLuc*)] (**b**) and *RLuc* mRNA levels, relative to the MS2-CP condition (**c**) ± s.d. (n = 6 independent experiments). ***P <0.001, ****P < 0.0001, ns: not significant compared to the MS2-CP condition by one sample *t* test. (**d**) Western blot analysis of the levels of MS2-EWSR1, MS2-NTD and MS2-CTD with anti-FLAG antibody. GAPDH is used as a loading control. Samples are total cell lysates from HeLa cells transfected with MS2-CP fused EWSR1, NTD or CTD constructs together with *RLuc-0* or *RLuc-4*. (**e**) Luciferase tethering assays in HeLa cells transfected with FLAG-Empty or a vector expressing FLAG-tagged CTD constructs together with *RLuc-4* or *RLuc-0* reporters. Results are means of normalized *RLuc/FLuc* activities [(*RLuc-4/FLuc*)/(*RLuc-0/FLuc*)] ± s.d. (n = 4 independent experiments). ns = not significant compared to the FLAG-Empty condition by one sample *t* test. (**f**) Western blot analysis of the levels of FLAG-CTD using an anti-FLAG antibody. β-Tubulin is used as a loading control. Samples are total cell lysates from HeLa cells transfected with FLAG-CTD together with *RLuc-0* or *RLuc-4* reporters constructs. (**g**) Schematic representation of the reporter *RLuc*-6xMS2-poly(A) with 6 MS2 binding sites used in panel (**h**). (**h**) Luciferase tethering assays in HeLa cells transfected with MS2-CP or MS2-CP fused EWSR1 or CTD constructs together with *RLuc*-6xMS2-poly(A) and pGL3-*Firefly* reporters. Results are means of normalized *RLuc/FLuc* activities ± s.d. (n = 3 independent experiments). *P < 0.05, ***P <0.001 compared to the MS2-CP condition by one sample *t* test. (**i**) Western blot analysis of the levels of MS2-EWSR1 and MS2-CTD proteins with anti-FLAG antibody. HSP90 is used as a loading control. Samples are total cell lysates from HeLa cells transfected with MS2-EWSR1 or MS2-CTD together with *RLuc*-6xMS2-poly(A) and pGL3-*Firefly* reporters constructs.

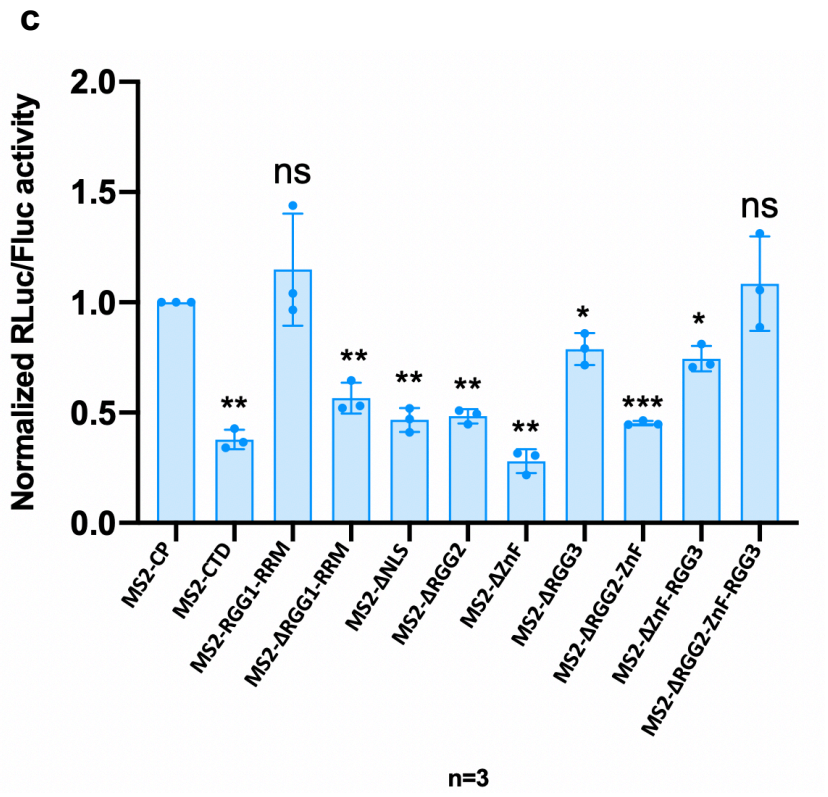
To better map the sub-region of the CTD implicated in the inhibition of translation, we generated

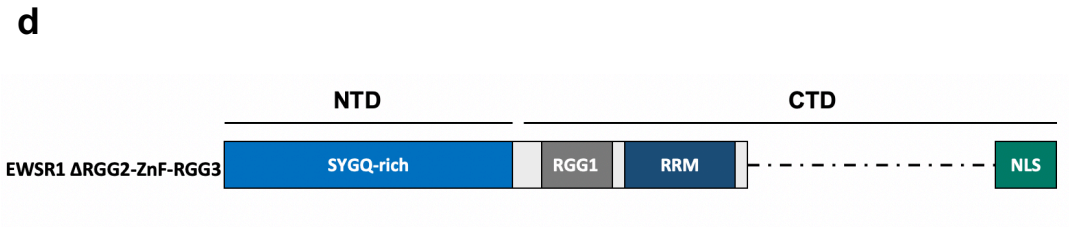
a series of 9 mutants lacking specific structural and/or functional domains within the EWSR1 CTD. We then tested individually each of these mutants CTD in the tethering reporter assay (**Figure 61a**). Among the tested mutants, which were all expressed at comparable levels (**Figure 61b**), two had lost the ability to repress expression of the *Renilla* luciferase, namely the RGG1-RRM and ΔRGG2-ZnF-RGG3 mutants (**Figure 61c**). These translationally inactive mutants lacked a region of the CTD spanning from the second to the third arginine/glycine/glycine (RGG)-rich motifs and including the ZnF (*i.e.*, the RGG2-ZnF-RGG3 region). The absence of the ZnF only had no effect, as the ΔZnF mutant retained full repressive activity. Strikingly, the presence of either RGG2 or RGG3 was sufficient to maintain some translational repression, although not to the level of the full CTD.

To confirm these observations, we deleted this region from the full-length EWSR1 (EWSR1-FL) protein (**Figure 61d**) and found that the resulting mutant (*i.e.*, EWSR1 ΔRGG2-ZnF-RGG3) had completely lost the ability to repress expression of *RLuc* in the tethering assay (**Figures 61e,f**). Based on these analyses, we identified the RGG2-ZnF-RGG3 region as being responsible for the repression of translation by the CTD, most predominantly because of the presence of both RGG2 and RGG3.




b
MS2
CP
CTD
RGG1-RRM
ΔRGG1-RRM
ΔNLS
ΔRGG2
ΔZnF
CP
CTD
ΔRGG3
ΔRGG2-ZnF
ΔZnF-RGG3
ΔRGG2-ZnF-RGG3
RLuc-0
RLuc-4
kDa
FLAG
GAPDH
70
55
40
15
35


c
Normalized RLuc/Fluc activity
2.0
1.5
1.0
0.5
0.0
ns
ns
**
**
**
**
**
*
***
*
MS2-CP
MS2-CTD
MS2-RGG1-RRM
MS2-ΔRGG1-RRM
MS2-ΔNLS
MS2-ΔRGG2
MS2-ΔZnF
MS2-ΔRGG3
MS2-ΔRGG2-ZnF
MS2-ΔZnF-RGG3
MS2-ΔRGG2-ZnF-RGG3
n=3


d
NTD
CTD
EWSR1 ΔRGG2-ZnF-RGG3
SYGQ-rich
RGG1
RRM
NLS

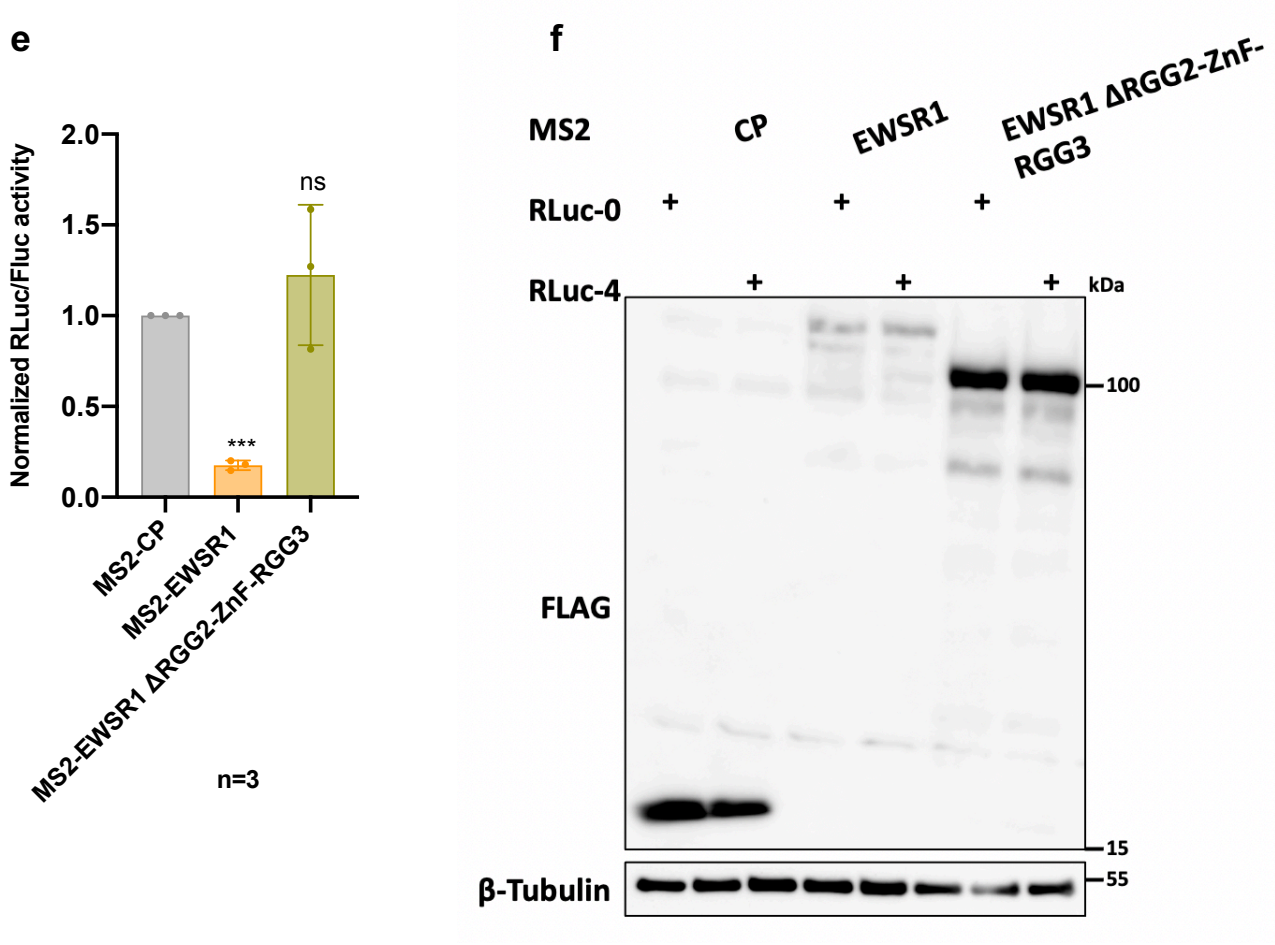


Figure 61: RGG2-ZnF-RGG3 is the region implicated in the repression of translation. (**a**) Schematic domain structure of CTD-deletion mutants. FL: full length. (**b**) Western blot analysis of the levels of MS2-CP, MS2-CTD and MS2-CTD-deletion mutants proteins with anti-FLAG antibody. GAPDH is used as a loading control. Samples are total cell lysates from HeLa cells transfected with MS2-CP or MS2-CP fused CTD or CTD-deletion mutants together with *RLuc-0* or *RLuc-4* reporters constructs. (**c**) Luciferase tethering assays in HeLa cells transfected with MS2-CP or MS2-CP fused CTD or CTD-deletion mutants constructs together with *RLuc-4* or *RLuc-0* reporters. Results are means of normalized *RLuc/FLuc* activities [(*RLuc-4/FLuc*)/(*RLuc-0/FLuc*)] ± s.d. (n = 3 independent experiments). **P* < 0.05, ***P* < 0.01, ****P* <0.001, ns: not significant compared to the MS2-CP condition by one sample *t* test. (**d**) Schematic domain structure of EWSR1-deletion mutant. (**e**) Luciferase tethering assays in HeLa cells transfected with MS2-CP or MS2-CP fused EWSR1 or EWSR1-deletion mutant constructs together with *RLuc-4* or *RLuc-0* reporters. Results are means of normalized *RLuc/FLuc* activities [(*RLuc-4/FLuc*)/(*RLuc-0/FLuc*)] ± s.d. (n = 3 independent experiments). ****P* <0.001, ns: not significant compared to the MS2-CP conditions by one sample *t* test. (**f**) Western blot analysis of the levels of MS2-CP, MS2-EWSR1 and MS2-EWSR1-deletion mutant with anti-FLAG antibody. β-Tubulin is used as a loading control. Samples are total cell lysates from HeLa cells transfected with MS2-CP or MS2-CP fused EWSR1 or EWSR1-deletion mutant proteins together with *RLuc-4* or *RLuc-0* reporters constructs.

3.5 EWSR1 associates with the 40S ribosomal subunit

Interestingly, database curation of the BioGRID[4.4] repository revealed that the EWSR1 interactome comprises many ribosome related factors, including notably both core ribosomal proteins (RPLs like RPL12 and RPL29 and RPSs like RPS10 and RPS18) and ribosome-associated proteins such as eIF4H and eIF5A. This raised the intriguing possibility that EWSR1 might associate with ribosomes to regulate translation. To test this hypothesis, we first examined whether EWSR1 might be present in the cytoplasm, as suggested by previous studies[9]. To this aim, we prepared nuclear and cytoplasmic fractions from HeLa cells and examined the presence of EWSR1 by western blot. As illustrated in **Figure 62**, EWSR1 was observed both in the nucleus and in the cytoplasm. As a control for specificity, the EWSR1 signal was reduced upon transfection with *siEWSR1#1*. Next, we tested whether cytoplasmic EWSR1 could associate with ribosomes, as suggested previously[113]. We conducted subcellular fractionation of HeLa cells following a previoulsy described protocol[609] leading to S30 (cytoplasmic fraction), S100 (cytoplasmic fraction free of ribosomes), R (ribosomes with associated factors fraction) and RSW (ribosomes without or with very little associated factors fraction) fractions (**Figure 63a**). As expected, core ribosomal proteins such as 40S (RPS6, RPS23 and RACK1) and 60S (RPL26, RPL4) components were specifically detected in the R and RSW fractions, but not in the S100 fraction (**Figure 63b**). The initiation factor eIF2α was enriched in the R fraction but barely visible in the RSW fraction, showing that associated factors, such as initiation factors were efficiently detached from the ribosomes by the high salt wash between the R and RSW steps, as expected. Cytoplasmic EWSR1 was detected in the S100 fraction indicating that it can be found outside of ribosomes. However, a large portion of EWSR1 was also detected in the R fraction, demonstrating that it is mostly associated with ribosomes when in the cytoplasm. Because very little EWSR1 could be found in the RSW fraction, we concluded that although EWSR1 sediments with ribosomal fractions, it is not an integral component of the ribosomes (**Figure 63b**).

Next, we studied the pattern of sedimentation of EWSR1 extracted from HeLa cells on linear sucrose density gradients, using polysome profiling. This approach generated a polysome profile including fractions of mRNAs outside of ribosomes (also known as “free mRNA pool”; fractions 1 to 6 in **Figure 63c**), isolated 40S (fractions 7 to 11 in **Figure 63c**), and 60S/80S (fractions 12 to 16 in **Figure 63c**) and polysomes (fractions 17 to 26 in **Figure 63c**). Gradient fractions were analyzed by western blotting. To detect the distribution of the 40S and 60S subunits throughout the gradient, we used RPS6 and RPL26, respectively (**Figure 63c**). EWSR1 was highly present in fractions corresponding to the non-translating (“free mRNA pool”), and although it was also visible in the fractions corresponding to isolated 40S, it was never observed in the 60S/80S or polysomes fractions (**Figure 63c**).

These results prompted us to assess the association of EWSR1 with cytoplasmic mRNAs. To this aim, we conducted an oligo(dT) purification approach. We irradiated HeLa cells with UV to

covalently cross-link direct RNA-protein interactions. We then pulled down poly(A)+ mRNPs, from both total and cytoplasmic extracts using oligo(dT)-tagged beads. RBPMS, a well described RBP[610], and GAPDH and Histone H3, both of which have no known mRNP-related function were used as positive and negative controls, respectively. As expected, GAPDH was observed in total and cytoplasmic fractions, whereas H3 was only detected in total cell lysates. EWSR1 was specifically detected in the oligo(dT) pulldowns from both the total and cytoplasmic fractions (**Figure 64**). The presence of EWSR1 was strongly diminished when pulldowns reactions were treated with RNase A. These analyses show that EWSR1 can associate to cytoplasmic polyadenylated mRNAs. Together with the results of our polysome profiling and in line with the results of a previous study[611] reporting that EWSR1 was among the RBPs either specifically or highly enriched in small, < 40 S mRNP complexes, we concluded that cytoplamsic EWSR1 is predominantly associated with non-translating mRNAs.

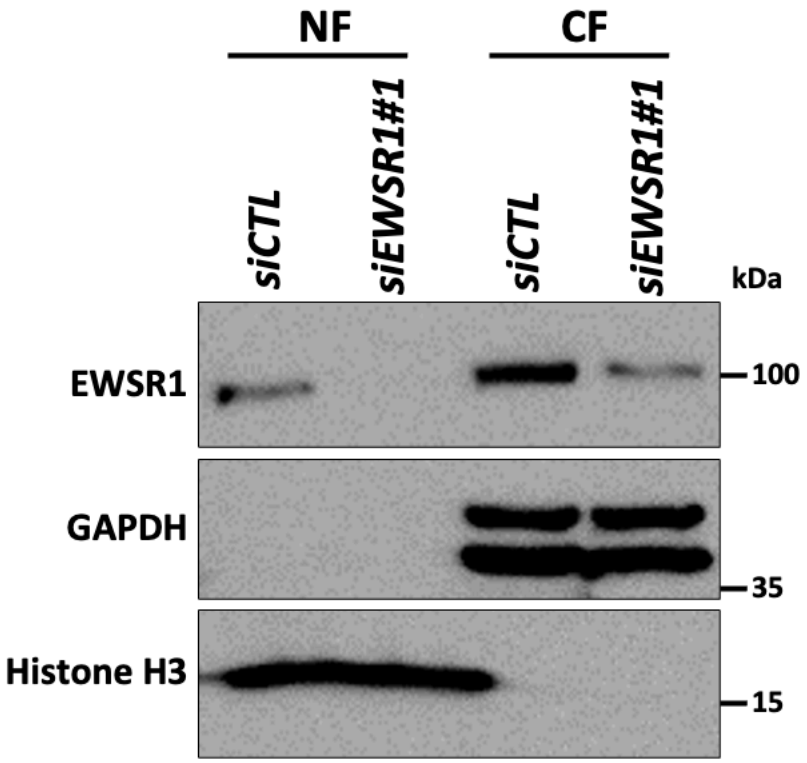


Figure 62: EWSR1 is localized in the nucleus and the cytoplasm. Western blotting of nuclear (NF) and cytoplasmic (CF) fractions from HeLa cells transfected with *siCTL* and *siEWSR1#1*. GAPDH (cytoplasmic control) and Histone H3 (nuclear control).

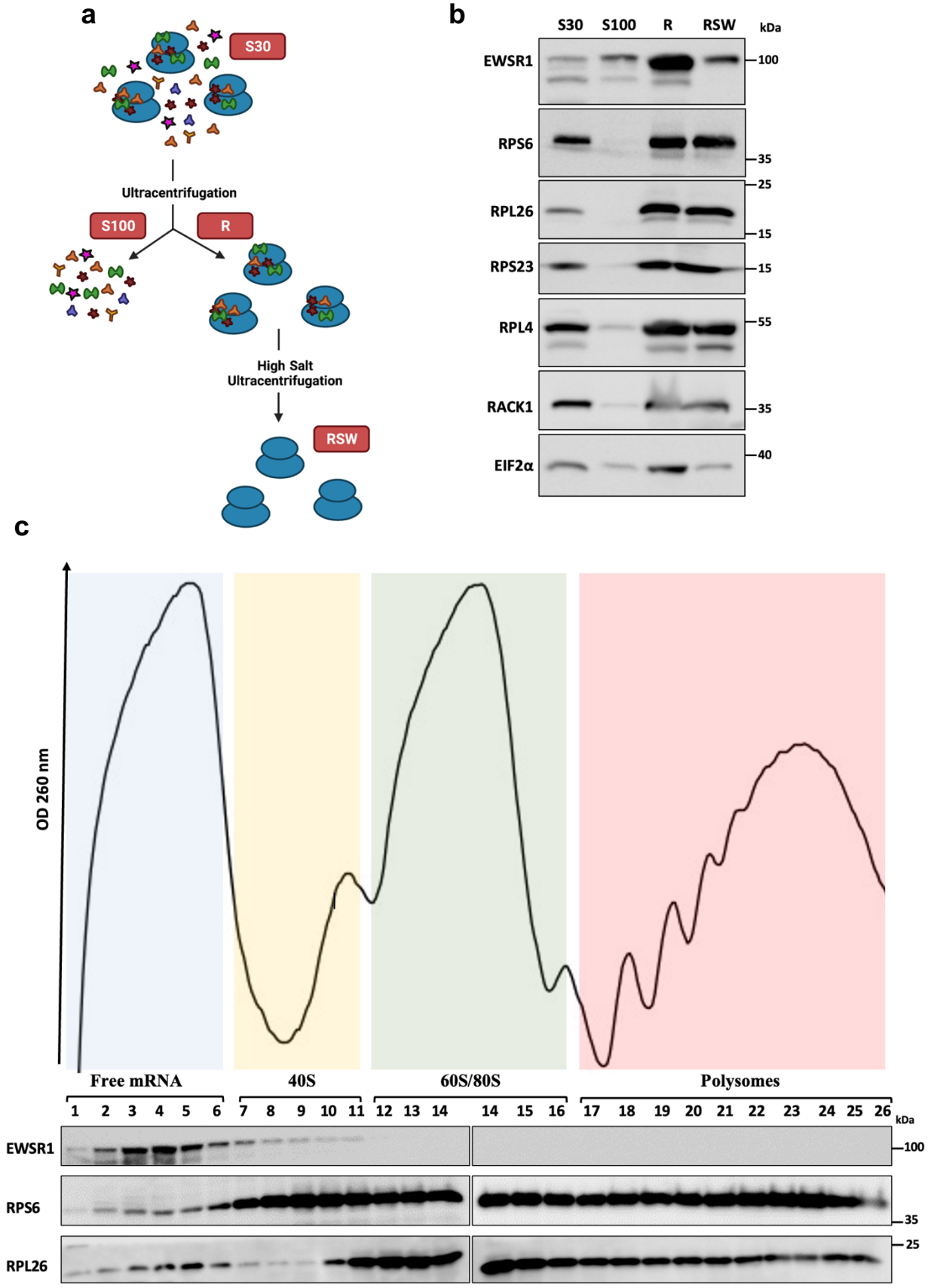

a
S30
Ultracentrifugation
S100
R
High Salt
Ultracentrifugation
RSW
b
S30
S100
R
RSW
kDa
EWSR1
RPS6
RPL26
RPS23
RPL4
RACK1
EIF2α
100
35
25
15
15
55
35
40
c
OD 260 nm
Free mRNA
40S
60S/80S
Polysomes
1 2 3 4 5 6 7 8 9 10 11 12 13 14 14 15 16 17 18 19 20 21 22 23 24 25 26
kDa
EWSR1
RPS6
RPL26
100
35
25

Figure 63: EWSR1 is present in ribosomal fractions. (**a**) Illustration of the subcellular fractionation protocol. Created with BioRender.com. S30: cytoplasmic fraction, S100: cytoplasmic fraction free of ribosomes, R: ribosomes with associated factors fraction, RSW: ribosomes without or with very little associated factors fraction. (**b**) Western blotting of S30, S100, R and RSW fractions from HEK293 cells. The presence of the indicated proteins was analyzed in each fractions using specific antibodies. (**c**) Absorbance profiles at 260 nm (above) and western blot analysis (below) of sucrose gradient sedimentation of HeLa cell cytoplasmic extracts. Fractions 1 to 6 corresponds to "free mRNA pool" not associated to ribosomes, fractions 7 to 11 correspond to isolated 40S particles, fractions 12 to 16 contain free 60S and 80S and fractions 17 to 26 correspond to polysomes. The distributions of EWSR1, RPS6 and RPL26 throughout the different fractions of the sucrose gradient were analyzed by western blotting.

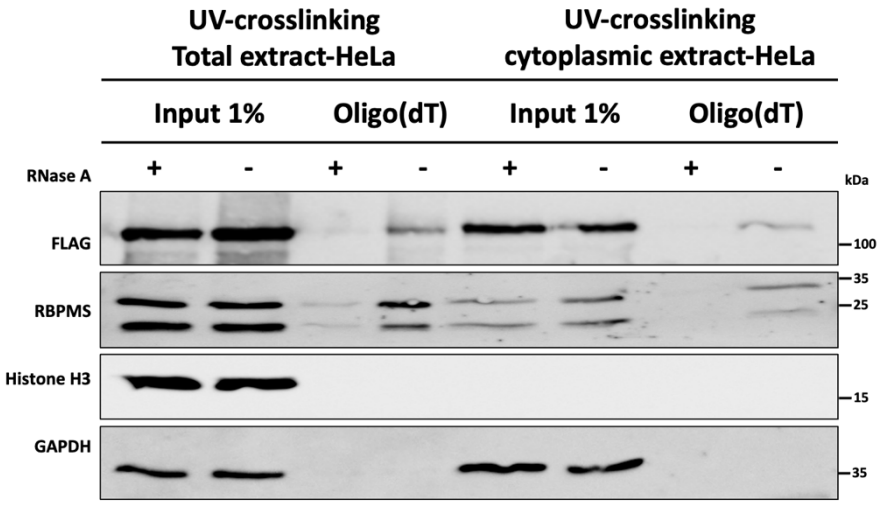


Figure 64: EWSR1 binds to total and cytoplasmic mRNA. Western blotting of oligo(dT)-RNA bound proteins with the indicated antibodies. Samples are total or cytoplasmic HeLa cells extracts transfected with FLAG-EWSR1 and cross-linked with UV. Samples are nontreated (-) or RNase A treated (+).

Our observations that EWSR1 i) is found in the R fraction together with ribosomal components, ii) co-sediments with 40S on polysome gradient and iii) associates with non-translated cytoplasmic mRNA suggest that EWSR1 might inhibit translation by interacting with the small ribosomal subunit. To test this hypothesis, we first performed co-immunoprecipitations between FLAG-tagged EWSR1 or its isolated NTD or CTD regions and endogenous ribosomal proteins from HeLa cells extracts. In these conditions, we found that both EWSR1 and its CTD associated with components of the 40S and 60S particles (**Figure 65**). In contrast, the NTD region of EWSR1 showed no association with the tested ribosomal proteins, indicating that the association between ribosomal components and EWSR1 is mediated by its CTD.

To discriminate between indirect interactions mediated by mRNAs and RNA-independent interactions, we performed co-immunoprecipitations in the presence of RNase A, which is known to preserve the integrity of mammalian 80S particles[612] (**Figure 66a**). EWSR1 and its CTD readily co-immunoprecipitated with both 40S (RPS6 and RPS23) and 60S (RPL26 and RPL22) ribosomal proteins in the absence of RNase A pre-treatment (**Figure 66b**). However, the levels

of RPL22 and RPL26 co-immunoprecipitating with EWSR1 or its CTD were significantly reduced under RNase A treatment, while associations with RPS6 and RPS23 remained largely unaffected. Similar experiments were repeated using RNase T1 which digests polyribosomes into monosomes, without affecting the integrity of the 80S ribosome[612]. Again, we found an RNase-insensitive interaction between EWSR1 or its CTD, with the 40S ribosomal protein, RPS6, but not with 60S protein, RPL22 (**Figure 66c**). Finally, association between EWSR1 and the 40S particle was confirmed at the endogenous level (**Figure 66d**). Altogether, these results thus suggest that EWSR1 associates with the 40S ribosomal particle, in an RNA-independent manner.

Finally, to further establish the association between EWSR1 and the 40S particle, we performed GST pulldown assays with a purified CTD fragment (GST-CTD) produced in *E. coli* as a GST fusion protein (GST-CTD) (**Figure 66e**). GST-CTD was incubated with 40S or 60S/80S fractions purified from polysome gradients. As shown by western blot analysis, the 60S/80S fractions contained both 40S (RPS6, RPS5) and 60S ribosomal proteins (RPL26, RPL4), but very little to no initiation factors, such as the 43S (eIF3B, eIF3η, eIF4E, eIF5). In contrast, the 40S fraction contained 40S ribosomal proteins, various initiation factors, but no detectable 60S ribosomal proteins. Analysis of proteins pulled down from the 40S and 60S/80S fractions by GST-CTD revealed that it associated specifically with 40S ribosomal proteins, such as RPS6 and RPS5, whereas the GST control did not. In contrast, no association was observed between the GST-CTD and 60S components, such as RPL26 and RPL4 (**Figure 66e**). Because the 60S/80S fraction most likely contains assembled 80S particles, these experiments strongly suggest that EWSR1 can bind the 40S particle in the absence of any associated 60S subunit. Initiation factors present in the 40S fractions, such as the cap-binding eIF4E, or the 43S preinitiation factor eIF3 were not pulled down by GST-CTD, suggesting that EWSR1 does not associate with the 43S pre-initiation complex (PIC).

The above observations prompted us to more finely map the 40S-interacting region of EWSR1 within the CTD. To this aim, we performed co-immunoprecipitations using the series of CTD deletion mutants that we previously generated (**see Figure 61a**). We tested the ability of these mutants to associate with endogenous 40S ribosomal proteins, such as RPS6 and RPS23. This analysis identified two mutants that were unable to associate with the 40S particle, namely the RGG1-RRM (lacking the RGG2-ZnF-RGG3-NLS region) and the ΔRGG2-ZnF-RGG3 mutants (**Figure 67a**). The ΔRGG2-ZnF and ΔZnF-RGG3 mutants were severely impaired in their ability to associate with RPS6 and RPS23, while the ΔRGG3 and ΔRGG2 mutants were not significantly affected. These results thus identify the RGG2-ZnF-RGG3 region of EWSR1 as the minimal region responsible for the interaction with the 40S particle. Indeed, an EWSR1 mutant lacking the RGG2-ZnF-RGG3 region did not associate with RPS6 (**Figure 67b**). Strikingly, the RGG2-ZnF-RGG3 region is the same region that we had identified as being involved in EWSR1-mediated translational repression (**see Figure 61c**). The observation that the 40S-interacting

region overlaps with the region functionally involved in translational repression supports the idea that the repression of translation by EWSR1 is correlated to its ability to associate with the 40S particle.

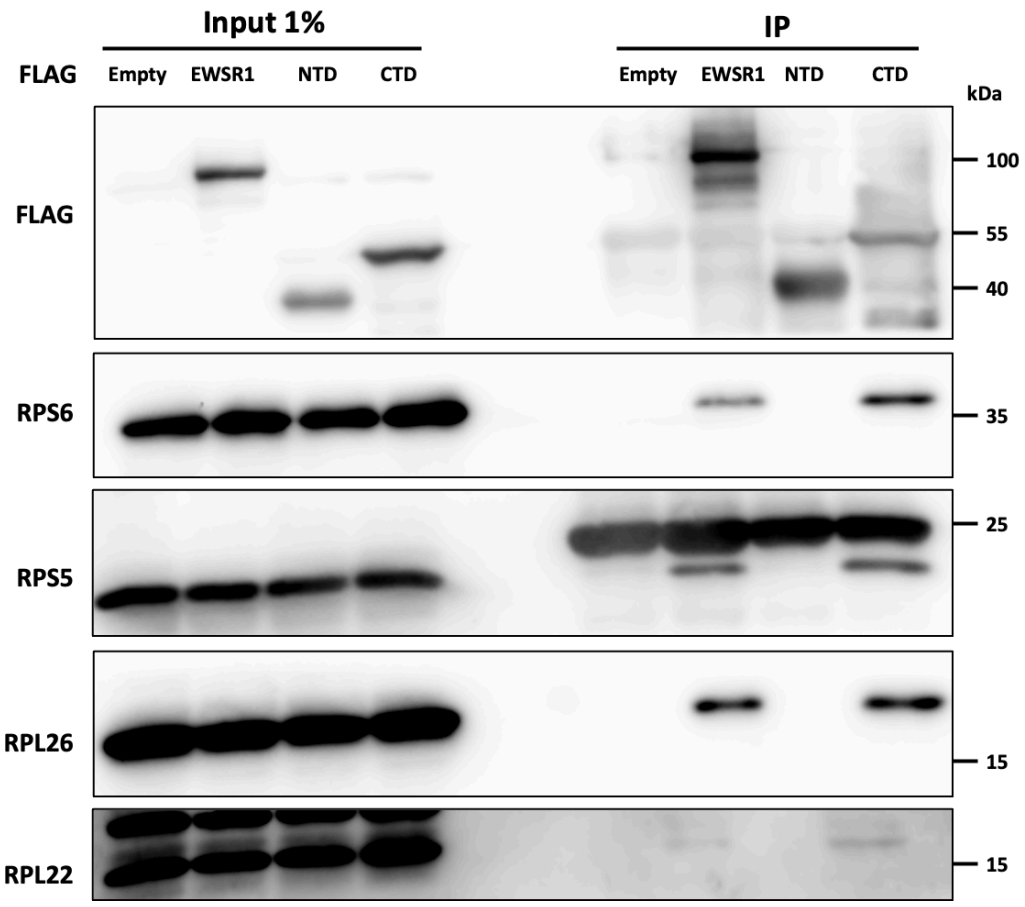
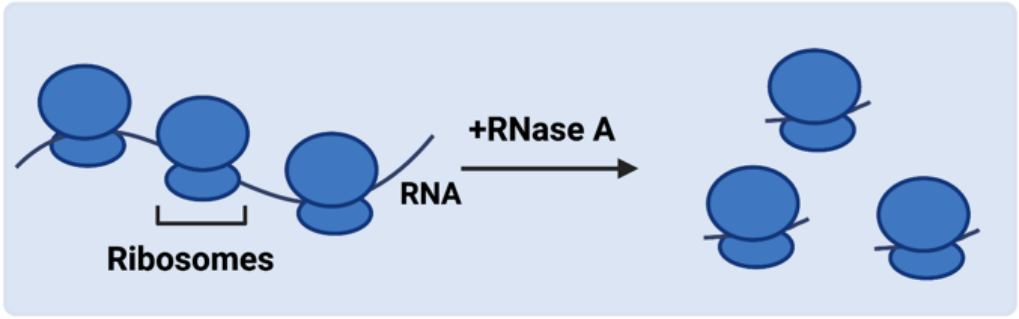


Figure 65: EWSR1 binds to the 40S subunit *via* its CTD. Immunoprecipitation (IP) of FLAG-tagged EWSR1, NTD and CTD and anti-FLAG, anti-RPS6, anti-RPS5, anti-RPL26 and anti-RPL22 western blotting. Samples are lysates from HeLa cells transfected with FLAG-EWSR1, FLAG-NTD, FLAG-CTD or control FLAG Empty-vector.

a



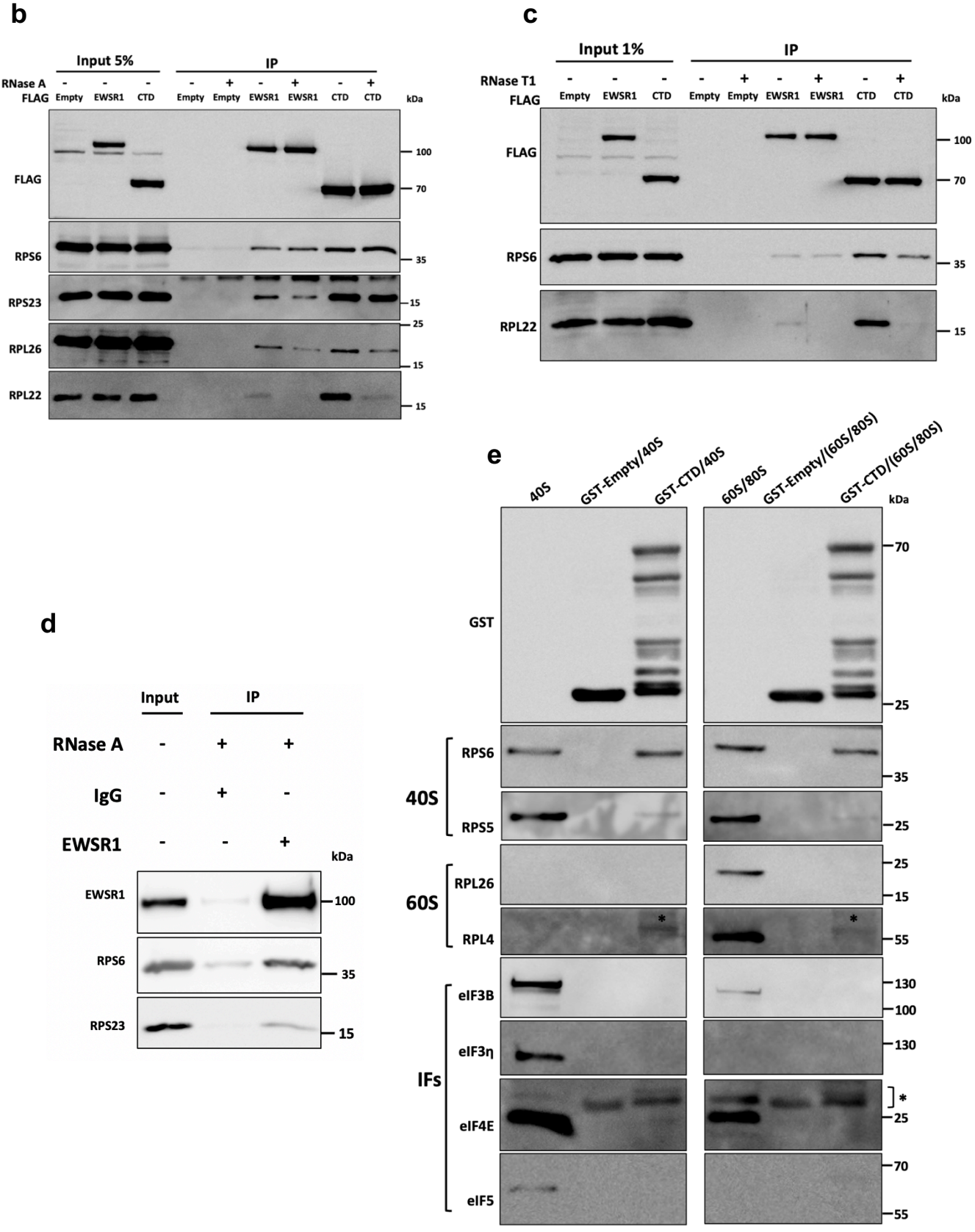

b
Input 5%
IP
RNase A
FLAG
Empty
EWSR1
CTD
kDa
FLAG
RPS6
RPS23
RPL26
RPL22
c
Input 1%
IP
RNase T1
FLAG
Empty
EWSR1
CTD
kDa
FLAG
RPS6
RPL22
d
Input
IP
RNase A
IgG
EWSR1
kDa
EWSR1
RPS6
RPS23
e
40S
GST-Empty/40S
GST-CTD/40S
60S/80S
GST-Empty/(60S/80S)
GST-CTD/(60S/80S)
kDa
GST
40S
RPS6
RPS5
60S
RPL26
RPL4
IFs
eIF3B
eIF3η
eIF4E
eIF5

Figure 66: EWSR1 binds to the 40S through RNase-insensitive interaction. (**a**) Schematic representation of RNase A treatment effect on polysomes. Created with BioRender.com. (**b**) Immunoprecipitation (IP) of FLAG-tagged EWSR1 and CTD and anti-FLAG, anti-RPS6, anti-RPS23, anti-RPL26, and anti-RPL22 western blotting. Samples are nontreated (-) or RNase A treated (+) lysates from HeLa cells transfected with FLAG-EWSR1, FLAG-CTD or control FLAG Empty-vector. (**c**) Immunoprecipitation (IP) of FLAG-tagged EWSR1 and CTD and anti-FLAG, anti-RPS6 and anti-RPL22 western blotting. Samples are nontreated (-) or RNase T1 treated (+) lysates from HeLa cells transfected with FLAG-EWSR1, FLAG-CTD or control FLAG Empty-vector. (**d**) Immunoprecipitation (IP) of endogenous EWSR1 from HEK293 cell lysates followed by western blotting for the indicated proteins. (**e**) Western blotting of GST pulldowns performed with GST-CTD or GST alone on 40S or 60S/80S fractions. GST-tagged proteins and the endogenous levels of 40S, 60S markers and initiation factors (IFs) are shown using specific antibodies. Asterisks indicate nonspecific bands.

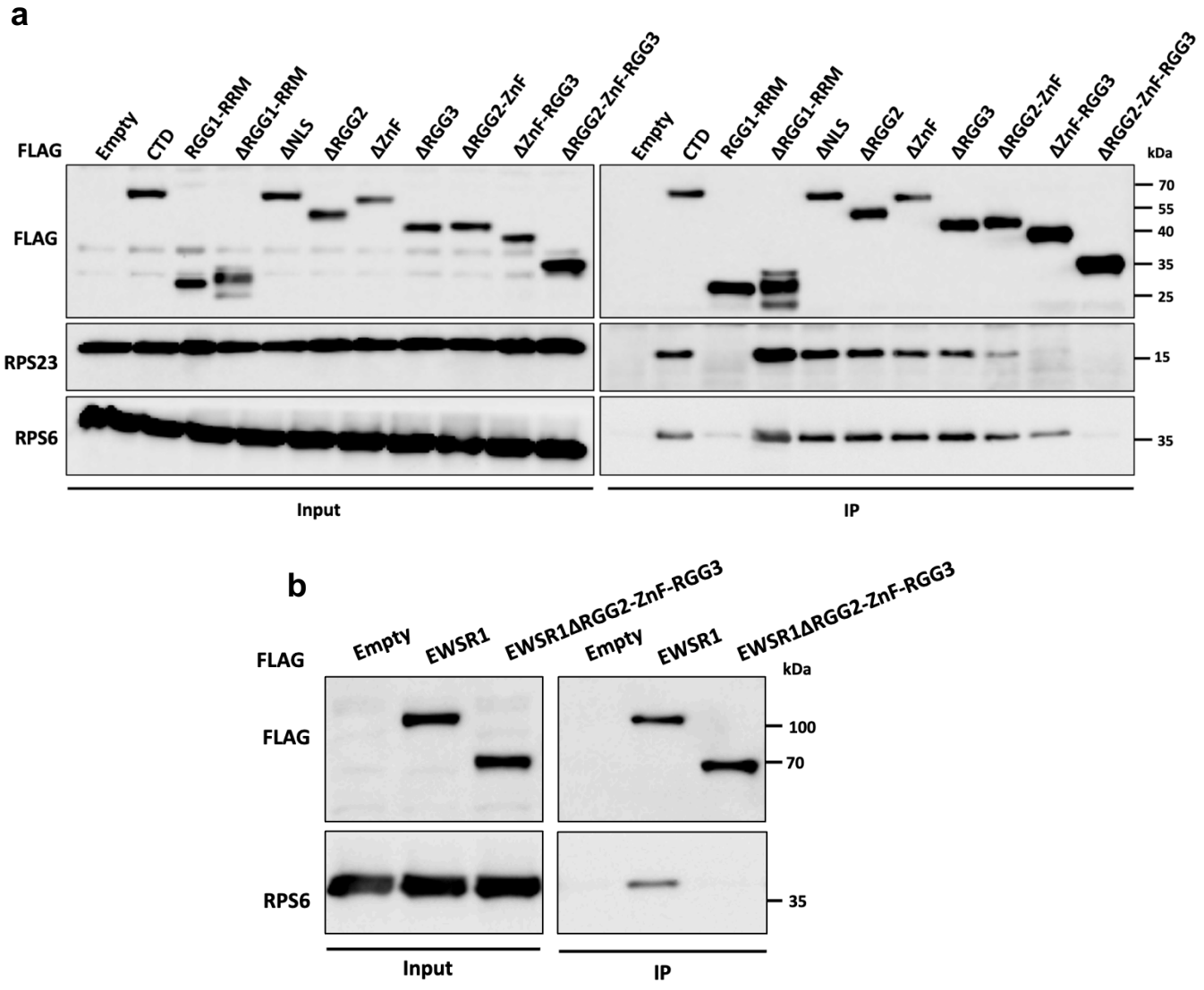


Figure 67: RGG2-ZnF-RGG3 is the CTD region implicated in the interaction with the 40S. (**a**) Immunoprecipitation (IP) of FLAG-tagged CTD and its deletion mutants and anti-FLAG, anti-RPS23 and anti-RPS6 western blotting. Samples are lysates from HeLa cells transfected with FLAG-CTD, FLAG-CTD-deletion mutants or control FLAG Empty-vector. (**b**) Immunoprecipitations (IP) of FLAG-tagged EWSR1 and its deletion mutant and anti-FLAG and anti-RPS6 western blotting. Samples are lysates from HeLa cells transfected with FLAG-EWSR1, FLAG-EWSR1-deletion mutant or control FLAG-Empty vector.

3.6 EWSR1 represses translation downstream of ATG scanning

As the rate-limiting step of the translation process, initiation is also often targeted by translation regulators[613]. We thus considered the possibility that EWSR1 might inhibit translation by interfering with initiation. First, we tested whether EWSR1 might interact and/or tamper with cap recognition using an m^7GTP-Agarose binding assay[351] (**Figure 68a**). While we were able to efficiently pull down eIF4E, eIF4G and subunits of the initiation factor eIF3, we found no evidence that endogenous EWSR1 associates with m^7GTP (**Figure 68b**). Consistent with our GST-CTD pulldown assay, in which the CTD showed no binding to eIF4E (**Figure 66e**), these observations support the model whereby EWSR1 does not interact with translation initiation factors to inhibit translation. Moreover, knockdown of EWSR1 had no effect on the association of the eIF4F components, including the cap-binding subunit eIF4E, the scaffolding subunit eIF4G and the associated initiation factors eIF3η and eIF3B to the immobilized m^7GTP, indicating that cap recognition by the 43S PIC does not require EWSR1 (**Figure 68b**).

Internal ribosome entry sites (IRES) from viral or rare cellular mRNA drive a non-canonical mode of translation initiation termed "cap-independent translation" that does not rely on the m^7GTP at the 5'end of mRNA, but might depend on few of the canonical initiation factors[614]. The Poliovirus (PV) IRES requires eIF2, eIF3, eIF4A, eIF4G, eIF4B, eIF1A, and a single IRES trans-acting factor (ITAF), the poly(C) binding protein 2 (PCBP2)[615]. The Cricket paralysis virus (CrPv) IRES does not require any translation initiation factors[616]. To more firmly establish that EWSR1 does not repress cap-dependent translation by interfering with the initiation step, we tested its effect on IRES-driven reporters. We cloned the PV or CrPv IRES in the 5'UTR of our bidirectional *RLuc* reporter with 8 MS2 binding sites and tested the effect of tethering EWSR1 or its CTD on IRES-driven *RLuc* translation (**Figure 69a**). Tethering of MS2-EWSR1 or MS2-CTD to the *RLuc* mRNA significantly reduced luciferase expression from both the CrPv and PV IRES reporters, similarly to our observations with cap-dependent reporters (**Figures 69b-e**).

The ability of EWSR1 to repress both cap-dependent and -independent translation suggests an EWSR1-mediated translation repression mechanism, acting downstream of the initiation step. Once assembled, the 43S PIC moves in the 3' direction, scanning for the proximal initiation codon. To assess whether EWSR1 might inhibit scanning by the 43S, we used a previously described *RLuc*-6xMS2 reporter whose translation occurs *via* a scanning-independent mechanism, using a TISU motif located in a 9 nt 5'UTR (TISU-*RLuc*-6xMS2-polyA)[617] (**Figure 70a**)**.** Upon tethering to the reporter transcript, we found that EWSR1 and its CTD significantly inhibited expression of *RLuc* from this reporter by 20% and 83%, respectively, suggesting that inhibition of translation by EWSR1 is independent of 5'UTR scanning by the 43S PIC (**Figures 70b,c**).

In parallel, we looked whether depletion of EWSR1 might interfere with translation by slowing down 5'UTR scanning or increase residency of the 43S PIC at the initiation codon. To this aim, we determined read distribution in 5'UTR and P-site occupancy at start codons in all detected mRNAs, using our RIBO-seq data and the same approach that was conducted by *Sun et al.*[618]. These analyses revealed no difference between *siCTL* and *siEWSR1#1* conditions for the relative P-site occupancy at start codon or read densities in 5'UTR between *siCTL* and *siEWSR1#1* (**Figure 71**).

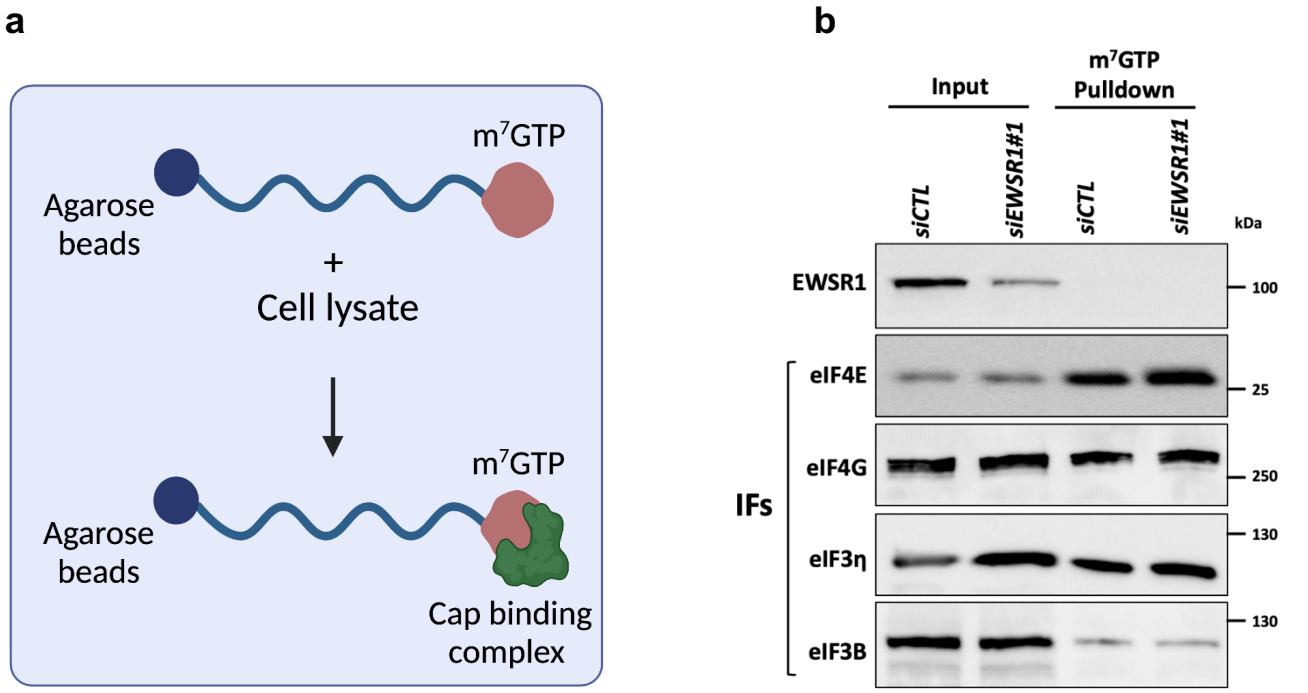
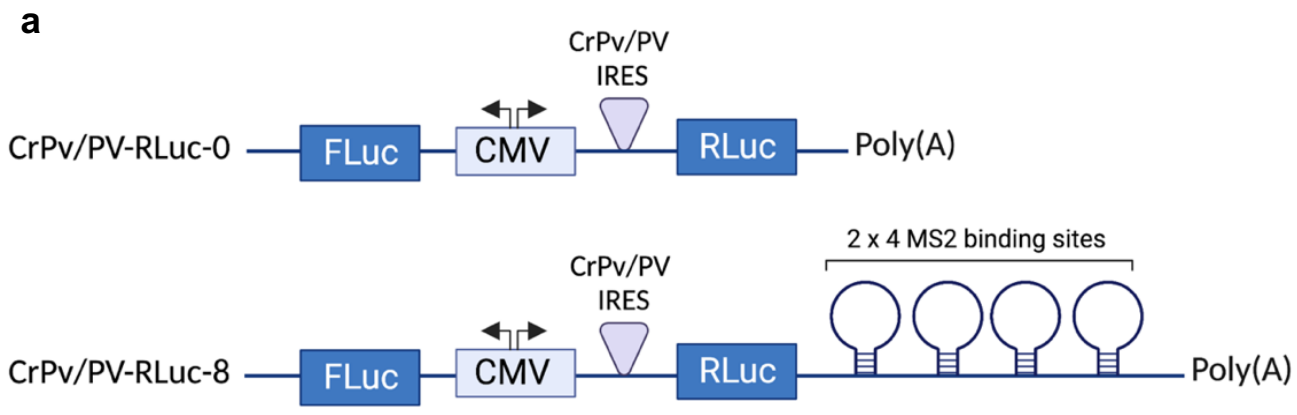
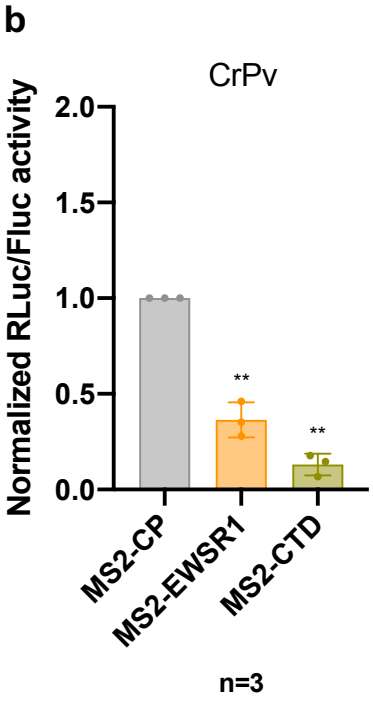
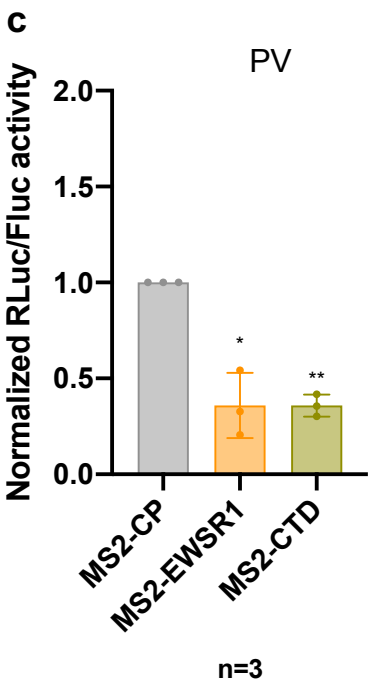
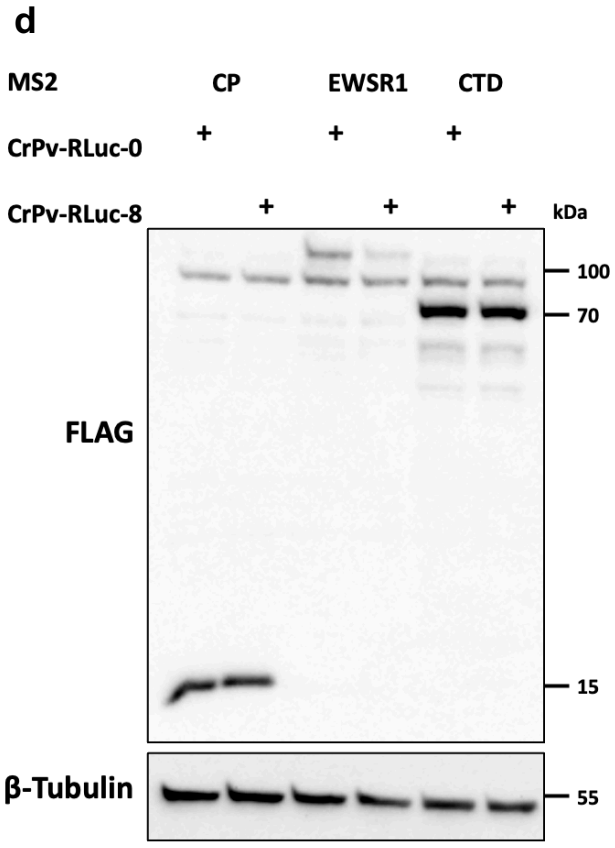


Figure 68: EWSR1 does not interact with the cap or initiation factors. (**a**) Schematic representation of m^{7}GTP pulldown. Created with BioRender.com. Cap binding complex from cell lysate associates to the cap bound to agarose beads. (**b**) Western blotting of EWSR1 and the indicated initiation factors. Samples are m^{7}GTP pulldowns from HeLa cells transfected with *siEWSR1#1* or *siCTL*. IFs: initiation factors. m^{7}GTP: 7-methylguanylate structure.




b
CrPv
Normalized RLuc/Fluc activity
2.0
1.5
1.0
0.5
0.0
**
**
MS2-CP
MS2-EWSR1
MS2-CTD
n=3


c
PV
Normalized RLuc/Fluc activity
2.0
1.5
1.0
0.5
0.0
*
**
MS2-CP
MS2-EWSR1
MS2-CTD
n=3


d
MS2
CP
EWSR1
CTD
CrPv-RLuc-0
+
+
+
CrPv-RLuc-8
+
+
+
kDa
100
70
FLAG
15
β-Tubulin
55

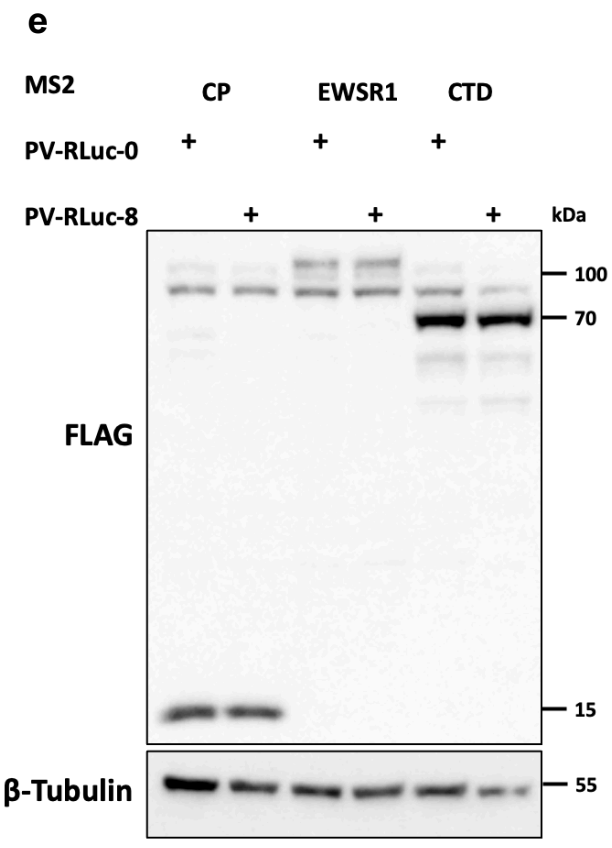

e
MS2
CP
EWSR1
CTD
PV-RLuc-0
+
+
+
PV-RLuc-8
+
+
+
kDa
100
70
FLAG
15
β-Tubulin
55

Figure 69: EWSR1 represses cap-independent translation. (**a**) Schematic representation of CrPv/PV IRES bidirectional luciferase reporters (with 8 MS2 binding sites). Created with BioRender.com. (**b, c**) Luciferase tethering assays in HeLa cells transfected with MS2-CP or MS2-CP fused EWSR1 or CTD constructs together with CrPv-*RLuc*-8 or CrPv-*RLuc*-0 (**b**) or PV-*RLuc*-8 or PV-*RLuc*-0 (**c**) reporters. Results are means of normalized *RLuc/FLuc* activities [(*RLuc-8/FLuc*)/(*RLuc-8/FLuc*)] ± s.d. (n = 3 independent experiments). **P* < 0.05, ***P* < 0.01 compared to the MS2-CP condition by one sample *t* test. (**d, e**) Western blot analysis of the levels of MS2-CP, MS2-EWSR1 and MS2-CTD with anti-FLAG antibody. HSP90 is used as a loading control. Samples are total cell lysates from HeLa cells transfected with MS2-CP or MS2-CP fused EWSR1 or CTD together with CrPv-*RLuc*-8 or CrPv-*RLuc*-0 (**d**) or PV-*RLuc*-8 or PV-*RLuc*-0 reporters constructs(**e**). IRES: Internal ribosome entry sites. PV: Poliovirus. CrPv: Cricket paralysis virus.

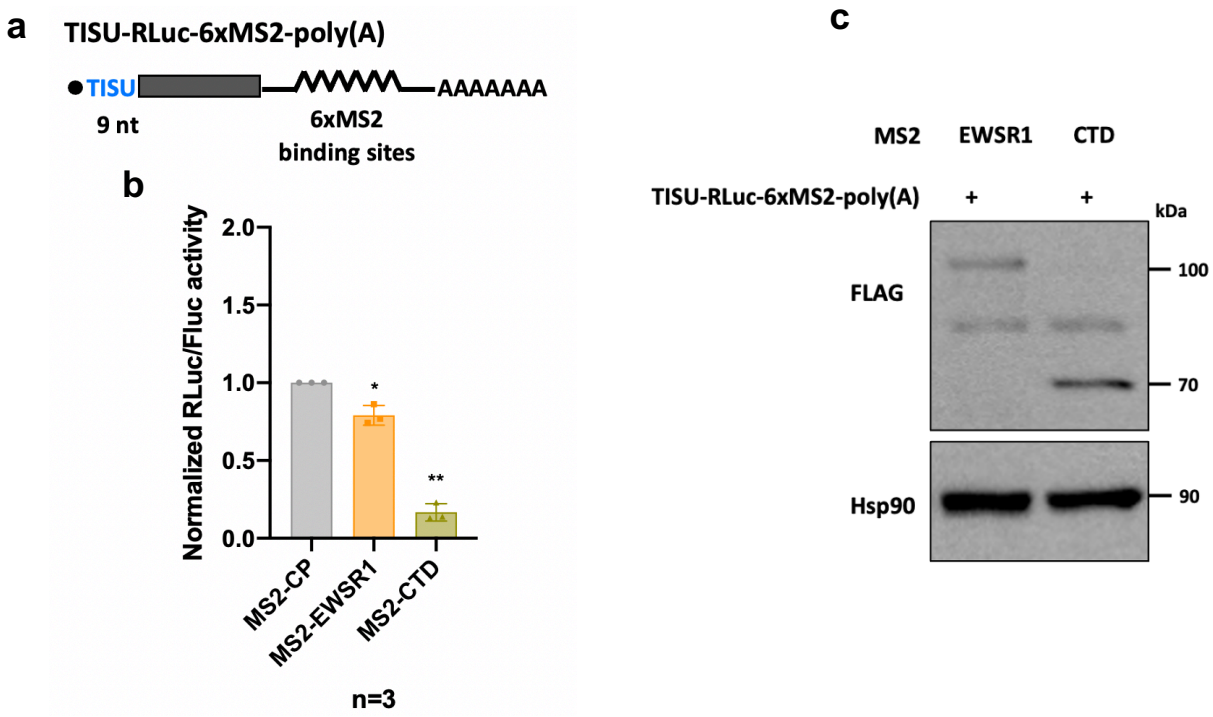


Figure 70: EWSR1 represses scanning-independent translation. (**a**) Schematic representation of the reporter TISU-*RLuc*-6xMS2-poly(A) with 6 MS2 binding sites used in panel (**b**). (**b**) Luciferase tethering assays in HeLa cells transfected with MS2-CP or MS2-CP fused EWSR1 or CTD constructs together with TISU-*RLuc*-6xMS2-poly(A) and pGL3-*Firefly* reporters. Results are means of normalized *RLuc/FLuc* activities ± s.d. (n = 3 independent experiments). **P* < 0.05, ***P* < 0.01 compared to the MS2-CP condition by one sample *t* test. (**c**) Western blot analysis of the levels of MS2-EWSR1 and MS2-CTD with anti-FLAG antibody. HSP90 is used as a loading control. Samples are total cell lysates from HeLa cells transfected with MS2-EWSR1 or MS2-CTD together with TISU-*RLuc*-6xMS2-poly(A) and pGL3-*Firefly* reporters constructs.

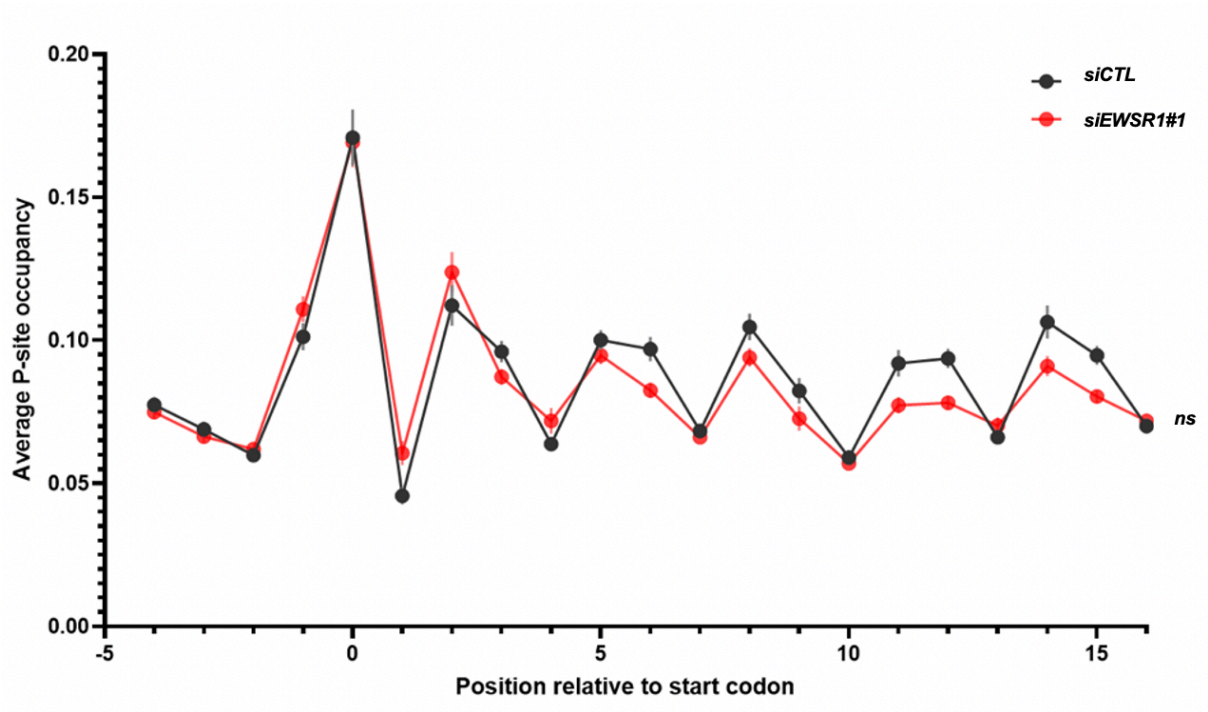


Figure 71: P-site occupancy. Average P-site occupancy at start codon between *siCTL* and *siEWSR1#1* conditions for all genes (n = 11,929). ns = not significant using Wilcoxon-Mann-Whitney test (n = 2 independent experiments).

3.7 The translational function of EWSR1 is linked to lipid homeostasis in cancer cells

To assess the biological relevance of EWSR1 function in translation, we performed Gene Ontology (GO) term enrichment analysis of genes whose TE is either up or downreguated by EWSR1 KD (*i.e.*, differential translation efficiency genes (DTEGs)). The 222 TE down genes were enriched in various RNA processing mechanisms (*i.e.*, RNA splicing, mRNA metabolic and catabolic processes, RNA localization and RNP assembly) and translation (*i.e.*, translational initiation and ribosomal subunit) related processes (**figure 72a**). Interestingly, analysis of the 595 TE up genes revealed significant enrichments in genes related to membrane structure and organization (*e.g.*, extracellular matrix organization, collagen-containing extracellular matrix, basement membrane, integral component of organelle membrane), lipid metabolism (*e.g.*, membrane lipid metabolic process, sphingolipid metabolic process), glycosylation metabolism (*e.g.*, glycoprotein metabolic cellular process, protein glycosylation, glycosaminoglycan metabolic process, protein-O-linked glycosylation, glycosaminoglycan catabolic process) and endoplasmic reticulum (ER) cellular processes (*e.g.*, protein localization to endoplasmic reticulum, endoplasmic reticulum lumen) (**Figure 72b**). Moreover, GO analysis of the 134 “direct EWSR1 translational targets” showed a similar enrichment to that of TE up genes (**Figure 72c**).

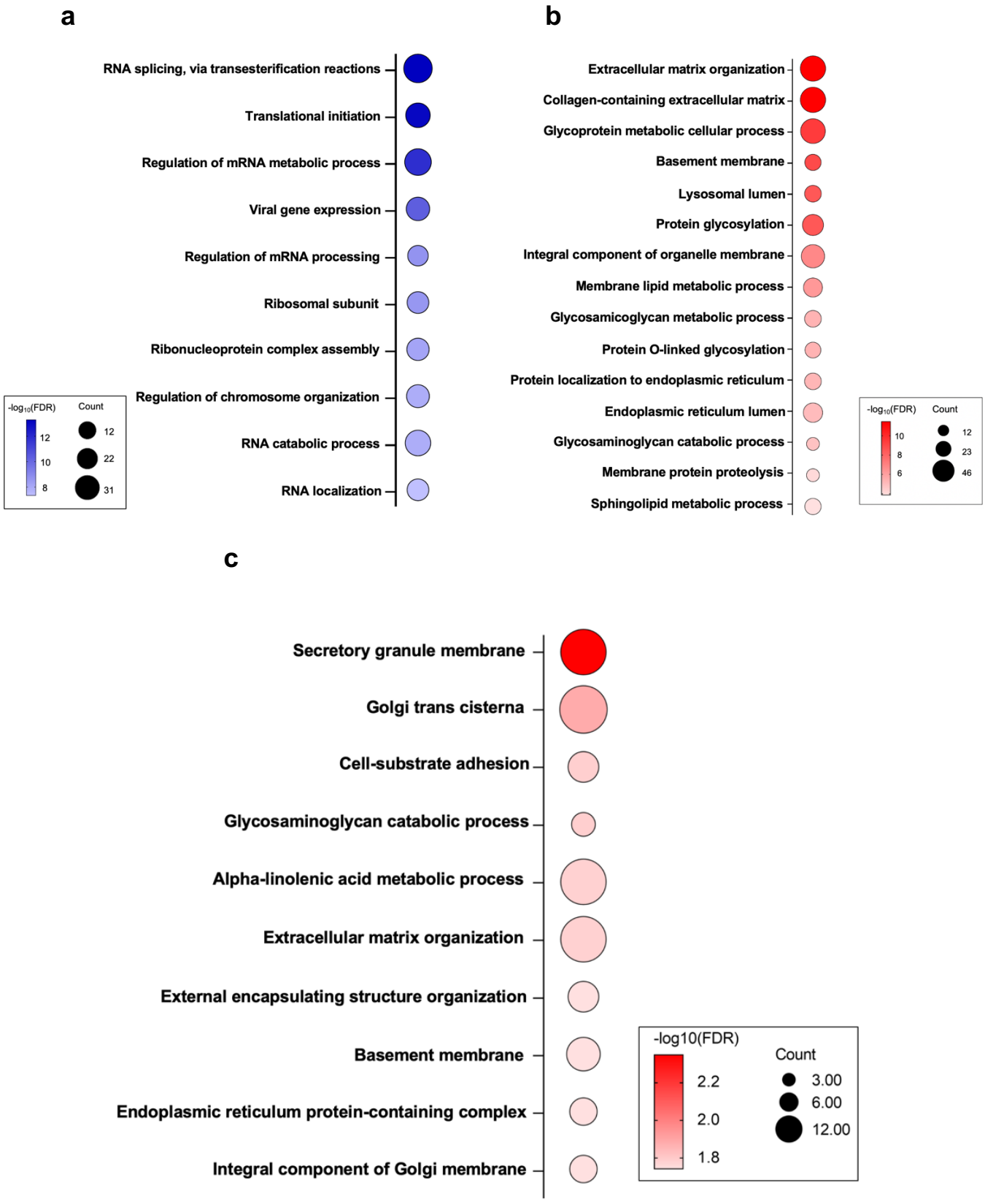

a
RNA splicing, via transesterification reactions
Translational initiation
Regulation of mRNA metabolic process
Viral gene expression
Regulation of mRNA processing
Ribosomal subunit
Ribonucleoprotein complex assembly
Regulation of chromosome organization
RNA catabolic process
RNA localization
-log10(FDR)
12
10
8
Count
12
22
31
b
Extracellular matrix organization
Collagen-containing extracellular matrix
Glycoprotein metabolic cellular process
Basement membrane
Lysosomal lumen
Protein glycosylation
Integral component of organelle membrane
Membrane lipid metabolic process
Glycosamicoglycan metabolic process
Protein O-linked glycosylation
Protein localization to endoplasmic reticulum
Endoplasmic reticulum lumen
Glycosaminoglycan catabolic process
Membrane protein proteolysis
Sphingolipid metabolic process
-log10(FDR)
10
8
6
Count
12
23
46
c
Secretory granule membrane
Golgi trans cisterna
Cell-substrate adhesion
Glycosaminoglycan catabolic process
Alpha-linolenic acid metabolic process
Extracellular matrix organization
External encapsulating structure organization
Basement membrane
Endoplasmic reticulum protein-containing complex
Integral component of Golgi membrane
-log10(FDR)
2.2
2.0
1.8
Count
3.00
6.00
12.00

Figure 72: GO analysis of EWSR1 TE targets. (**a, b,c**) Bubble charts of Gene Ontology (GO) Biological Process terms in TE down (blue colored bubbles) (**a**) or TE up (red colored bubbles) genes (**b**) or genes from the "direct EWSR1 translational targets" dataset (**c**) (FDR < 0.05). Bubble's size reflects the number of DTEGs and color scale corresponds to *P*-values. Enrichment analyses were performed with DAVID and *P*-values were calculated using FDR correction.

Interestingly, many of the translational targets of EWSR1 are linked to lipid metabolism. Therefore, we hypothesized that knockdown of EWSR1 might change the intracellular lipid composition. To test this, we conducted a mass spectrometry-based lipidomic analysis of cells treated with *siEWSR1#1*. We compiled the identified eighteen lipid classes into three major lipid categories: storage, membrane and signaling lipids (**Figure 73**). Storage lipids comprise triacylglycerides and cholesterol esters. Glycerolphospholipids, sphingolipids and steroids are lipid components of cell membranes. Signaling lipids correspond to numerous classes such as LPC (lysophosphatidylcholine), DG (diacylglycerides) and CER (ceramides).

Western blot analysis of the three independent replicates used for the lipidomic experiment confirmed knockdown of EWSR1 (**Figures 74a,b**)**.** At the relative level (when looking at the proportions of lipid species relative to the overall lipid composition), PCA revealed no significant clustering (**Figure 74c**). In contrast, clustering was clearly observed between the two conditions (*siEWSR1#1 vs siCTL*), when data were examined at the quantitative level (absolute amounts of each lipid species) (**Figure 74d**). Quantitative data were expressed as nanomoles of a specific lipid species, normalized to DNA content to account for differences in cell number between the two conditions. Except for CE, LPC, DG, DCER, we observed a significant increase of all lipid classes in the *siEWSR1#1* treated cells (**Figure 75**). Strikingly, membrane lipids such as PC, PS, PE, PI, PG and SM were among the most highly increased lipids in the EWSR1 KD condition.

Triacylglycerides (TGs) are neutral lipids that represent the major form of storage and transport of fatty acids within cells. They are deposited in lipid droplets, which play important roles in several cellular events[619]. Remarkably, we found that triacylglycerides were the most upregulated lipid class upon EWSR1 depletion, pointing to an important function of EWSR1 in the regulation of lipid storage in the cell. This remarkable rise of TGs could be explained by an overall increased lipid production, in EWSR1 KD condition, which can be converted into TGs, and might be deposited subsequently in lipid droplets.

As shown in **Figure 72b**, mRNA coding for proteins linked to the cellular lipid metabolism are significantly enriched in DTEG upon EWSR1 KD. This connection with lipid metabolism is specific to the translational function of EWSR1, as no lipid-related GO was enriched in DEG (**see Figure 49a**). To support the idea that EWSR1 directly controls the translation of transcripts related to lipid metabolism, we tested its ability to bind such mRNA. Among the 134 "direct EWSR1 translational targets" list, we identified 8 transcripts coding for important factors in lipid metabolism: *FADS1, FADS2, ELOVL5, ELOVL6, LPCAT1, LAMP1, SLC39 A1, HSD17B12*.

Using RIP-qPCR, we found all tested lipid-related transcripts to be significantly enriched in the EWSR1 immunoprecipitation (**Figure 76**). As control, *GAPDH* mRNA did not show a significant enrichment.

The ER is the main site of lipid biosynthesis. In addition, changes in lipid homeostasis have an impact on ER shape and function[452,620]. In light of our results, we assessed the possible role of EWSR1 in ER biology. To this aim, we first performed ultrastructural analysis of ER using transmission electron microscopy (TEM) in EWSR1 KD HeLa cells. Interestingly, TEM analysis revealed that ER from EWSR1 KD cells had an altered morphology, appearing as dilated structures with expended lumen (**Figure 77**). In contrast, no morphological difference was detected for the nucleus, Golgi, lysosomes, or mitochondria between control and EWSR1-depleted conditions. The specific dilated mophology of the ER in EWSR1 KD condition indicates that EWSR1 is a key protein for the maintenance of a normal ER shape.

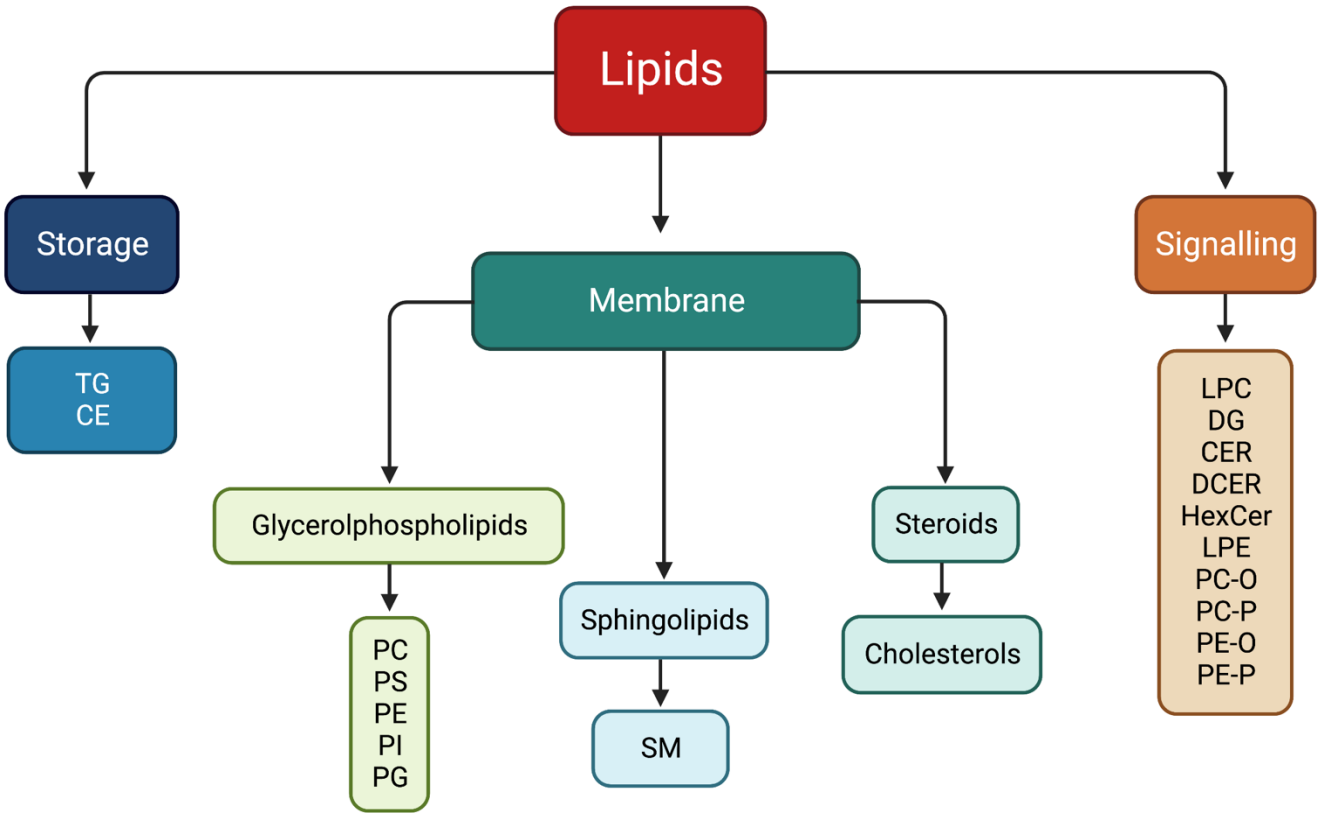


Figure 73: Lipids classification into three main categories: storage (blue), membrane (green) and signaling (orange). Created with BioRender.com. TG: Triacylglyceride, CE: Cholesterol ester, PC: phosphatidylcholine, PS: phosphatidylserine, PE: phosphatidylethanolamine, PI: phosphatidylinositol, PG: phosphatidylglycerol, SM: Sphingomyelin, LPC: lysophosphatidylcholine, DG: diacylglycerides, CER: ceramides, DCER: dihydroceramides, HexCer: hexosylceramides, LPE: lysophosphatidylethanolamine, PC-O: 1-alkyl,2-acylphosphatidylcholine, PC-P: -alkenyl,2-acylphosphatidylcholine, PE-O: 1-alkyl,2-acylphosphatidylethanolamine, PE-P: 1-alkenyl,2-acylphosphatidylethanolamine.

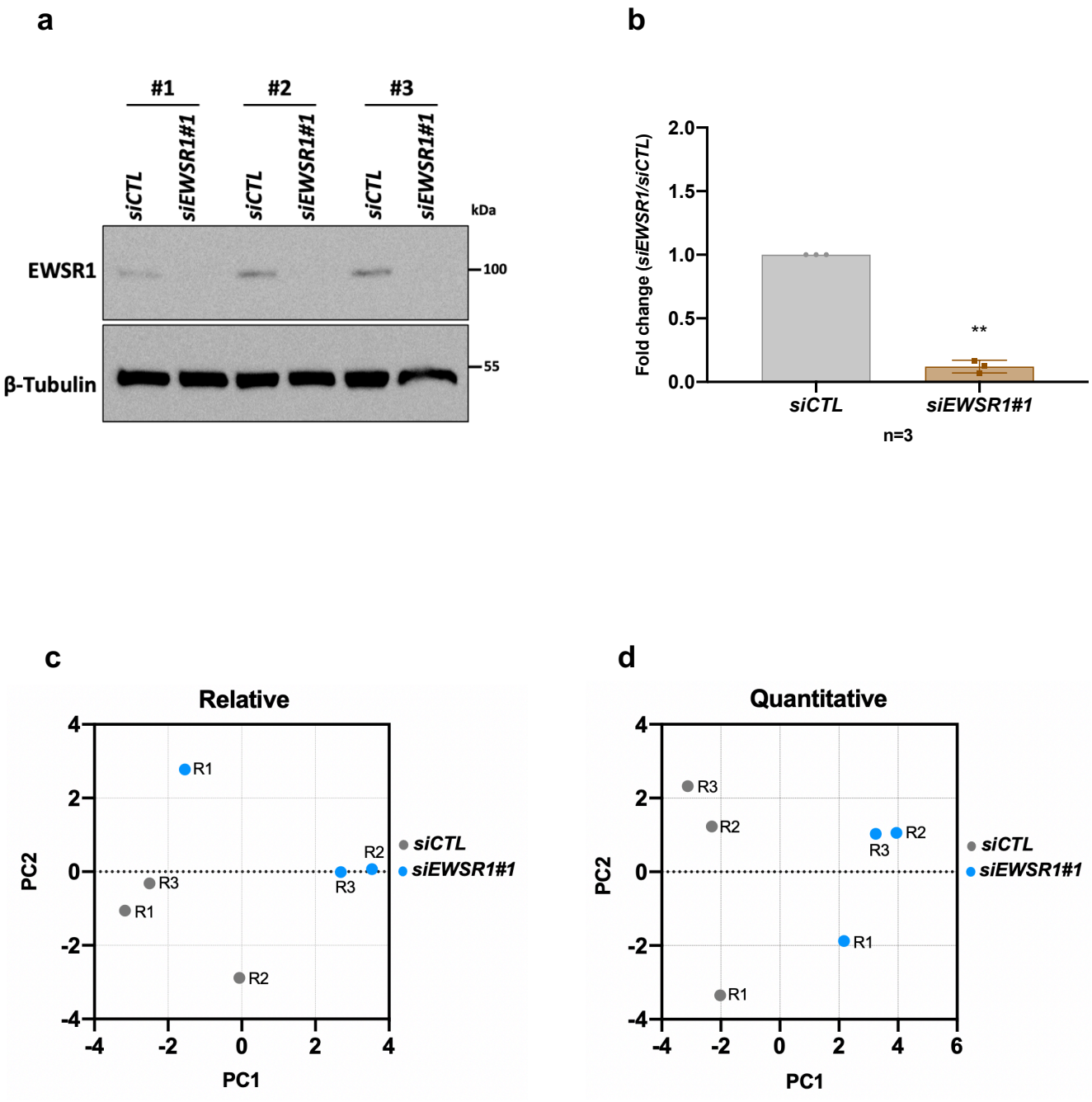


Figure 74: Quality control of the lipidomic analysis. (**a**) Western blotting of EWSR1 and β-Tubulin (loading control). Samples are total cell lysates from HeLa cells transfected with *siEWSR1#1* or *siCTL*. (**b**) The level of EWSR1 was quantified by band densitometry analysis using ImageJ. Results are shown as means ± sd (n = 3 independent experiments) after normalization to β-Tubulin. $^{**}P < 0.01$ compared to the *siCTL* condition by one sample *t* test. (**c, d**) PCA analysis of relative (**c**) and quantitative (**d**) lipidomic data (n = 3 independent experiments).

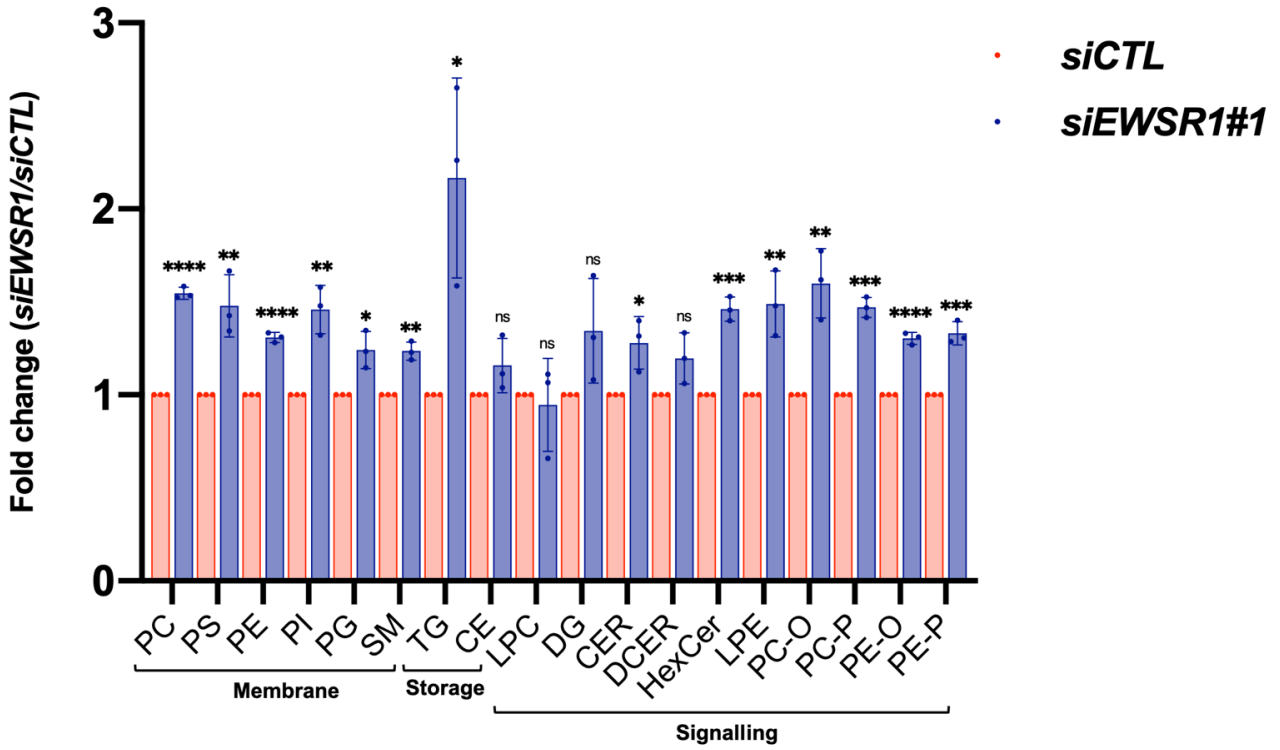


Figure 75: Accumulation of lipids upon EWSR1 KD. Refer to **Figure 73** for all lipid classes abbreviations. Fold change (*siEWSR1*/*siCTL*) of the 18 classes of lipids. Results are shown as means ± sd (n = 3 independent experiments). **P* < 0.05, ***P* < 0.01; ****P* < 0.001, *****P* < 0.0001, ns = not significant compared to the *siCTL* condition by two-tailed unpaired Student's *t* test.

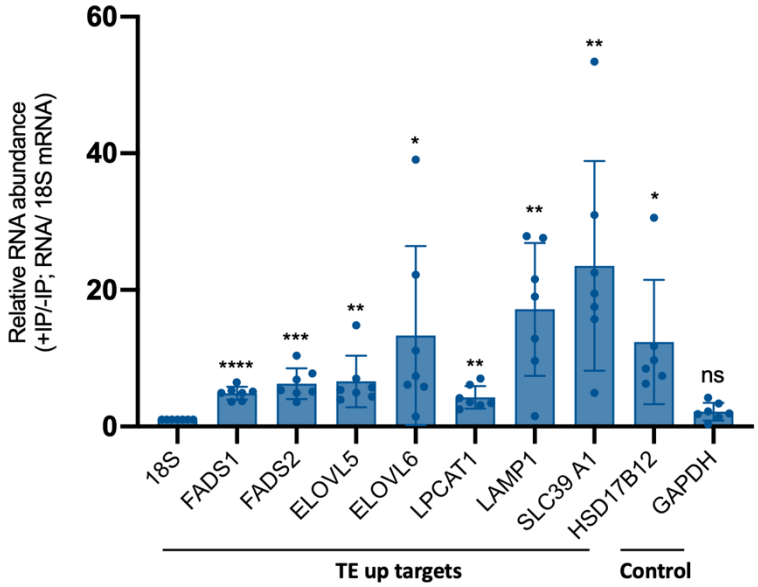
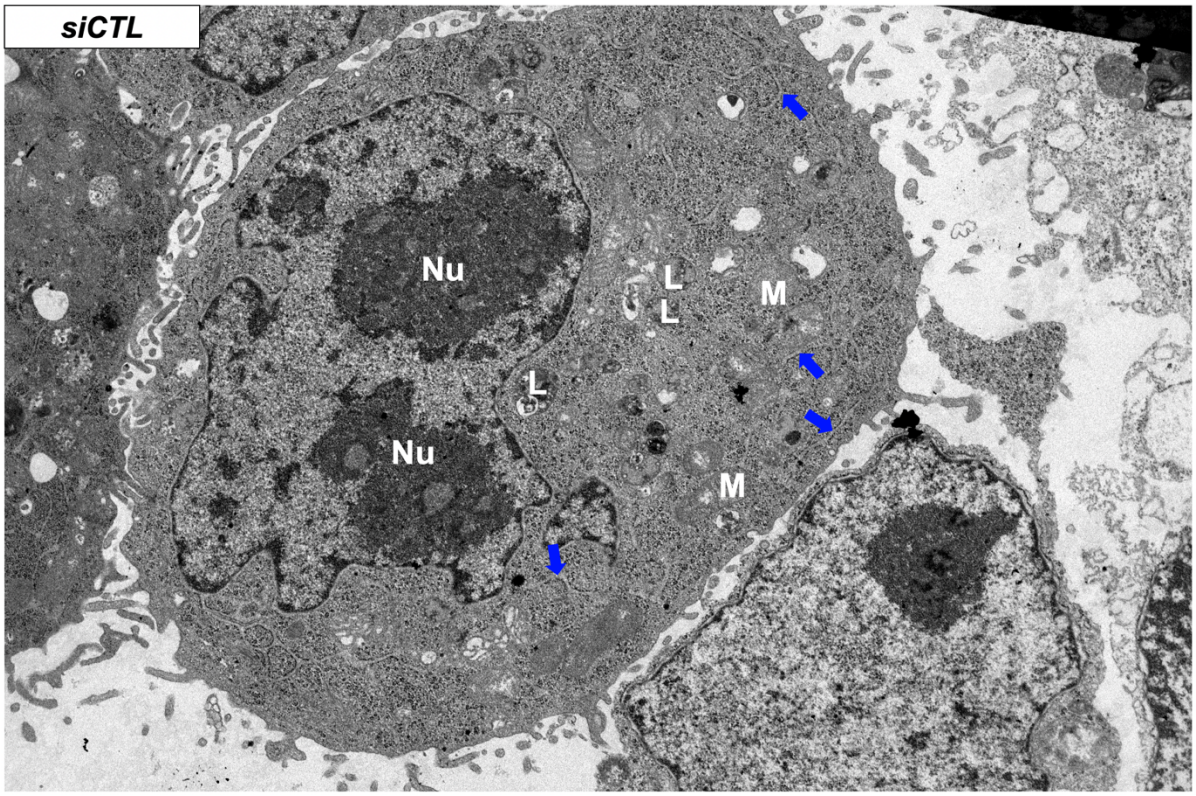
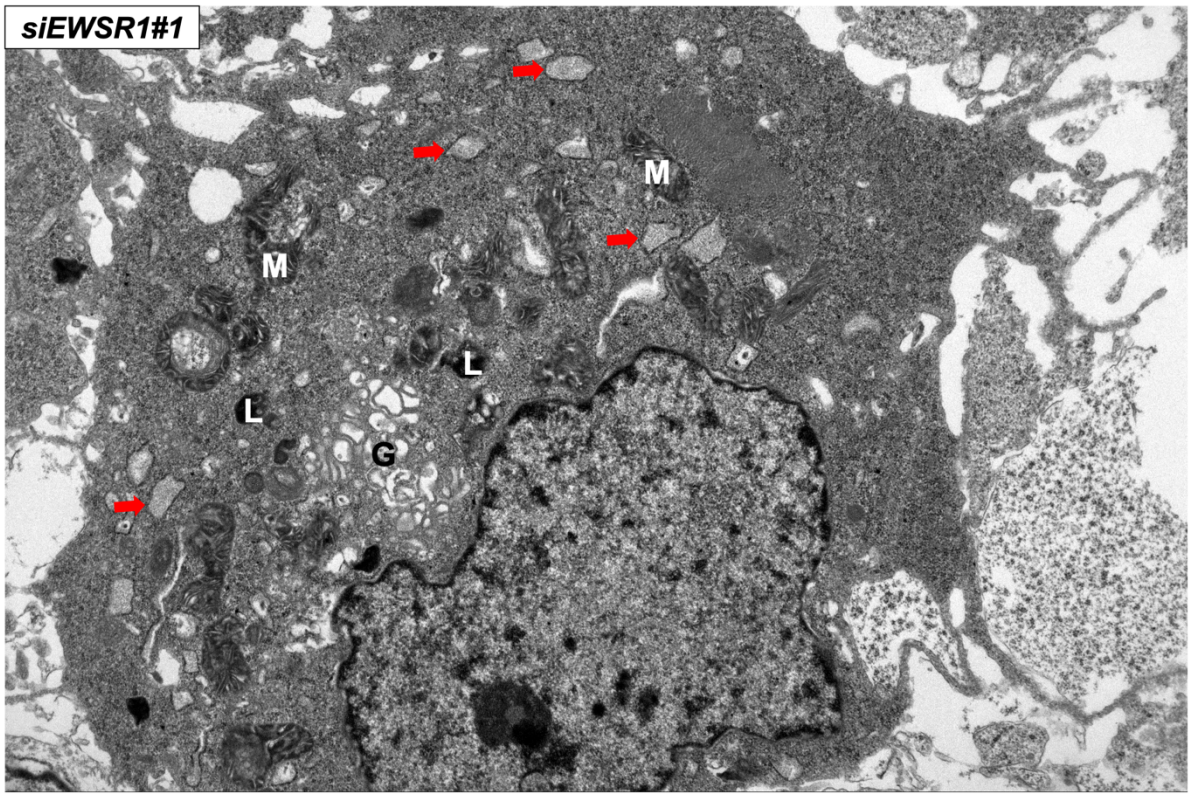


Figure 76: EWSR1 binds to mRNA targets associated with lipid metabolism. RT-qPCR detection of various EWSR1 TE up targets mRNA and control gene (GAPDH) in HeLa cells (n = 7 independent experiments). **P* < 0.05, ***P* < 0.01; ****P* < 0.001, *****P* < 0.0001, ns = not significant using one sample *t* test.

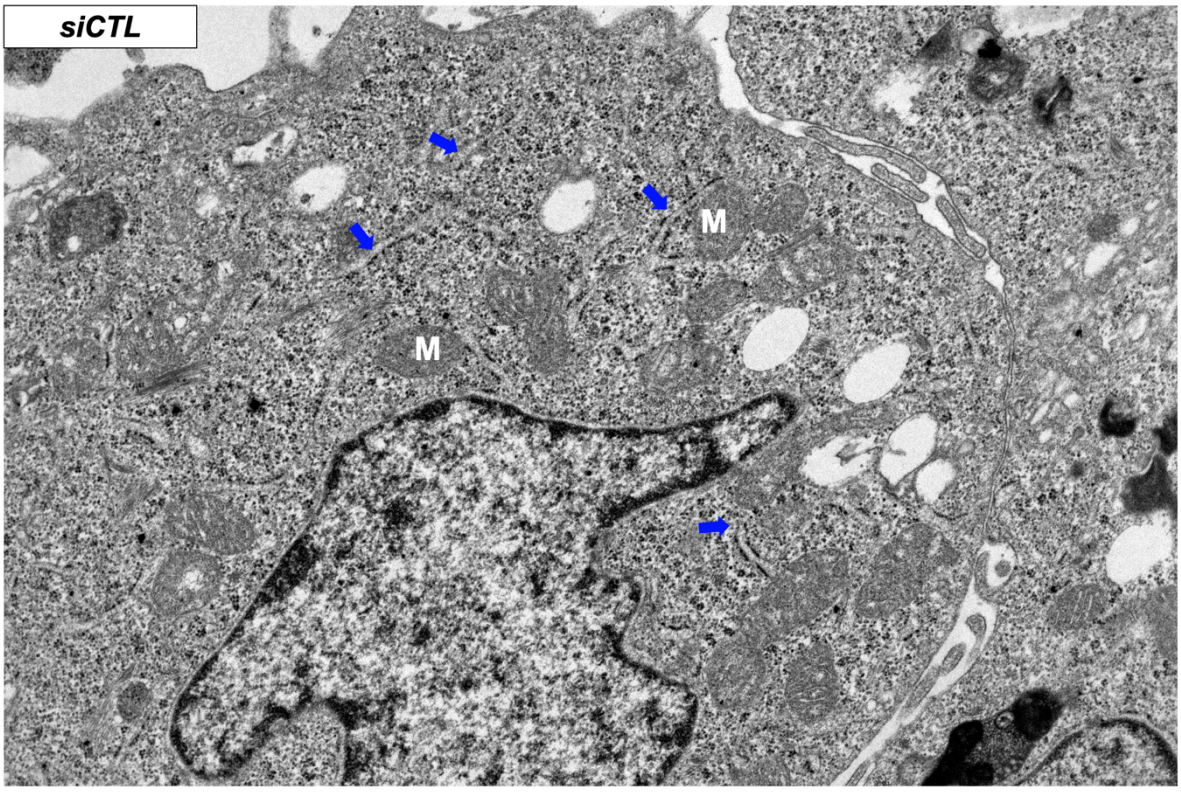
a



b



c



d

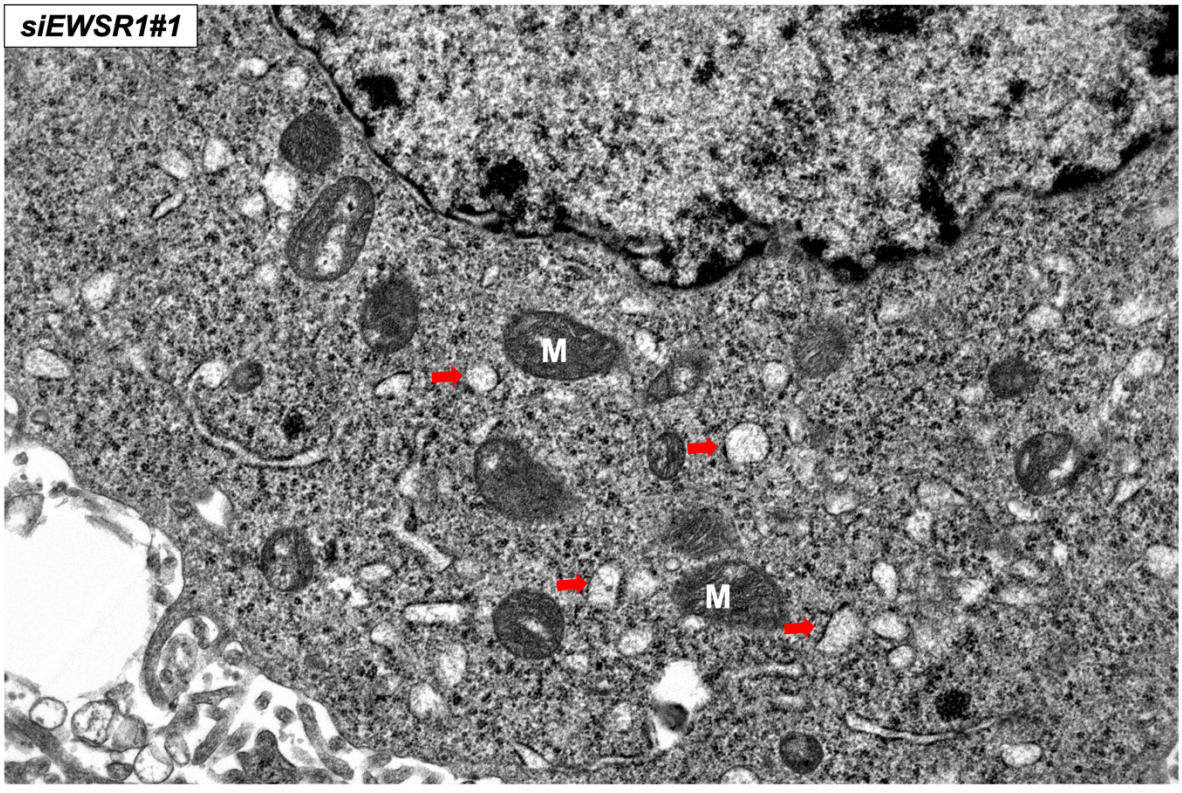


Figure 77: Changes of ER shape upon EWSR1 knockdown. (**a-d**) TEM images showing ER changes in HeLa cells upon knockdown of EWSR1 with *siEWSR1#1*. Blue arrows points to ER in *siCTL* condition and red arrows to ER in *siEWSR1#1* condition. **L**: Lysosomes, **M**: Mitochondria, **G**: Golgi, **Nu**: Nucleolus. (**a**) Scale bars: 2 µm; Mag.: 3000x. (**b**) Scale bars: 2 µm; Mag.: 4000x. (**c, d**) Scale bars: 2 µm; Mag.: 6000x.

3.8 Deletion of EWSR1 impairs the activation of the unfolded protein response of the ER (UPR^{ER})

Signaling sensors within the ER detect perturbations in ER homeostasis and trigger an adaptive stress response known as the unfolded protein response (UPR^{ER})[621]. Interestinlgy, recent findings show that UPR^{ER} sensors are also sensitive to perturbations in lipid homeostasis[622]. Based on our observations that EWSR1 affects lipid homeostasis (**Figure 75**) and ER morphology (**Figure 77**), we investigated the UPR^{ER} status in EWSR1 KD cells.

Activation of the UPR^{ER} involves a set of three transmembrane ER-resident proteins, including IRE1α, PERK, and ATF6, each of which triggers a specific downstream signaling pathway (**see Figure 36 in the Introduction**). Using classical markers of the three UPR^{ER} branches, we monitored the UPR^{ER} status of *siCTL*- and *siEWSR1*-treated cells[623]. With regards to the PERK pathway, we found that KD of EWSR1 correlated with a significant reduction in PERK levels (27%). This also correlated with an almost 50% reduction (47%) of the phosphorylation of its downstream target eIF2α and a more than 70% decrease (72%) in protein levels of ATF4 (**Figures 78a,b**).

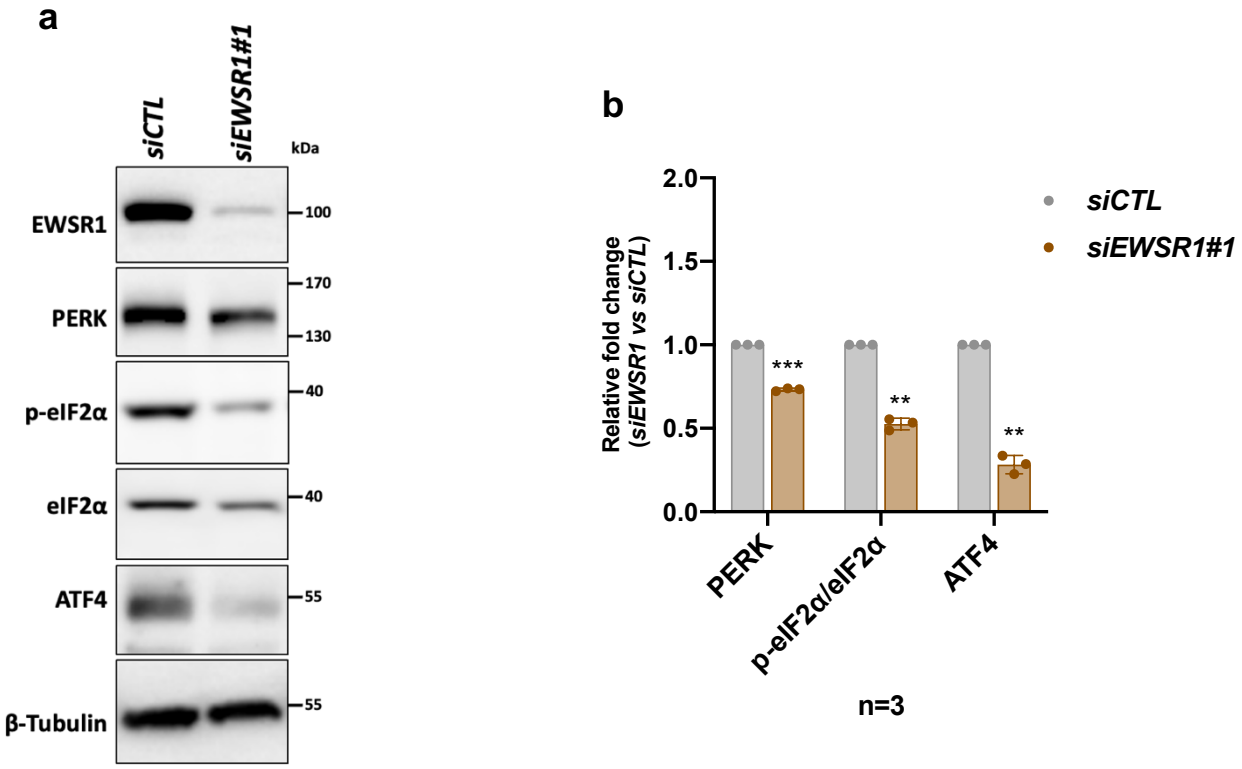
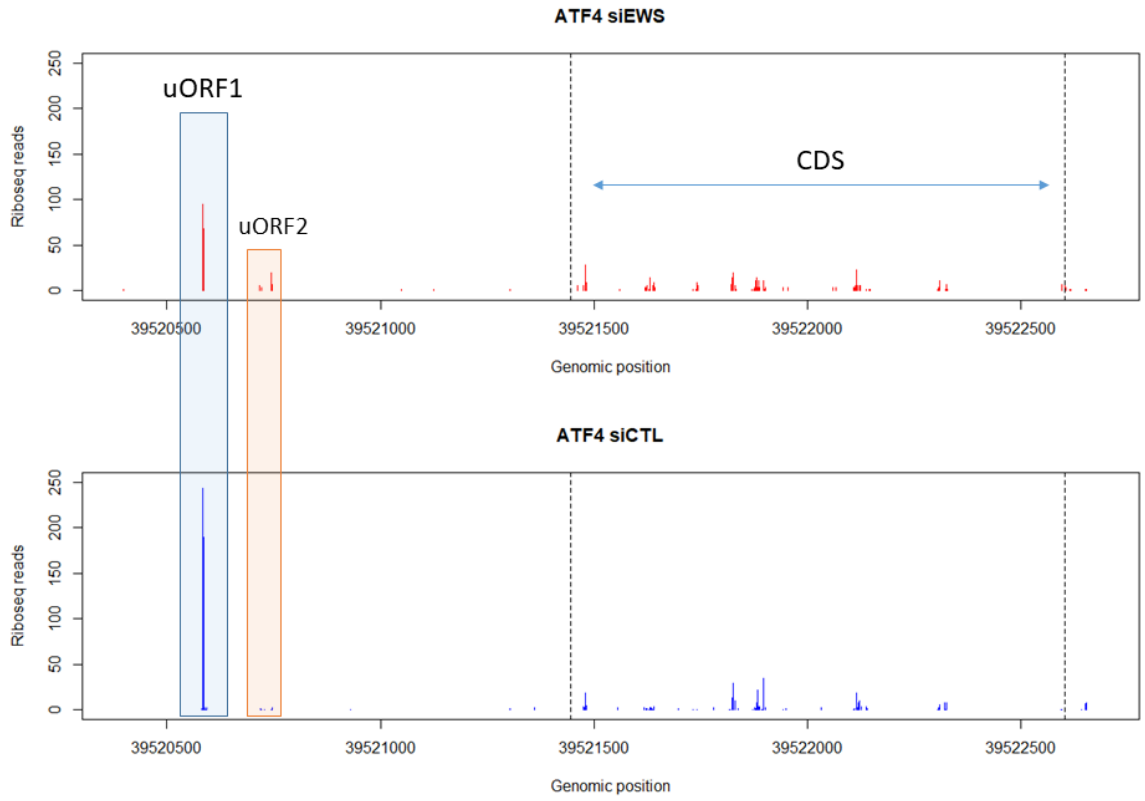


Figure 78: EWSR1 knockdown impairs the activation of the PERK axis of the UPR^{ER}. (**a**) Representative western blot analysis of EWSR1, PERK, p-eIF2α, eIF2α, ATF4 and β-Tubulin (loading control). Samples are total cell lysates from HeLa cells transfected with *siEWSR1#1* or *siCTL*. (**b**) Protein levels were quantified by band densitometry analysis using ImageJ. Results are shown as means ± sd (n = 3 independent experiments) after normalization to β-Tubulin and relative to *siCTL* condition. ***P* < 0.01; ****P* < 0.001 compared to the *siCTL* condition by one sample *t* test.

At the translational level, ATF4 expression is controlled by two short upstream open reading frames (uORF1 and uORF2) within the 5'UTR of its mRNA[509]. Higher translation of uORF2 prevents ATF4 from being properly translated, as it overlaps with the start codon of the ATF4 CDS. Using our RIBO-seq data, we assessed the ribosome distribution profile across the *ATF4* mRNA and quantified the proportion of RPF within uORF1 and uORF2. While no RPF could be mapped to uORF2 in the *siCTL* condition, ribosomal footprints were detected in uORF2 in EWSR1 KD cells, suggesting that ATF4 translation is decreased in cells KD for EWSR1 (**Figure 79a**). These observations confirmed our WB analyses (**see Figure 78a**) and show that lack of EWSR1 correlates with a reduction of ATF4 expression through the classical uORF-mediated regulation, downstream of PERK. Interestingly, we found that classical ATF4 targets, such as *HERPUD1*, *PSAT1*, *TRIB3*, *MTHFD2*, were downregulated in EWSR1 KD cells (**Figure 79b**). Together, these results show that activation of the PERK/ATF4 branch is reduced in EWSR1 KD cells.

a



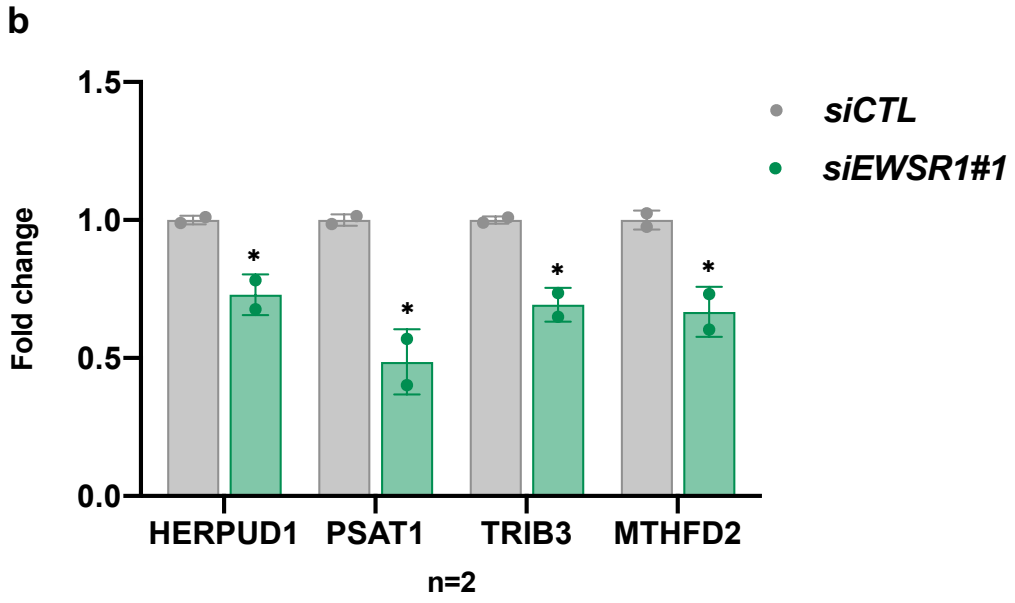


Figure 79: EWSR1 promotes expression of ATF4 and its downstream targets. (**a**) RIBO-seq reads of ATF4 at uORF1, uORF2 and CDS at genomic position. (**b**) Fold change (*siEWSR1/siCTL*) of the four downstream targets of ATF4 (*HERPUD1, PSAT1, TRIB3, MTHFD2*) analyzed from the RNA-seq library. *$P < 0.05$ compared to the *siCTL* condition by multiple *t* test.

In parallel, we also observed that KD of EWSR1 correlated with an inhibition of the ATF6 branch of the UPR^{ER}, as assessed by a decrease in protein level of ATF6 (**Figures 80a,b**). Finally, EWSR1 KD cells showed a significant decrease in the phosphorylation of IRE1α and splicing of *XBP1* (**Figures 81a-c**). Overall, these results indicate that EWSR1 KD impairs the activation of the three branches of the UPR^{ER} in HeLa cells.

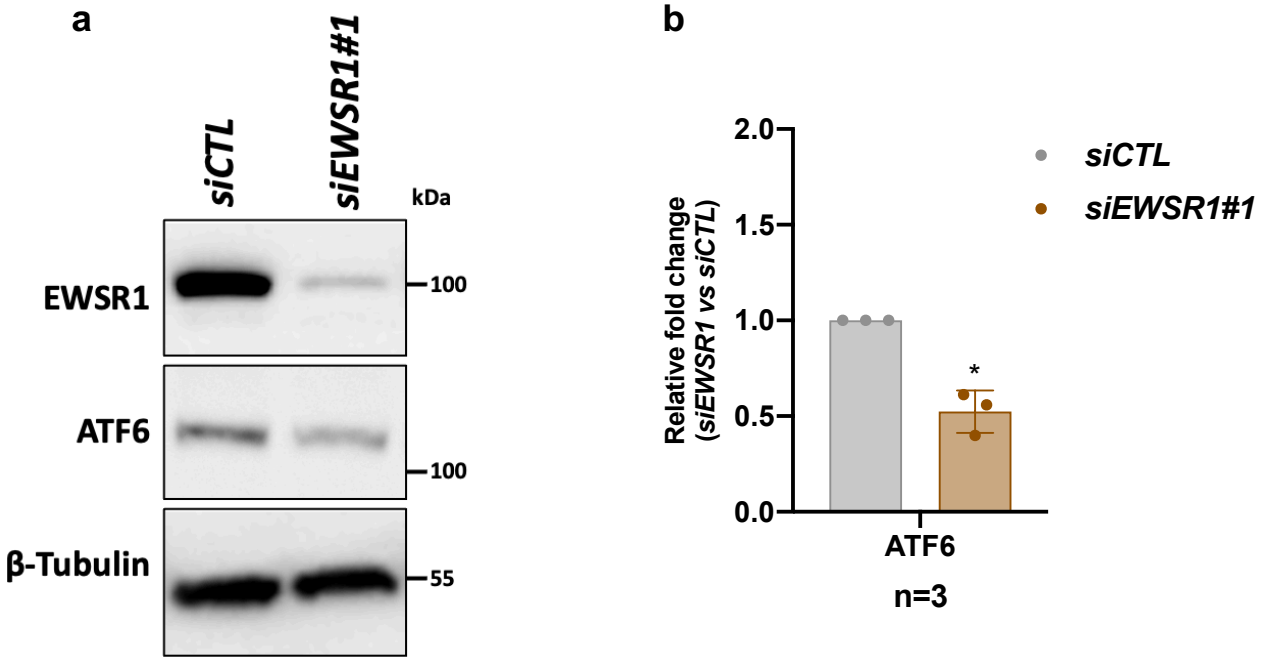


Figure 80: EWSR1 knockdown impairs the activation of ATF6 axis. (a) Representative western blot analysis of EWSR1, ATF6 and β-Tubulin (loading control). Samples are total cell lysates from HeLa cells transfected with *siEWSR1#1* or *siCTL*. (**b**) The levels of the proteins were quantified by band densitometry analysis using ImageJ. Results are shown as means ± sd (n = 3 independent experiments) after normalization to β-Tubulin and relative to the *siCTL* condition. **P* < 0.05 compared to the *siCTL* condition by one sample *t* test.

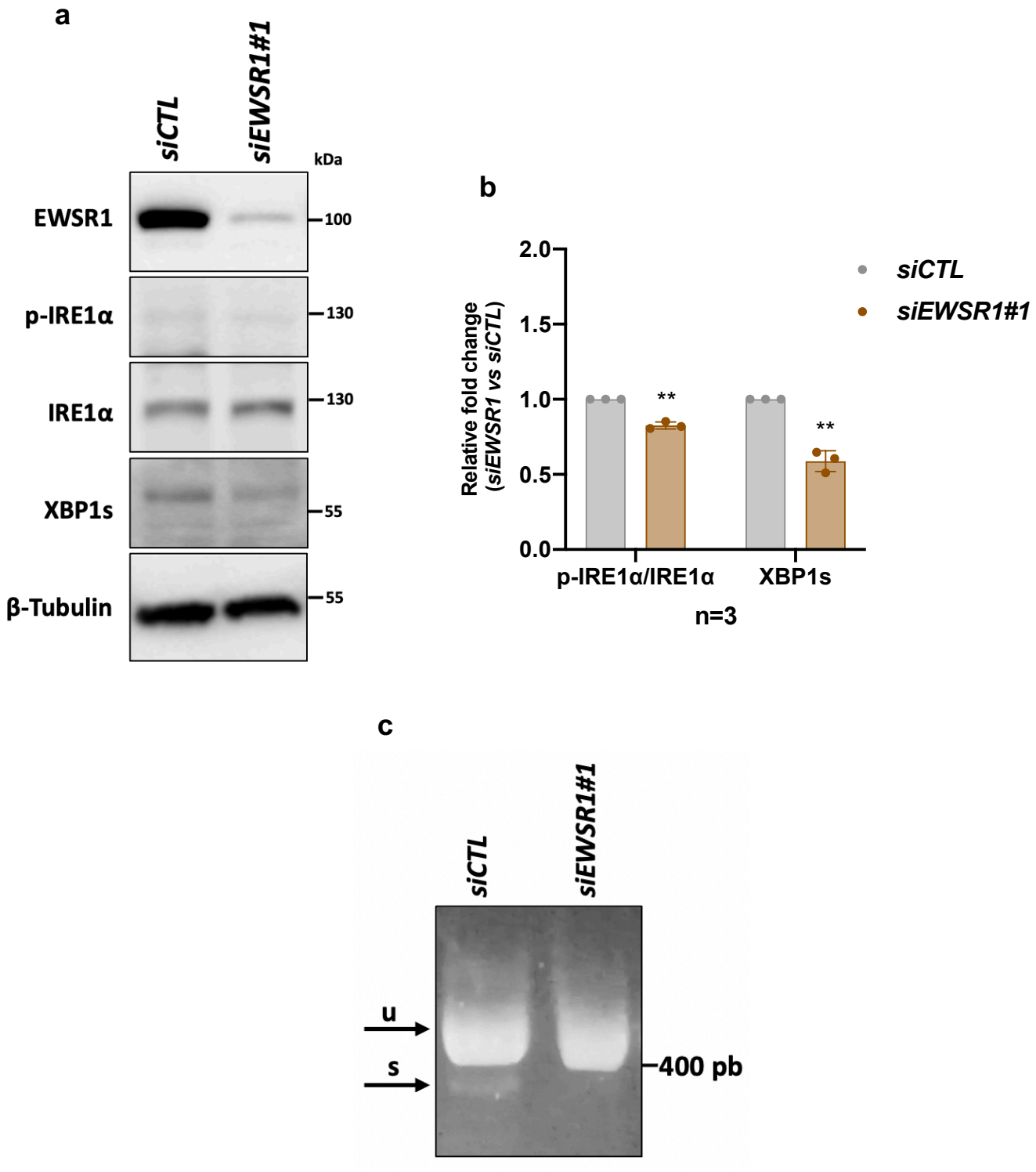
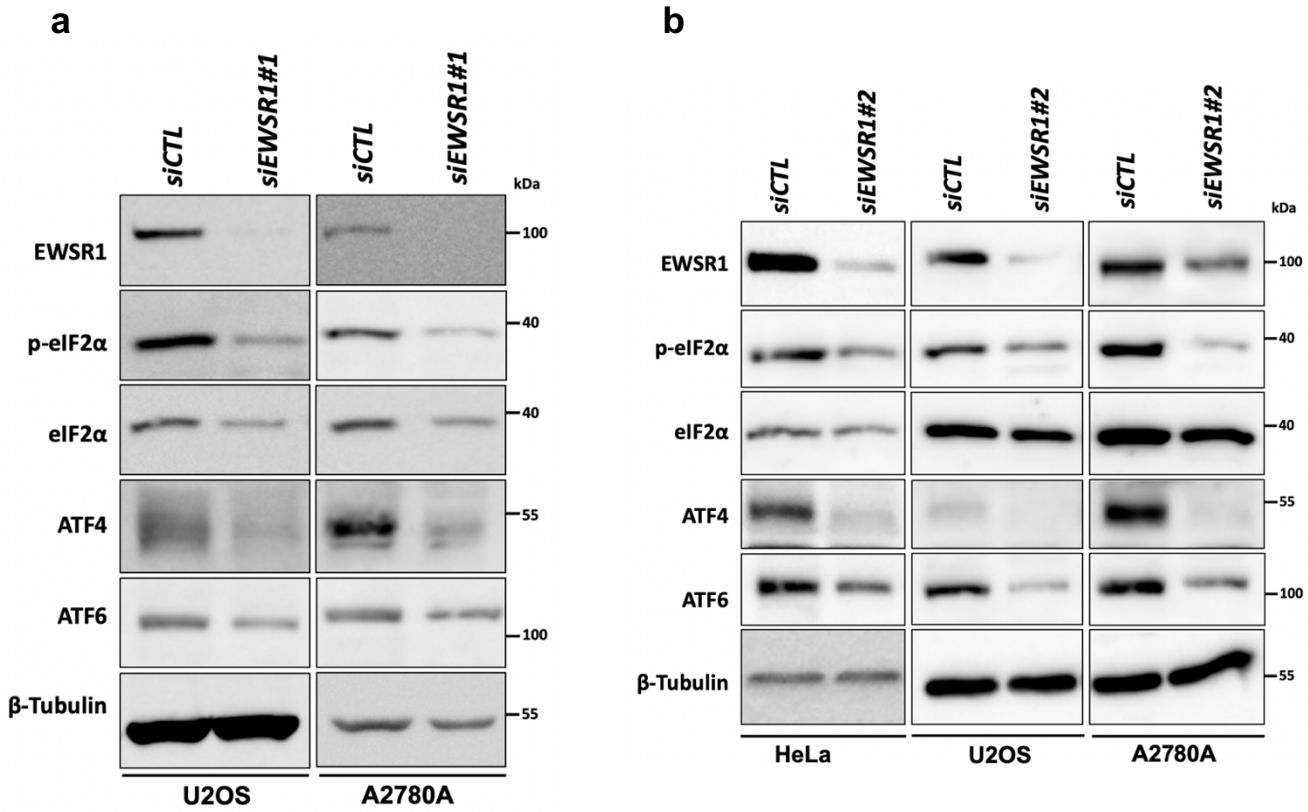


Figure 81: EWSR1 knockdown impairs the activation of IRE1α axis. (**a**) Representative western blot analysis of EWSR1, IRE1α axis and β-Tubulin (loading control). Samples are total cell lysates from HeLa cells transfected with *siEWSR1#1* or *siCTL*. (**b**) The levels of the proteins were quantified by band densitometry analysis using ImageJ. Results are shown as means ± sd (n = 3 independent experiments) after normalization to β-Tubulin and relative to the *siCTL* condition. ***P* <0.01 compared to the *siCTL* condition by one sample *t* test. (**c**) MetaPhor® Agarose gel (2.5%) analysis showing the amplicons corresponding to unspliced (u) and spliced (s) XBP1 obtained by RT-PCR from RNA extracted from HeLa cells transfected with *siEWSR1#1* or *siCTL*. The RT-PCR products are illustrated.

To assess if the effects of EWSR1 KD on the UPR^{ER} were limited to HeLa cells, we checked the consequences of depleting EWSR1 in non HeLa cells. We used two other unrelated human cancer cell lines: human osteosarcoma (U2OS) and ovarian cancer (A2780A) cells. As for HeLa cells, we observed a global decrease in the UPR^{ER} following knockdown of EWSR1 (**Figure 82a**). Importantly, these observations were replicated with a second EWSR1 siRNA (*siEWSR1#2*) (**Figures 82b,c**).

Taken together, our findings suggest that expression of EWSR1 sustains activation of the UPR^{ER} under basal conditions, pointing towards a role of EWSR1 in the maintenance of basal cellular homeostasis in cancer cells.



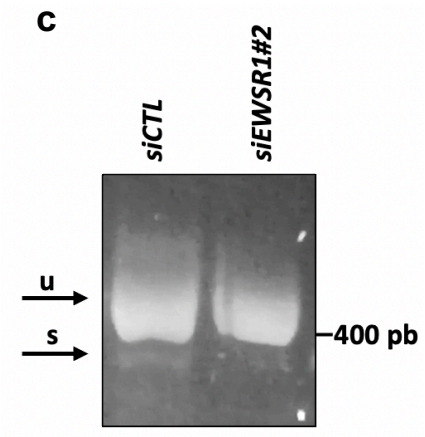


Figure 82: EWSR1 knockdown impairs the activation of UPR^{ER} in U2OS and A2780A cell lines. (**a**) Western blotting of EWSR1, p-eIF2α, eIF2α, ATF4, ATF6 and β-Tubulin (loading control). Samples are total cell lysates from U2OS and A2780A cells transfected with *siEWSR1#1* or *siCTL*. (**b**) Western blotting of EWSR1, p-eIF2α, eIF2α, ATF4, ATF6 and β-Tubulin (loading control). Samples are total cell lysates transfected with *siCTL* or *siEWSR1#2* from HeLa, U2OS and A2780A cells. (**c**) MetaPhor® Agarose gel (2.5%) analysis showing the amplicons corresponding to unspliced (u) and spliced (s) XBP1 obtained by RT-PCR from RNA extracted from HeLa cells transfected with *siEWSR1#2* or *siCTL*. The RT-PCR products are illustrated.

3.9 EWSR1 KD leads to impairment activation of UPR^{ER} under ER stress

Based on the observation that depletion of EWSR1 correlates with defective UPR^{ER} in basal conditions, we sought to test the ability of EWSR1 KD cells to trigger the UPR^{ER} in response to stress. Several molecules that induce ER stress through a variety of mechanisms, have been identified. ER-stress inducers such as tunicamycin (Tu) inhibits N-linked glycosylation of proteins[624], whereas thapsigargin (Tg), a sesquiterpene lactone, targets the Sarco/ER Ca^{2+}-ATPase (SERCA) and reduces Ca^{2+} concentrations in the ER, which impairs its protein folding capacity[625]. Besides proteotoxic stresses caused by thapsigargin and tunicamycin[523], exposure to saturated free fatty acids such as palmitic acid (PA), increases *de novo* biosynthesis of saturated phospholipids leading to aberrant phospholipid metabolism, which may contribute to palmitate-induced ER stress and lipotoxicity[626].

To investigate the impact of EWSR1 on the stress-induced UPR^{ER}, we incubated HeLa cells with tunicamycin (Tu, 0.5 µg/ml for 5 hours), thapsigargin (Tg, 500 nM for 24 hours) or palmitic acid (PA, 800 µM for 24 hours). As expected, each stress dramatically activated the UPR^{ER}, although they affected the three branches differently (**Figure 83**). For instance, the ATF6 branch was more activated by thapsigargin than by tunicamycin, in agreement with previous observations[627].

Nevertheless, no matter the applied ER-stress, depletion of EWSR1 significantly reduced and sometimes completely prevented activation of the UPR^{ER}.

C/EBP homologous protein (CHOP), also known as growth arrest and DNA damage-inducible protein 153 (DDIT3/GADD153) is a stress-responsive transcription factor onto which converge all three branches of the mammalian UPR^{ER}, although the PERK/eIF2α/ATF4 axis is thought to be predominant in CHOP activation[628]. Using expression of CHOP as a proxy for distal activation of the UPR^{ER}, we found that CHOP expression was systematically reduced in EWSR1 KD, both in nontreated (NT) or ER-stress induced conditions (**Figure 83**). When activated by the UPR^{ER}, CHOP represses the expression of the cell cycle regulator $p21^{Cip1}$[629]. In agreement with lower levels of CHOP, we found higher levels of $p21^{Cip1}$ in EWSR1 KD cells, both in the absence or presence of ER stresses (**Figure 83**).

Altogether, these observations demonstrate that EWSR1 is necessary for full activation of proximal and distal UPR^{ER} signaling and strongly suggest that it might promote the ability of cells to respond to ER stress.

Activation of UPR^{ER} is considered as a rapid adaptative response, that can have contrasting outcomes on cell survival and be either a pro-apoptotic or pro-survival pathway, depending on the nature and duration of the stress, but also on the cell type[630]. For instance, while chronic UPR^{ER} usually leads to self-destruction through apoptosis in most cells, it is beneficial to cancer cells, as it is thought to enhance their survival, proliferation, metastatic capacity, drug resistance and angiogenic properties[631]. We examined the ability of HeLa cells to cope with ER stress in the absence of EWSR1. In the absence of ER stressors, we observed no differences in apoptosis between *siCTL*- and *siEWSR1#1*-transfected cells using Annexin V staining and flow cytometry analysis. This indicates that knockdown of EWSR1 does not significantly affect cell survival in the absence of acute ER stress (**Figures 84a-c**). As expected, thapsigargin, tunicamycin or palmitic acid treatments all induced significant apoptosis in *siCTL* cells[632–634]. The effects of tunicamycin (**Figure 84a**) and thapsigargin (**Figure 84b**) were not statistically different in control and EWSR1 KD cells, indicating that depletion of EWSR1 does not sensitize cells towards decreased levels of Ca^{2+} or accumulation of unfolded glycoproteins in the ER. In contrast, cells KD for EWSR1 were significantly more sensitive to PA treatment (**Figure 84c**). A prominent event during apoptosis is the selective cleavage of poly (ADP-ribose) polymerase (PARP) by a specialized family of cysteinyl-aspartate proteases (caspases), especially by caspase-3[635]. Western blot analysis showed no significant changes in PARP expression, between *siCTL* and *siEWSR1#1* in nontreated cells (NT) or following treatments with tunicamycin or thapsigagin (**Figure 83**). In contrast, PARP cleavage induced by PA treatment was more pronounced in EWSR1 KD cells compared to control cells (**Figure 83**), confirming that cells depleted for EWSR1 are more sensitive to excess lipids. This specificity towards PA-induced apoptosis is remarkable, as it correlates with our above lipidomic observations (**Figure 75**) and suggests that because they are already overloaded with lipids, EWSR1 KD cells might be more sensitive to lipitoxicity.

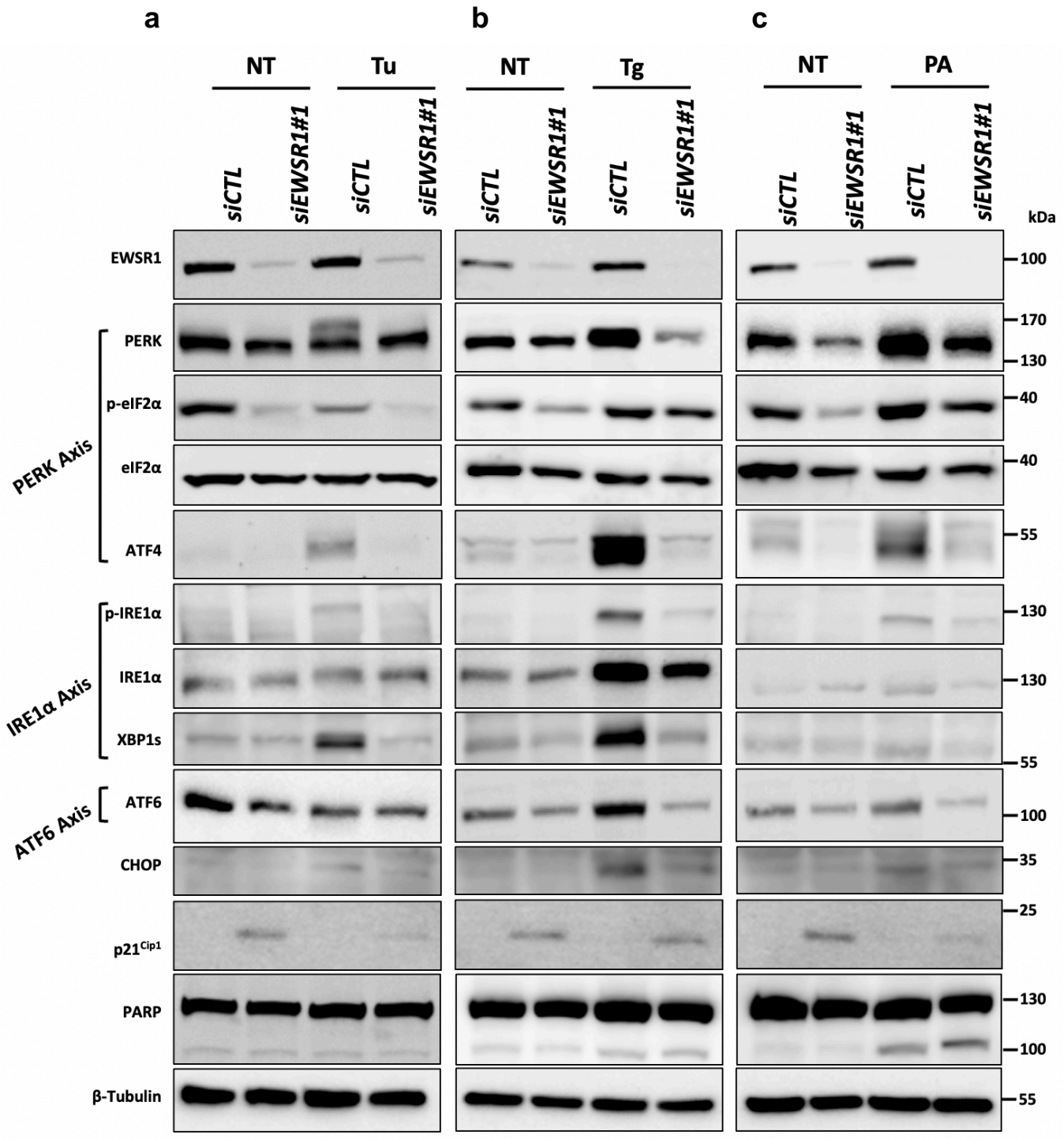


Figure 83: Knockdown of EWSR1 impairs the activation of the UPR^{ER} under ER stress. (**a, b, c**) Western blot analysis of markers of the PERK, IRE1α and ATF6 axis of the UPR^{ER}, CHOP, p21^{Cip1}, PARP and β-Tubulin (loading control) in cells transfected with *siEWSR1#1* or *siCTL*. Samples are total cell lysates from HeLa cells in nontreated condition (NT) or treated with tunicamycin (Tu; 0.5 µg/ml – 5 h) (**a**), thapsigargin (Tg; 500 nM – 24h) (**b**) or palmitic acid (PA; 800 µM – 24h) (**c**).

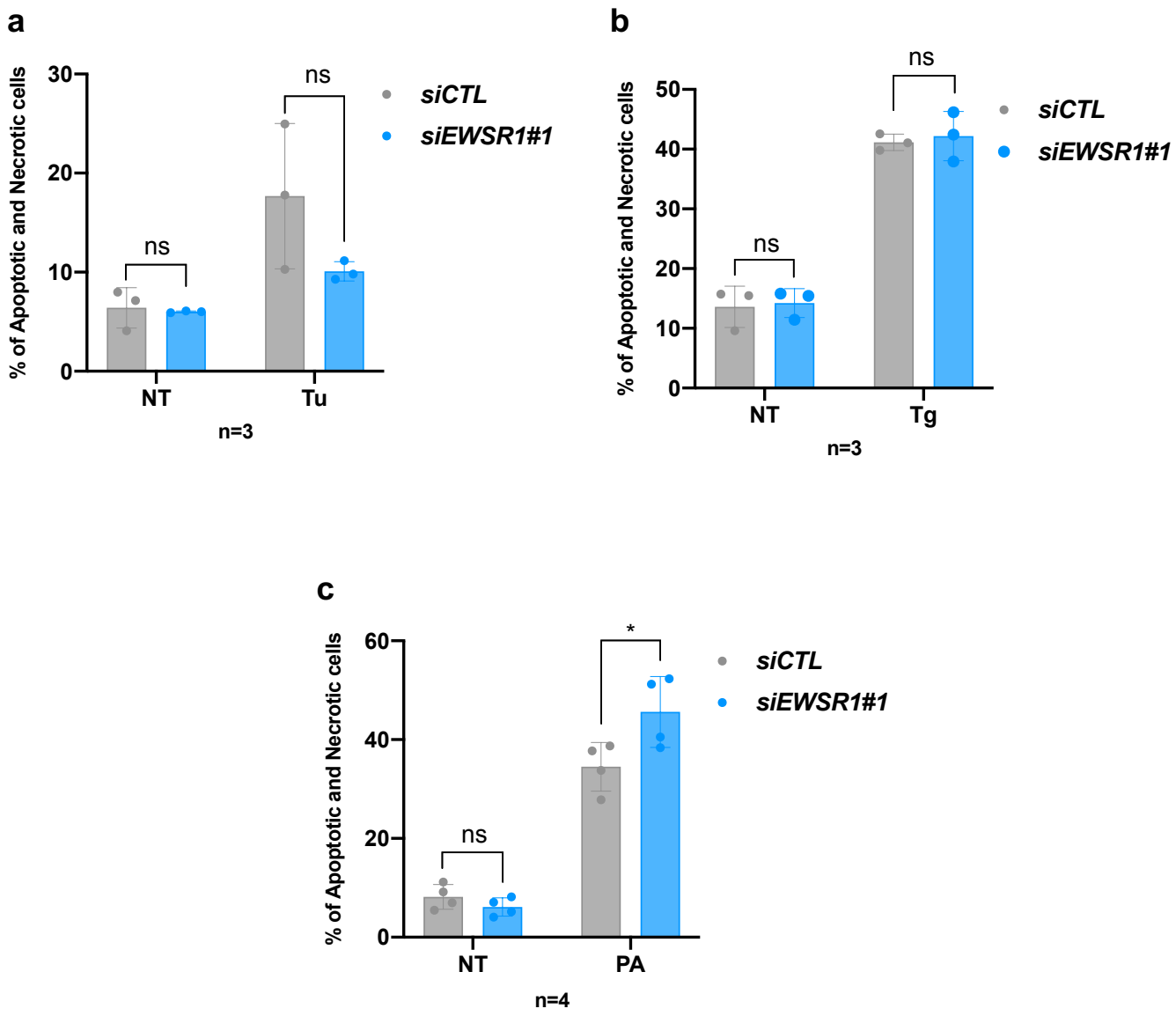


Figure 84: EWSR1 KD leads to increased apoptosis in PA treated condition. (**a, b, c**) Bar plots showing the percentage of apoptotic and necrotic cells. Samples are HeLa cells transfected with *siEWSR1#1* and *siCTL*, nontreated (NT) or treated with tunicamycin (Tu; 0.5 µg/ml – 48 h) (**a**), thapsigargin (Tg; 500 nM – 48h) (**b**) or palmitic acid (PA; 800 µM – 24h) (**c**). Results are means ± sd (n ≥ 3 independent experiments). *$P < 0.05$, ns = not significant compared to the *siCTL* condition by two-tailed unpaired Student's *t* test.

4 DISCUSSION AND PERSPECTIVES

In this section, we provide further results that we did not include in the "**RESULTS**" section but are useful for the discussion. We plan not to include them in our publication that will follow the same order of the above "**3 RESULTS**" section. We also cover future perspectives to this work. Some aim to consolidate the present findings by suggesting alternative experiments, others might provide guidelines for future research projects.

4.1 EWSR1 is an inhibitor of mRNA translation

Gene expression levels and protein abundance are subjected to a strict regulation that determines the appropriate time and place at which transcripts will be translated. This sophisticated control is crucial to maintain diverse cellular processes, such as development, metabolism and cancer progression[636]. In contrast to the old-fashion view, gene expression is now considered as a streamlined process, in which mRNA synthesis and processing events are coordinated to one another. Another emerging concept is that of multifunctionality, which is becoming the rule rather than the exception in the world of proteins[637]. It can be exemplified by RNA binding proteins (RBPs) that orchestrate transcription and mRNA metabolism.

Being a multifunctional protein, EWSR1 has been implicated in transcriptional and post-transcriptional mechanisms[123]. Particularly, the implication in post-transcriptional regulation prompted us to investigate EWSR1 function in mRNA translation. Although previous reports pointed towards a possible link between EWSR1 and translation, by showing i) its binding to ribosomes[113], ii) its role in mRNA nuclear retention or transport[25,120], two molecular mechanisms that can affect the rate of translation and iii) its localization to SGs[6], the cytoplasmic foci in which untranslated mRNA are stored; this work represents the first demonstration that EWSR1 is actively involved in mRNA translation (**Figure 85**). Interestingly, our findings show that EWSR1 is an important repressor of a specific subset of mRNA (**Figure 51a**). We also identified the RGG2-ZnF-RGG3 within the CTD of EWSR1 as the region involved in both functional repression of translation (**Figures 61c,e**) and association with the 40S ribosomal subunit (**Figure 67**). Indeed, a mutant lacking this region (*i.e.,* the ΔRGG2-ZnF-RGG3 mutant) was not able to functionally repress the translation, nor to associate with the 40S.

Importantly, we found that EWSR1 by regulating the translation of specific mRNAs (*e.g.*, mRNAs related to lipid metabolism) (**Figure 72b**), affected lipid homeostasis and consequently, impacted ER shape (**Figure 77**) and the unfolded protein response (UPR^{ER}) and its downstream effects (**Figures 78-82**). Therefore, our findings provide a better understanding of *bona fide* roles of EWSR1 in maintaining cellular homeostasis, by controlling specific mRNA translation.

Particularly, our findings might lead to relevant therapeutic opportunities for EWSR1-related disorders for two reasons. Firstly, we uncovered the essential region for EWSR1-mediated translational regulation, the CTD (**Figures 60, 65, 66**), which was associated with many diseases. Secondly, we showed that translational dysregulation by EWSR1 was coupled with perturbated ER homeostasis (**Figure 77**) and improper activation of the UPR^{ER} (**Figures 78-82**), which were linked to the emergence of many diseases. Indeed, as we previously described (**see Introduction; "1.3.10 ER stress and diseases"**), it is well known that impaired ER homeostasis results in aberrant cellular responses, contributing to the pathogenesis of various diseases, and that improper activation of the UPR^{ER} is dangerous, as it can destroy the cell or protect it against death. Overall, this suggests that translational dysregulation by EWSR1, due to CTD mutations or loss, might be a critical factor in the emergence of several diseases (**see below; "4.6 Deregulation of EWSR1 translational activity might be implicated in several diseases"**).

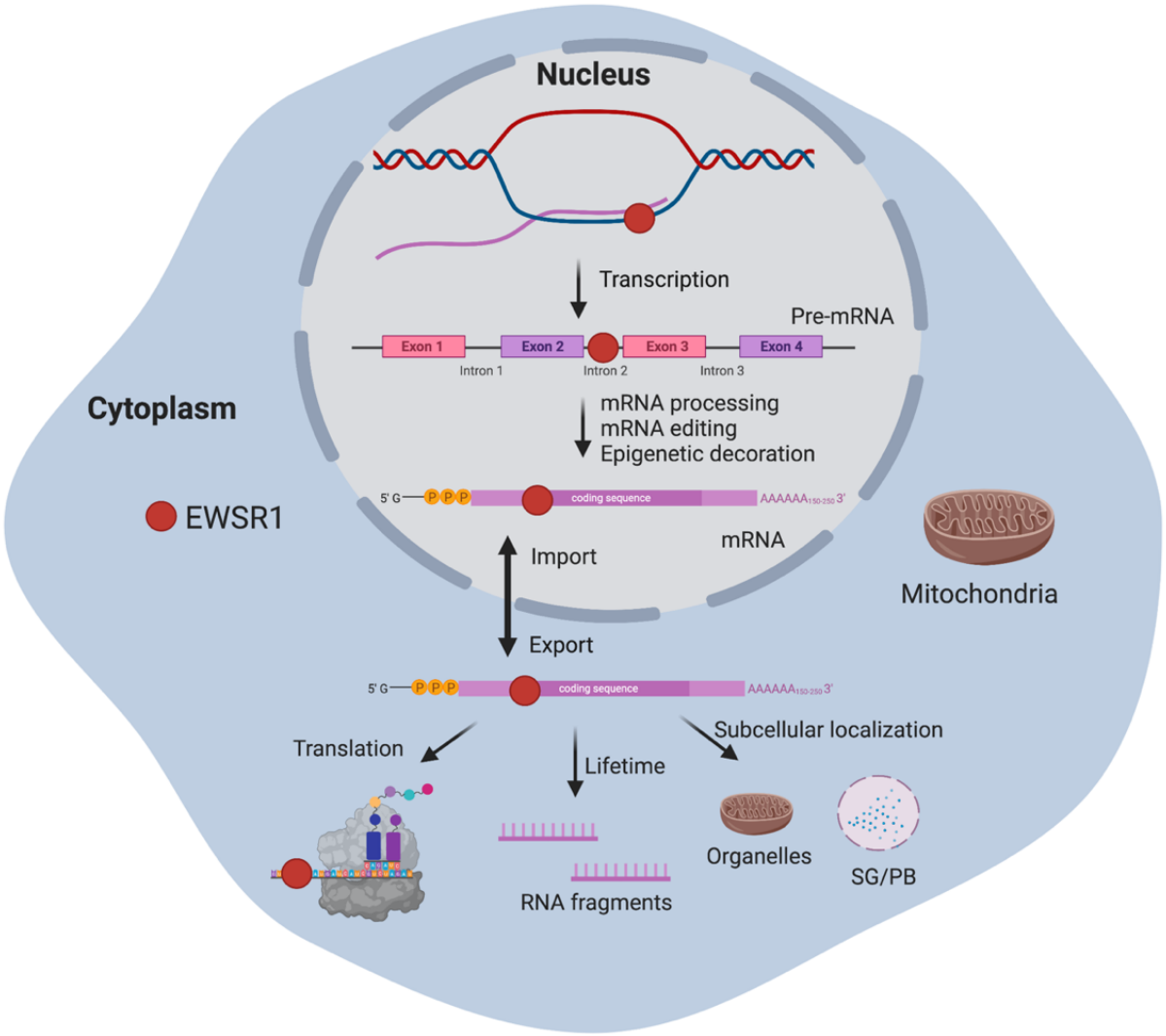


Figure 85: Illustration of the implication of EWSR1 in transcriptional and post-transcriptional regulation. Created with BioRender.com. SG: stress granules, PB: processing bodies.

4.2 RIBO-seq: a powerful tool to determine the translational regulation by EWSR1

Although many details of post-transcriptional gene regulation (PTGR) remain to be revealed, the rapid development of next-generation sequencing-based methods, such as ribosome profiling (RIBO-seq) aided to investigate PTGR processes (*e.g.,* transport, stability, translation, degradation of RNA) in various cellular systems[638]. Ribosome profiling has revolutionized the field of translation by revealing many of its underlying mechanisms (**see Introduction; "1.2.6.4 Ribosome profiling (RIBO-seq)"**). Therefore, being an ideal method for sensitive in-depth quantification of temporal changes in the translatome, we conducted RIBO-seq to characterize the translational regulation by EWSR1.

As RIBO-seq allows to measure mRNA abundance by mRNA-seq, we were able to gain additional insights about the regulation of mRNA expression by EWSR1. Interestingly, our RNA-seq data highlighted that EWSR1 showcases promising roles in several cell survival pathways, such as signaling, cell cycle and proliferation, by controlling the expression of its RNA targets (**Figures 48-50**), which correlated with previous reports[120,123,586]. In addition, we used translational efficiency (TE) (**Figure 51a**), which is commonly used in the literature, as a key metric of measuring translational control. However, it is important to note that TE is not a direct measure of protein output, as RIBO-seq measures "to be translated mRNAs". Therefore, RIBO-seq may be more correlated with newly synthesized proteins than total proteins (newly synthesized plus pre-existing proteins), which are usually detected with mass spectrometry[639], as will be further discussed below.

Intriguingly, our genome wide analysis of TE by combining RIBO-seq data with corresponding RNA-seq data revealed that 595 mRNA showed increased TE (TE up) and 222 showed decreased TE (TE down) (**Figure 51a**). Remarkably, DTEGs presented enrichment in processes (**Figure 72**) different from that of DEGs (**Figure 49**) and GO of TE up genes (**Figure 72b**) were different from TE down genes (**Figure 72a**). Many TE up mRNAs were enriched with several processes related to lipid metabolism and ER function (**Figure 72b**). Accordingly, we showed that EWSR1 regulates lipid homeostasis (**Figure 75**), which impacts ER shape and function (**Figure 77**). On the other hand, TE down genes were related to translation (**Figure 72a**). This was surprising since our results demonstrate that EWSR1 is a translational inhibitor, but this might result from compensatory mechanisms.

Overall, RIBO-seq was a valuable approach to study the mRNA translation landscape in the presence/absence of EWSR1.

While transcription contributes to coordinated gene expression in time and space, several studies have reported the discordance between levels of mRNA (transcriptome) and protein

production (proteome), both at the steady state and dynamically[640]. The discrepancy implies that the bulk of gene expression regulation must occur post-transcriptionally, for instance at the translational level, which explains the surge of interest in studying the translatome in the recent years. Indeed, translatome estimations of gene expression levels correlate better with proteomic data than transcriptomic analysis[641]. However, while the relationship between footprint occupancy and protein synthesis has been shown to be largely true in bacteria[642], it remains less clear in the more complex translational system of eukaryotes[643]. Therefore, an important perspective for this work would be to perform pulsed-SILAC (p-SILAC)[644] to directly monitor the synthesis of new proteins between *siCTL* and *siEWSR1* conditions. p-SILAC has been widely used to monitor new proteins synthesis in various systems using shotgun proteomics approaches. It consists of adding to the cellular media, stable isotope-labeled amino acids, which are subsequently incorporated into all newly synthesized proteins[644]. Thereafter, combining RIBO-seq and p-SILAC data permits to evaluate to which level ribosome footprint density is quantitatively reflective of protein synthesis in our model.

To show how changes in TE could be caused by mRNA up or down or translation up or down, or both, a relevant perspective would be to break down the different classes of mRNAs in **Figure 51a,** following the analysis of **Figures 27d,e**.

4.3 FET proteins and the repression of translation

EWSR1 is a member of the FET family, including FUS and TAF15, which has been implicated in several aspects of DNA and RNA regulation[9]. In this work, we showed that EWSR1 represses specific mRNA translation (**Figures 51a, 52, 53**). An interesting question is whether this activity of EWSR1 is also shared with FUS and TAF15. To this aim, we tested the translational functions of these two related FET members using the MS2-tethering reporter assay. Interestingly, we found that tethering of MS2-FUS or MS2-TAF15 reduced the expression of the *RLuc* protein by 77% and 79%, respectively, with no effect on the abundance of the *RLuc* transcript (**Figure 86**). Moreover, we did not observe an inhibition of *RLuc* translation when transfecting the untethered version of the proteins, thus indicating that FUS-/TAF15-mediated translational repression requires their presence on the target transcript **(Figure 87**)**.** In addition, AGO2 KD did not affect the ability of MS2-FUS or MS2-TAF15 to inhibit translation of the *Renilla* luciferase, suggesting that, like EWSR1, the translational inhibition by FUS and TAF15 is not miRNA mediated (**Figure 88**).

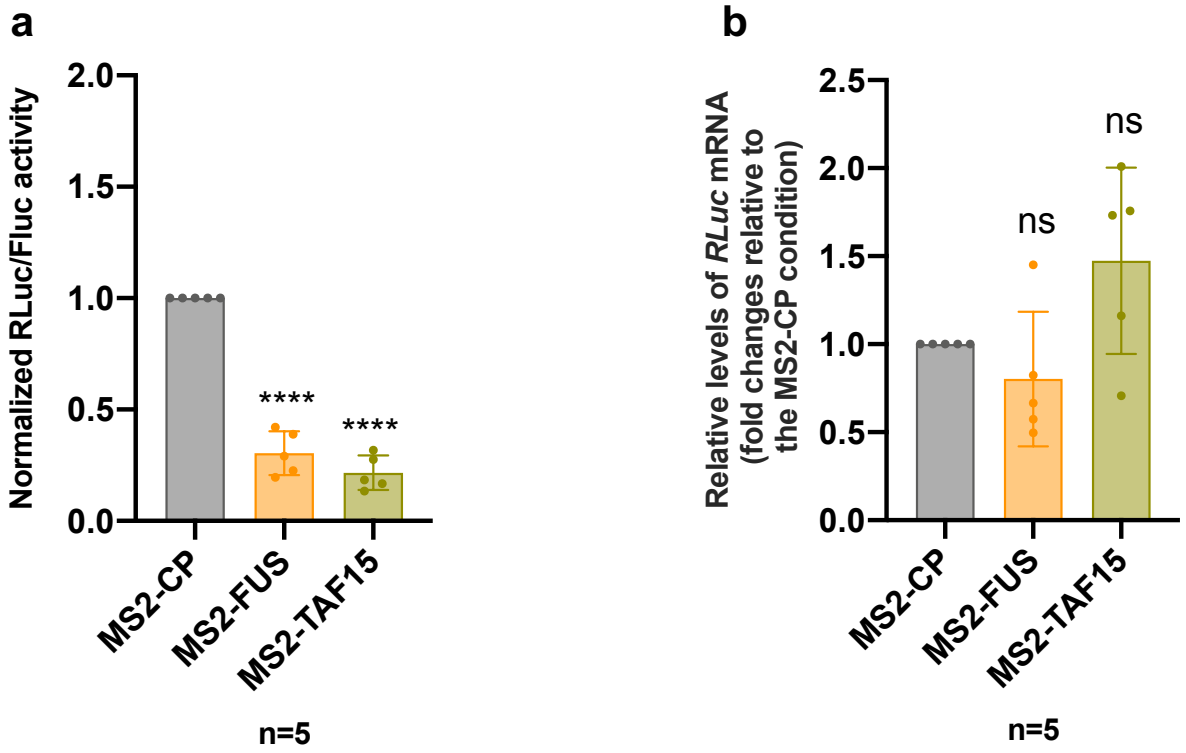
For FUS, these results were expected as several studies revealed its involvement in mRNA translation, besides its numerous roles in the regulation of transcription[645], splicing[646], DNA damage repair[647], RNA transport and stability[648] and microRNA processing[649]. Indeed, previous studies reported that ALS-linked *FUS* mutations were recruited to ribonucleoprotein granules, thus FUS was speculated to be involved in protein translation[650,651]. Another study used three

independent experiments (*Renilla* reporter gene, *in vitro* ^{35}S-methionine (^{35}S-Met) incorporation and SUnSET assays) and showed that mutant FUS impaired protein translation and activated the non-mediated decay (NMD)[122]. Moreover, stress conditions such as glutamate excitotoxicity[652], heat shock, sodium arsenite, or sorbitol treatments induce the localization of FUS to cytoplasmic stress granules (SGs) composed of messenger ribonucleoproteins and stalled mRNAs[650,653,654].

Our results suggest a potential implication of TAF15 in translational control. This is especially interesting considering the paucity of evidence pointing to such a role. To date, a single study suggested a potential involvement of TAF15 in transport and/or local RNA translation based on the observation that it associates with a minor subset of RNA containing granules in the cytoplasm of HT22 cells[655]. Another report showed that FUS and TAF15 localize to SGs in a stress dependent manner and to a larger extend than EWSR1[656].

Furthermore, co-immunoprecipitation experiments indicated that FUS, but not TAF15, was able to associate with ribosomal components, such as RPS6 and RPL26 (**Figure 89**). This correlates with a recent study showing that FUS mediates translational repression through mTOR-dependent signaling and its association with polyribosomes[29], but also indicates the specificity of the binding between EWSR1 or FUS and ribosomal components. At first, the inability of TAF15 to associate with ribosomal components might seem surprising because of the similarity between FET proteins. However, a recent study reported that RGG boxes (which we showed to be primordial for the activity of EWSR1 in the repression of translation (**Figures 61, 67**)) of the CTD within FET proteins can be functionally distinguished, with 75% of RGG boxes in TAF15 were present within prevalent YGGDR(G/S)G repeats that were totally absent from EWSR1 and FUS[10]. This may suggest that TAF15 can repress the translation through a mechanism independent of its interaction with ribosomal particles.

As much progress has been done concerning the translational regulation by FUS[29], an interesting future research project is to better characterize this activity for TAF15. This will give an idea about the way these three related proteins exert their effect on translation. A starting point could be to determine which region, the NTD or the CTD of TAF15, is implicated in the functional repression of translation.


a
Normalized RLuc/Fluc activity
2.0
1.5
1.0
0.5
0.0
****
****
MS2-CP
MS2-FUS
MS2-TAF15
n=5
b
Relative levels of RLuc mRNA (fold changes relative to the MS2-CP condition)
2.5
2.0
1.5
1.0
0.5
0.0
ns
ns
MS2-CP
MS2-FUS
MS2-TAF15
n=5

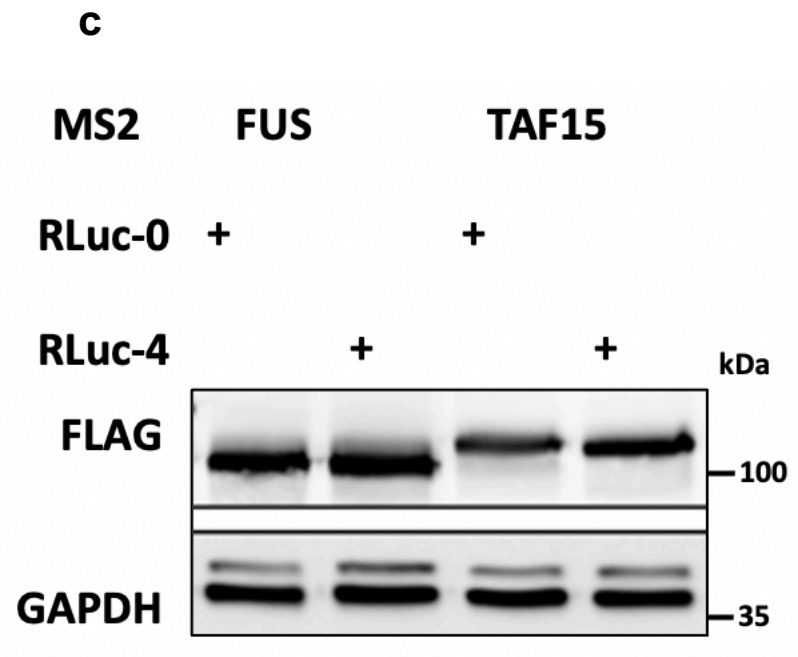

c
MS2
FUS
TAF15
RLuc-0 +
+
RLuc-4
+
+
kDa
FLAG
100
GAPDH
35

Figure 86: Tethering of FUS or TAF15 repress translation of the *Renilla luciferase* mRNA reporter. (**a, b**) Luciferase tethering assays in HeLa cells transfected with MS2-CP or MS2-CP fused FUS or TAF15 constructs together with *RLuc-4* or *RLuc-0* reporters. Results are means of normalized *RLuc/FLuc* activities [(*RLuc-4/FLuc*)/(*RLuc-0/FLuc*)] (**a**) and *RLuc* mRNA levels, relative to the MS2-CP condition (**b**) ± s.d. (n = 5 independent experiments). ****$P < 0.0001$, ns = not significant compared to the MS2-CP condition by one sample *t* test. (**c**) Western blot analysis of the levels of MS2-FUS and MS2-TAF15 with anti-FLAG antibody. GAPDH is used as a loading control. Samples are total cell lysates from HeLa cells transfected with MS2-FUS or MS2-TAF15 together with the *RLuc-0* or *RLuc-4* reporters constructs.

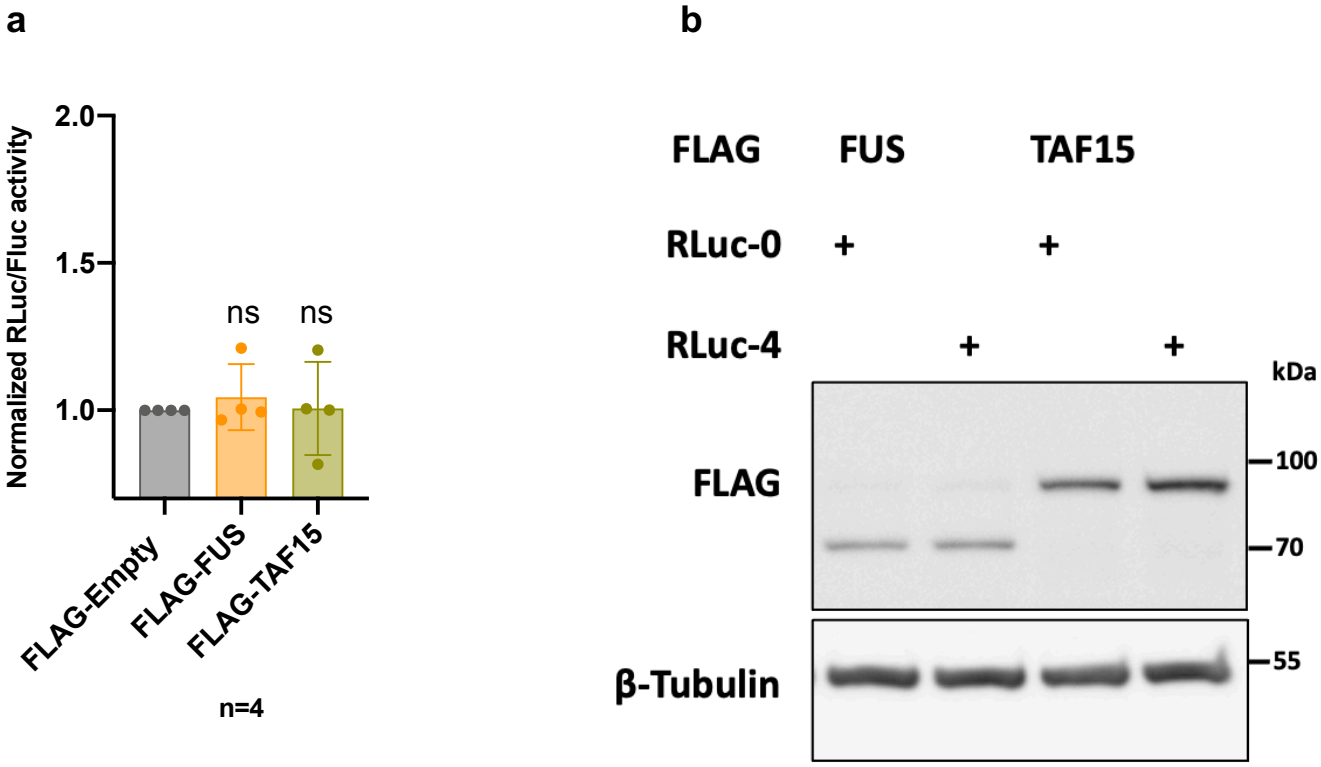


Figure 87: Non-tethered FLAG-FUS or FLAG-TAF15 do not repress the translation. (**a**) Luciferase tethering assays in HeLa cells transfected with a control empty vector (FLAG-Empty) or a vector expressing FLAG-tagged FUS or TAF15 constructs together with *RLuc-4* or *RLuc-0* reporters. Results are means of normalized *RLuc/FLuc* activities [(*RLuc-4/FLuc*)/(*RLuc-0/FLuc*)] ± s.d. (n = 4 independent experiments). ns = not significant compared to the FLAG-Empty condition by one sample *t* test. (**b**) Western blot analysis of the levels of FLAG-FUS and FLAG-TAF15 proteins using an anti-FLAG antibody. β-Tubulin is used as a loading control. Samples are total cell lysates from HeLa cells transfected with FLAG-FUS or FLAG-TAF15 together with *RLuc-0* or *RLuc-4* reporters constructs.

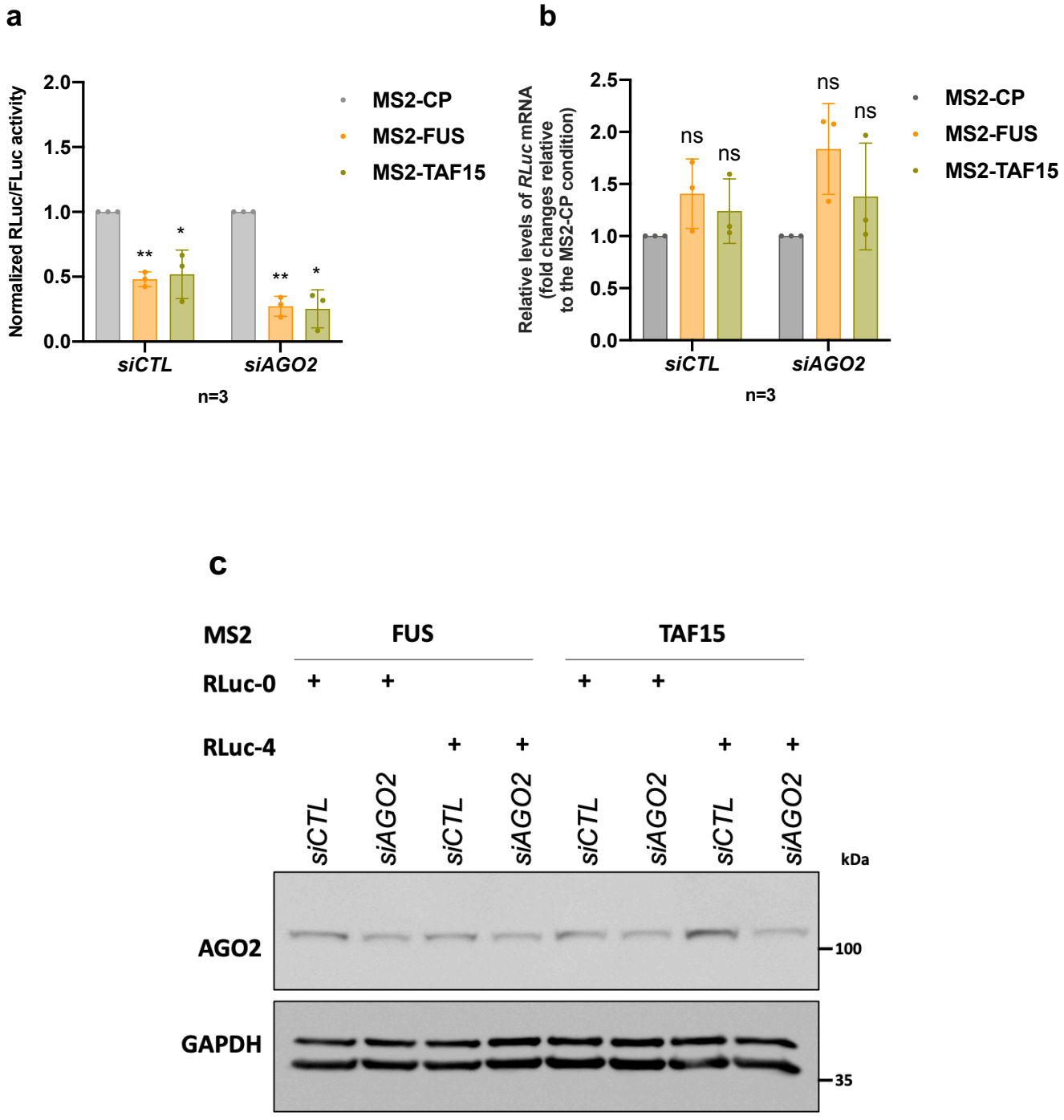


Figure 88: Repression of translation by FUS or TAF15 does not rely on the miRNA machinery. (**a, b**) Luciferase tethering assays in HeLa cells transfected with MS2-CP or MS2-CP fused FUS or TAF15 constructs together with *RLuc-4* or *RLuc-0* reporters, in *siCTL* or *siAGO2* conditions. Results are means of normalized *RLuc/FLuc* activities [(*RLuc-4/FLuc*)/(*RLuc-0/FLuc*)] (**b**) and *RLuc* mRNA levels, relative to the MS2-CP condition (**c**) ± s.d. (n = 3 independent experiments). *$P < 0.05$, **$P < 0.01$, ns = not significant compared to the MS2-CP condition by one sample *t* test. (**c**) Western blot analysis of the levels of AGO2. GAPDH is used as a loading control. Samples are total cell lysates from HeLa cells transfected with MS2-FUS or MS2-TAF15 together with the *RLuc-0* or *RLuc-4* reporters constructs, in *siCTL* or *siAGO2* conditions.

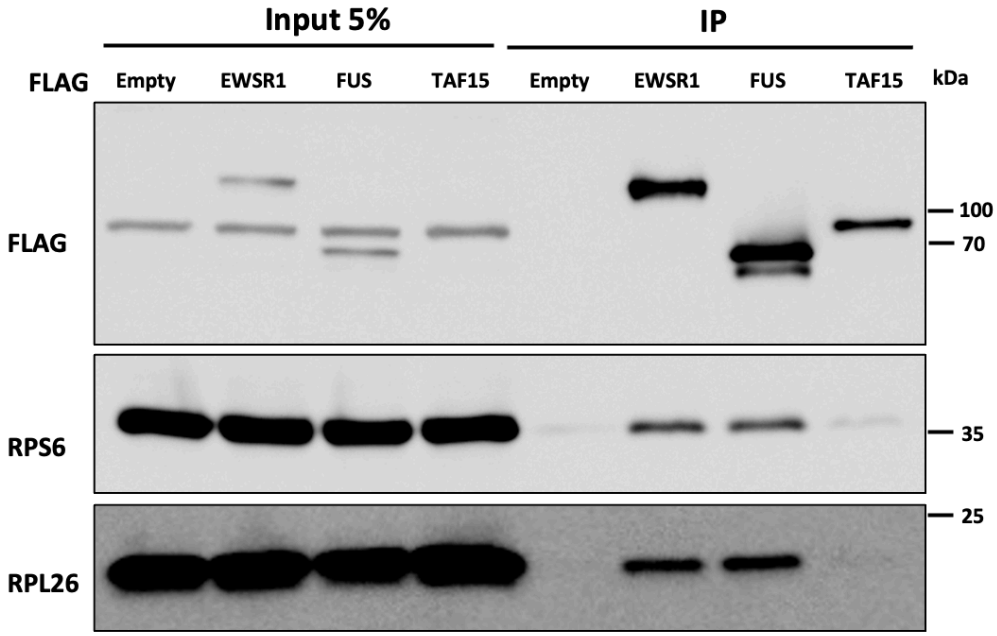


Figure 89: FET proteins bind to ribosomal subunits. Immunoprecipitation (IP) of FLAG-tagged EWSR1, FUS and TAF15 and anti-FLAG, anti-RPS6 and anti-RPL26 western blotting. Samples are lysates from HeLa cells transfected with FLAG-EWSR1, FLAG-FUS, FLAG-TAF15 or control FLAG Empty-vector.

4.4 Specific features of EWSR1 translational target mRNAs

To date, studies have suggested that EWSR1 binding is complex and have yet to provide a common model regarding the nature of the RNA sequences and structural features bound by EWSR1[9]. Interestingly, in this study, we found that several characteristics of EWSR1 translational targets are compatible with the sequences that EWSR1 is supposed to bind. Indeed, we found that EWSR1 TE up targets are enriched with specific structural and sequence features, such as high GC- and G-content or high incidence of rG4 structures (**Figure 54**), which correlates with previous data demonstrating that EWSR1, specifically *via* its RGG3, binds to and stabilizes G-quadruplex with structure specificity[49]. Of note, it has been reported that the high GC content causes repression of translation[594], and that G-rich sequences facilitate the formation of RNA G-quadruplex region (rG4) structures[657], which are also implicated in translational inhibition[658]. Overall, these findings suggest that EWSR1 might play the role of an rG4-stabilizing protein by binding to folded rG4 structures or by facilitating their formation by binding to G-rich sequences, in order to inhibit the translation of its targets. Therefore, EWSR1 might create a favorable RNA structure to enhance repression of translation. Additionally, our bioinformatic analysis indicated enrichment of EWSR1 RBNS (**Figure 55e**) and FUS bipartite motifs (**Figure 55d**) as two potential RNA-recognition elements for EWSR1 in the context of translation inhibition, which might add another layer of EWSR1 binding selectivity to RNA.

Although past studies mainly focused on the action of individual RBPs on their RNA targets to control gene expression, we are now observing an increasing amount of evidence describing the regulatory interplay phenomenon occurring between RBPs[596]. Strikingly, our RBP enrichment analysis revealed that EWSR1 might collaborate with other RBPs reported to be implicated in the repression of translation (**Figures 55a-c**)**.** Therefore, EWSR1 might also serve as a binding platform for RBPs that reinforce EWSR1-mediated repression of translation.

Importantly, our MS2-tethering assay (**Figure 57**) and RIP-qPCR results (**Figure 56c**) pointed towards a model whereby EWSR1 binding to its RNA targets is crucial for the functional repression of translation. Therefore, it will be interesting to first validate this hypothesis and second to test whether RNA binding and translational repression by EWSR1 are inseparable functions, or in contrast are conducted by different regions. For instance, we can choose some RNA from the “direct EWSR1 translational targets” list and evaluate which domain or combination of domains of EWSR1 bind to RNA, by performing electrophoretic mobility shift assays (EMSAs or band shift assays) using *in vitro* transcribed RNA as a substrate and the expressed and purified domain of EWSR1, as described in[659,660]. These experiments should be followed by tethering assay to test whether *in vitro* binding studies have implications for the repression of EWSR1 mRNA targets *in vivo*[660].
Additionally, we can test the affinity of the identified binding region to RNA possessing a range of sequences and structures, such as G-quadruplex, complex RNA, simple hairpains and single

strand RNA (ssRNA)[66]. This will help to first determine if this region displays a stronger preference to a specific RNA sequence, and second if a particular domain presents selective binding towards specific RNA sequences and structures.

Other features can be also looked at, such as Gibbs free energy values (dG) and Kozak sequence of 5'UTR, which can give more insights about EWSR1 translational targets.

As we found that some RBPs might interact with EWSR1 to repress translation (**Figures 55a-c**), it might be important to confirm the interaction between EWSR1 and these RBPs by performing, for instance, coimmunoprecipitations experiments. Furthermore, it might be worth to explore to which level these RBPs might affect the RNA binding, the translational function, and the association of EWSR1 with the 40S, especially that RBPs that bind to EWSR1 3'UTR are distinct from others binding to the 5'UTR (**Figures 55a,b**). The finding may unravel further features of selectivity and specificity about EWSR1 translational activity.

To test whether EWSR1 promotes the formation of rG4s, it will also be relevant to study rG4s in cells, for instance, by using G4-specific antibodies in combination with fluorophore tagged secondary antibodies, to evaluate whether these structures can be more detected in the presence of EWSR1[661].

Finally, as we found that EWSR1 and its CTD can repress mRNA translation when tethered to the 3'UTR of the *luciferase* mRNA reporter in our MS2-tethering assay (**Figures 57, 60**), it might be relevant to test other reporters with MS2 loops in their CDS or 5'UTRs, as described in[351], to examine whether EWSR1 binding position on mRNA is important for its ability to repress translation.

4.5 EWSR1 binds to ribosomes *via* its CTD

In this study, we found that EWSR1 binds to ribosomal subunits (**Figures 65, 66**). Of note, EWSR1 with FLAG-tag placed in its C-terminal was also able to interact with the RPS6, confirming that the tag does not interfere with EWSR1 association with ribosomal subunits (**Figure 90**).

Particularly, we identified the RGG2-ZnF-RGG3 as the sub-region of the CTD to be involved in the association with the 40S, as ΔRGG2-ZnF-RGG3 mutant totally lost the association with 40S ribosomal proteins, such as RPS6 and RPS23 (**Figure 67**). This is consistent with previous reports about the binding of several RBPs to ribosomes through their RGG motifs[29,662,663]. Additionally, the ZnF can play a key role in the association with ribosomes[664]. The observation that ΔZnF, ΔRGG2, ΔRGG3 mutants maintain association with the 40S, while ΔZnF-RGG3 and ΔRGG2-ZnF presented reduced association, especially with the RPS23 (**Figure 67a**), suggests

that each domain of the region RGG2-ZnF-RGG3 can contribute to ribosomal binding, without being by itself sufficient for this binding.

Remarkably, we found that ΔZnF-RGG3 and ΔRGG2-ZnF mutants presented a reduced interaction with RPS23 compared to the RPS6 (**Figure 67a**). This suggests that the RGG2-ZnF-RGG3 might bind selectively to different ribosomal proteins of the 40S.

An important perspective is to study protein-protein interaction networks of EWSR1, using protein complementation assays, such as gPCA or NanoLuc two-hybrid[665] to determine to which proteins of the 40S subunit can EWSR1 bind directly and specifically. Another interesting perspective is to determine the three-dimensional (3D) binding position of EWSR1 on the ribosome, by cryo-EM. We performed cryo-EM of the 80S•GST-CTD complex, but it was not successful, maybe due to the destabilization of the 80S by the CTD (**see below; "4.7 Possible EWSR1 mechanism in the repression of translation"**). Therefore, it might be relevant to test the 40S•GST-CTD complex rather than the 80S•GST-CTD complex.

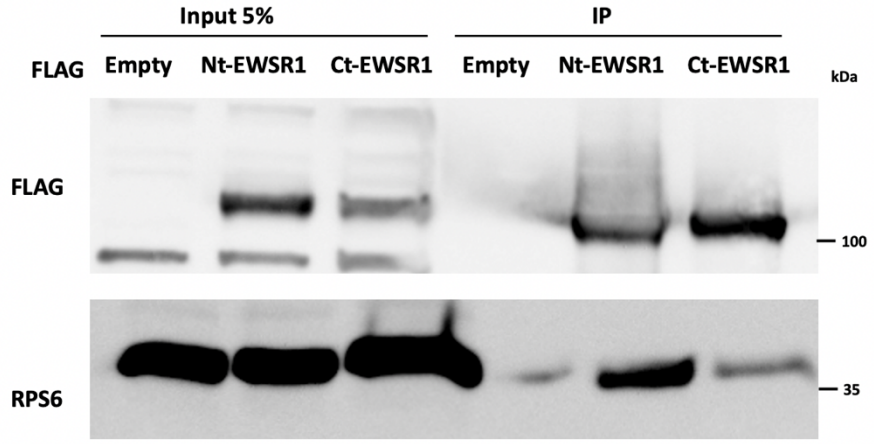


Figure 90: Ct-EWSR1 binds to ribosomal proteins. Immunoprecipitation (IP) of FLAG-tagged EWSR1 with FLAG-tag in the N-terminal region (Nt-EWSR1) and FLAG-tagged EWSR1 with FLAG-tag in the C-terminal region (Ct-EWSR1) and anti-FLAG and anti-RPS6 western blotting. Samples are lysates from HeLa cells transfected with FLAG-Nt-EWSR1, FLAG-Ct-EWSR1 or control FLAG Empty-vector.

4.6 Deregulation of EWSR1 translational activity might be implicated in several diseases

In this study, we uncovered a new function of the CTD in the repression of mRNA translation (**Figures 60, 61, 65, 66, 67, 69, 70**), besides its published functions in diverse transcriptional and post-transcriptional processes[666]. Importantly, it is well known that mutations in the CTD, particularly in RGG boxes causes diseases, for instance, neurological diseases[119], which may

raise the question about the implication of translational deregulation, caused by CTD mutations, in the development of these diseases.

In addition, to further highlight the importance of the CTD in the repression of translation, we evaluated the translational activity of the fusion protein EWSR1-FLI1 that lost the CTD (**see Introduction; "1.1.6 EWSR1 and sarcomas"**). As expected, we found that tethering of EWSR1-FLI1 did not present an effect on the expression of the *RLuc* protein, nor on the abundance of its transcript (**Figure 91**). Furthermore, EWSR1-FLI1 showed no interaction with ribosomal proteins, such as RPS6 and RPL26 (**Figure 92**), and it was only present in the "free mRNA pool", as it was absent from 40S, 60S, 80S and polysomes fractions (**Figure 93**). Primarily, these results strengthen our notion that both the functional repression of translation and association with ribosomes are ascribable to the CTD of EWSR1. Importantly, as most studies about Ewing sarcoma have focused on the function of the *EWSR1*-fusion genes, with only little understanding of the role that loss of an *EWSR1* allele has on pathogenesis; these findings might unveil another aspect of the contribution of EWSR1 haploinsufficiency towards the pathogenesis of *EWSR1*-fusion associated sarcomas[82,217]. Therefore, it is tempting to speculate that the loss or reduced function of EWSR1, herein in the repression of translation, under *EWSR1*-fusion associated sarcomas might contribute to Ewing sarcomagenesis. Particularly, we clearly demonstrated that impaired translational function of EWSR1 is associated with deleterious effects on cells, such as aberrant lipid homeostasis (**Figure 75**), perturbated ER function and shape (**Figure 77**) and impairment of the UPR^{ER} (**Figures 78-82**), suggesting that these biological processes might be highly perturbated in the context of Ewing sarcoma. Overall, both the fusion protein EWSR1-FLI1 and the haploinsufficiency of EWSR1 could drive tumorigenesis by deregulating transcriptional and post-transcriptional programs. Therefore, improved understanding of the oncogenic mechanisms employed by both EWSR1-FLI1 and EWSR1 haploinsufficiency to disturb normal cellular programming will uncover potential novel approaches to pharmacologically block development of the disease. This concept might also be broadened to other *FET*-fusion proteins.

Taken together, EWSR1 is primordial for the regulation of transcriptional and post-transcriptional mechanisms, which underscores its implication in diverse roles in physiological cellular functions. However, EWSR1-related disorders, such as CTD mutations and *EWSR1*-fusion genes, can cause perturbations of cellular functions and lead to various diseases, including cancer and neurodegenerative diseases (**Figure 94**).

Given the importance of mRNA translation in rapidly defining cellular proteome in a spatiotemporal manner to ensure cell homeostasis, it is not surprising that any translational deregulation may lead to disease. Indeed, growing evidence suggests that dysregulation in mRNA translation plays an important role in the pathogenesis of several diseases[667], such as

neurodevelopmental diseases and neurodegenerative disorders (*e.g.*, autism spectrum disorder (ASD) and fragile X syndrome (FXS), Alzheimer's disease (AD))[668].

Therefore, it will be interesting to investigate potential deregulations of the translatome in the context of EWSR1-related disorders, which will help in conceiving new therapeutic strategies.

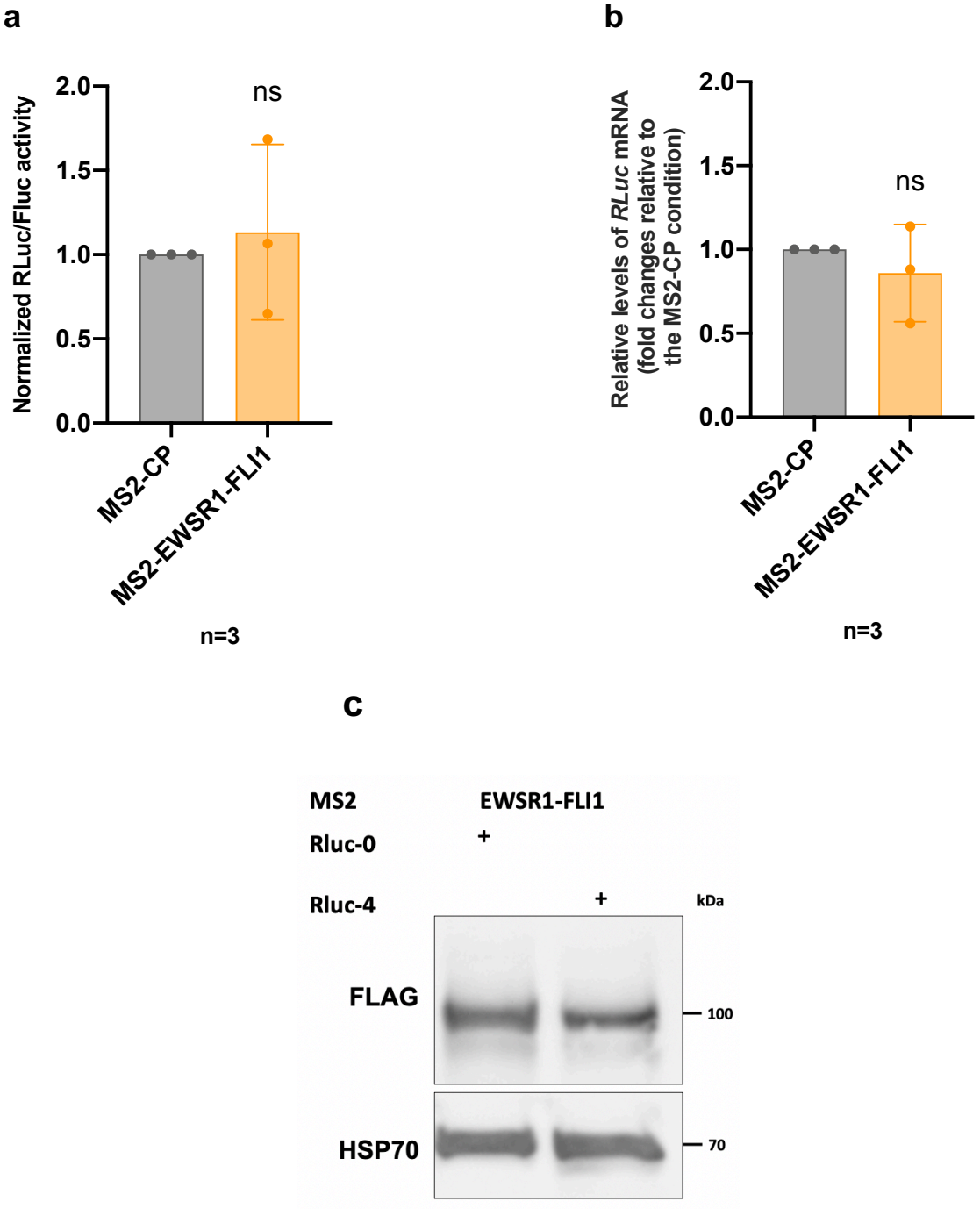


Figure 91: Tethering of EWSR1-FLI1 has no effect on the translation of the *Renilla luciferase* mRNA reporter. (**a, b**) Luciferase tethering assays in HeLa cells transfected with MS2-CP or MS2-CP fused EWSR1-FLI1 constructs together with *RLuc-4* or *RLuc-0* reporters. Results are means of normalized *RLuc/FLuc* activities [(*RLuc-4/FLuc*)/(*RLuc-0/FLuc*)] (**a**) and *RLuc* mRNA levels, relative to the MS2-CP condition (**b**) ± s.d. (n = 3 independent experiments). ns = not significant compared to the MS2-CP condition by one sample *t* test. (**c**) Western blot analysis of the levels of MS2-EWSR1-FLI1 with anti-FLAG antibody. HSP70 is used as a loading control. Samples are total cell lysates from HeLa cells transfected with MS2-EWSR1-FLI1 together with the *RLuc-0* or *RLuc-4* reporters constructs.

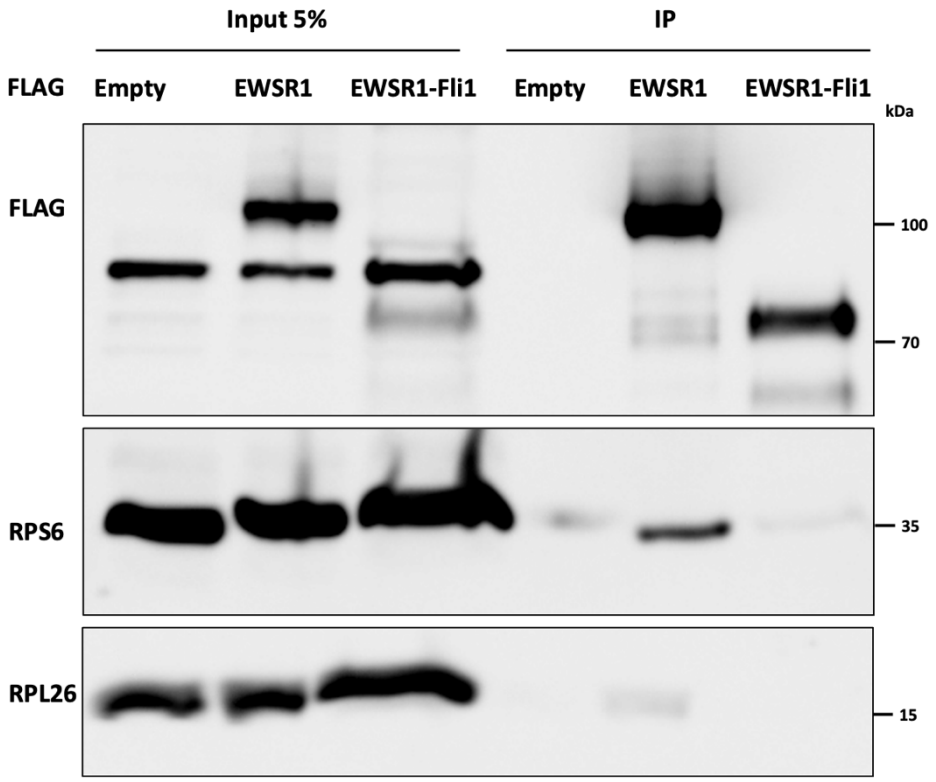


Figure 92: EWSR1-FLI1 does not bind to ribosomal proteins. Immunoprecipitation (IP) of FLAG-tagged EWSR1 and EWSR1-FLI1 and anti-FLAG and anti-RPS6 western blotting. Samples are lysates from HeLa cells transfected with FLAG-EWSR1, FLAG-EWSR1-FLI1 or control FLAG Empty-vector.

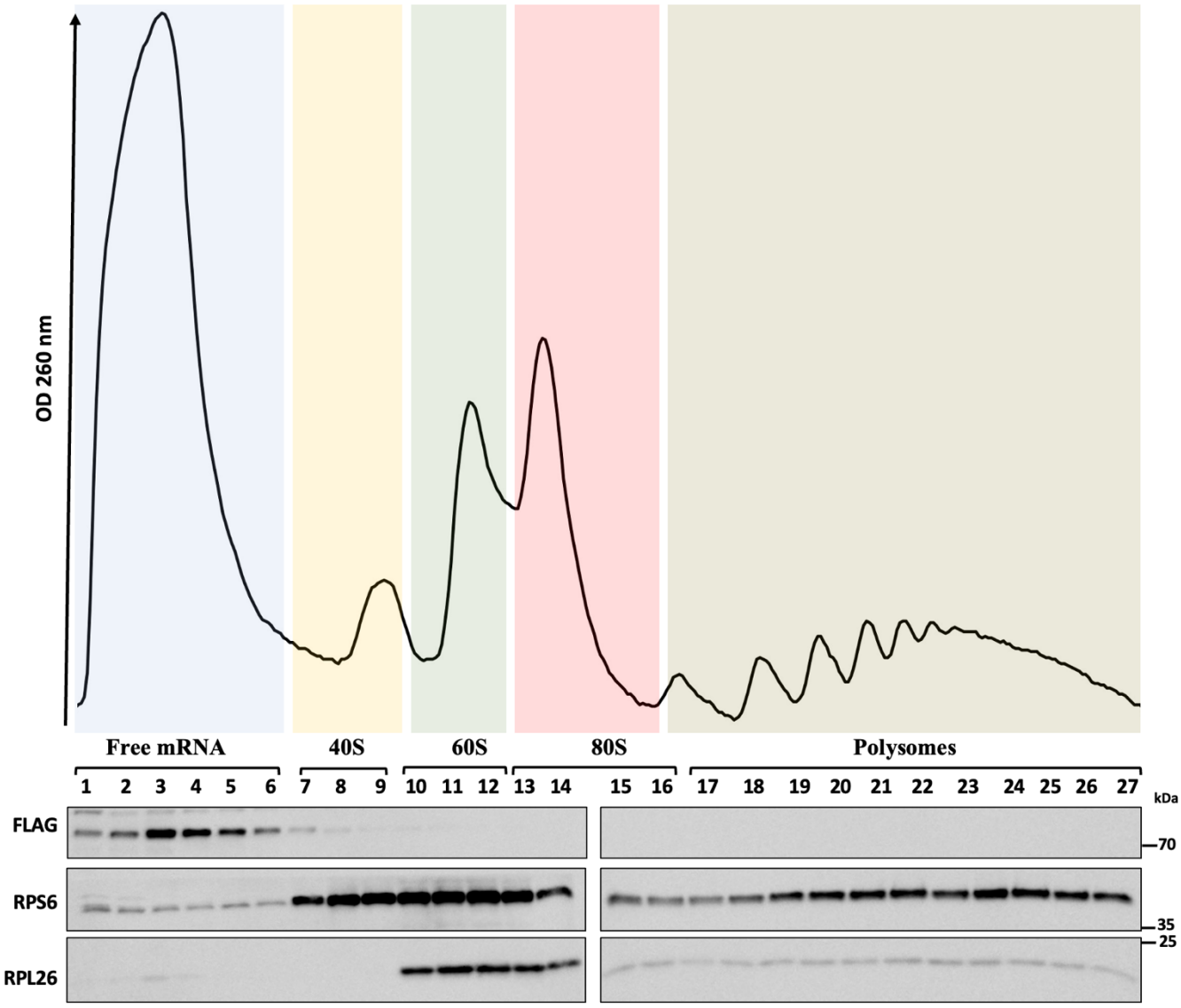


Figure 93: EWSR1-FLI1 is not present in ribosomal fractions. Absorbance profiles at 260 nm (above) and western blot analysis (below) of sucrose gradient sedimentation of HeLa cell cytoplasmic extracts. Fractions 1 to 6 corresponds to “free mRNA pool” not associated to ribosomes, fractions 7 to 9 correspond to isolated 40S particles, fractions 10 to 12 contain free 60S, fractions 13 to 16 correspond to 80S and fractions 17 to 27 correspond to polysomes. The distributions of FLAG-EWSR1-FLI1, RPS6 and RPL26 throughout the different fractions of the sucrose gradient were analyzed by western blotting.

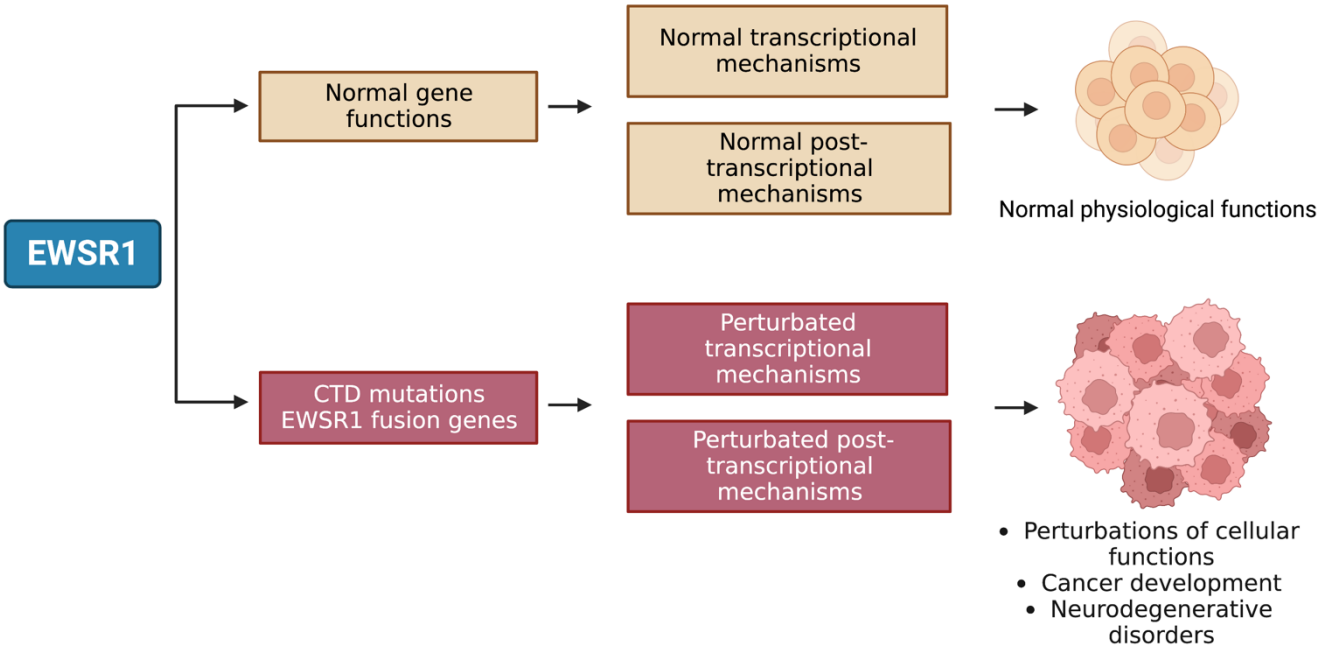


Figure 94: Outcome of the normal or perturbated transcriptional and post-transcriptional mechanisms mediated by EWSR1. Created with BioRender.com.

4.7 Possible EWSR1 mechanism in the repression of translation

We recapitulated the main steps of regulation of translation in **Figure 95**, which in coordination with our results, will help to reason a possible mechanism of EWSR1 in the inhibition of translation.

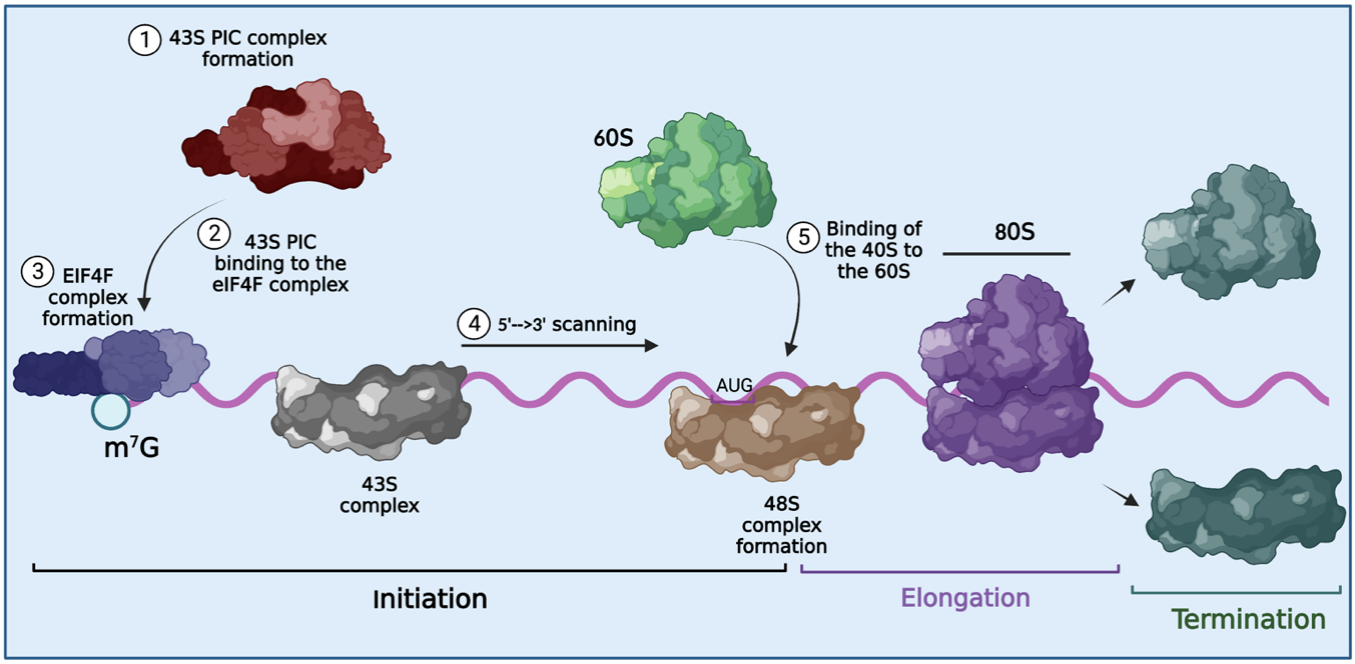


Figure 95: Illustration of main steps of mRNA translation regulation. Created with BioRender.com. PIC: pre-initiation complex.

First, we found that EWSR1 is implicated in the repression of translation of a specific subset of mRNA, through its binding onto its RNA targets (**Figures 51-60**). This indicates that the binding of EWSR1 to its RNA targets is necessary for the functional repression of translation. Next, we dissected the RNA-binding domains of EWSR1 to determine the essential domains for inhibiting translation and we found that the RGG2-ZnF-RGG3 region of the CTD, is responsible for the functional inhibition of translation (**Figure 61c**). Additionally, we demonstrated, by the series of binding experiments (*e.g.*, pulldowns, co-IPs) that EWSR1 is able to associate with the 40S ribosomal particle, remarkably in an RNase-insensitive manner (**Figures 65, 66b,c**). This implies that EWSR1 associates with the 40S to inhibit the translation. Importantly, we found again that the RGG2-ZnF-RGG3 was the 40S-interacting region (**Figure 67**). So far, these data point towards a model whereby association of EWSR1 with both the 40S and mRNA may be important for synergistically inhibiting translation.

Our results also indicate that EWSR1 inhibit the translation downstream of initiation for the several following reasons. First, our m^7GTP pulldown showed that EWSR1 does not associate

with the cap or any initiation factors, nor it affects cap-recognition by the 43S PIC (**Figure 68**). Therefore, EWSR1 does not affect steps 1, 2 or 3 of translation initiation in **Figure 95**. Second, TISU reporter assays indicated that inhibition of translation by EWSR1 is independent of 5'UTR scanning by the 43S PIC (**Figure 70**), ruling out the regulation of step 4 in **Figure 95**. Third, IRESs reporters, especially CrPv IRES, revealed that EWSR1 can also repress cap-independent translation without the need of any initiation factor (**Figure 69**). This result suggests that EWSR1 acts downstream of translation initiation and might affect translation elongation or termination. Particularly, the observations that i) EWSR1 is present in 40S fraction but mainly absent from 60S/80S and polysomes fractions (**Figure 63c**) and ii) that it is pulled down with the 40S in the absence of any 60S (**Figure 66e**) strongly suggest that EWSR1 might inhibit a very early step of elongation, most likely through its binding to the 40S and preventing the joining of the 60S (**step 5 in Figure 95**) and (**Figure 96a**). Alternatively, EWSR1 might be able to dissociate elongating 80S ribosomes, which might explain its absence from the 80S and polysomes fractions in our polysome profiling experiment (**Figure 96b**).

Collectively, our results so far clearly establish a role for EWSR1 in repressing mRNA translation downstream of initiation and through a mechanism implicating its association with both the 40S ribosomal subunit and its mRNA target (**Figure 96**).

Further analyses in cell-based and -free systems will more precisely define the mechanism of EWSR1 function in the repression of translation. Currently, we are performing an *in vitro* translation experiment, in collaboration with the CIRI, Centre International de Recherche en Infectiologie, (Team Ohlmann), Lyon. The system consists of using specific inhibitors of various stages of protein synthesis, such as puromycin and MDMP, as described in[669], to first confirm our previous results that EWSR1 does not affect the initiation of translation, and second to investigate the mechanism of translational repression by EWSR1. Another alternative is to perform toeprinting assay[670].

To assess whether EWSR1 blocks the recruitment of the 60S, an additional precipitation of biotinylated mRNA assay can be performed, upon completion of the translation repression reactions, wherein reporter mRNAs can be precipitated with streptavidin beads. Then, precipitates can be subjected to Northern blot analysis for ribosomal RNAs, to calculate the ratio of 60S rRNA:40S rRNA and therefore we can evaluate whether the presence of EWSR1 might lead to reduced 60S ribosome subunit loading on target mRNAs[671].
We can also test whether EWSR1 can dissociate the 80S by conducting two interesting approaches. First, by checking whether the interaction between BiFC-tagged RPs which are in different subunits, such as RPS18-YN and RPL11-YC, might be diminished in the presence of EWSR1[672]. Second, by testing whether EWSR1 can lead to ribosome dissociation in a "ribosome dissociation assay", as described in[673].

Interestingly, from our RIBO-seq data we can also retrieve many fundamental insights into the determinants of translation elongation kinetics (**see Introduction; "1.2.6.4 Ribosome profiling (RIBO-seq)"**). Currently, we are applying several pipelines to study whether EWSR1 might cause stalling/pausing of ribosomes[674–676]. This is highly important as these determinants directly coordinate downstream processes on the ribosome, such as co-translational protein folding and degradation and interaction of nascent chains with chaperones or ribosome-associated targeting factors[677].

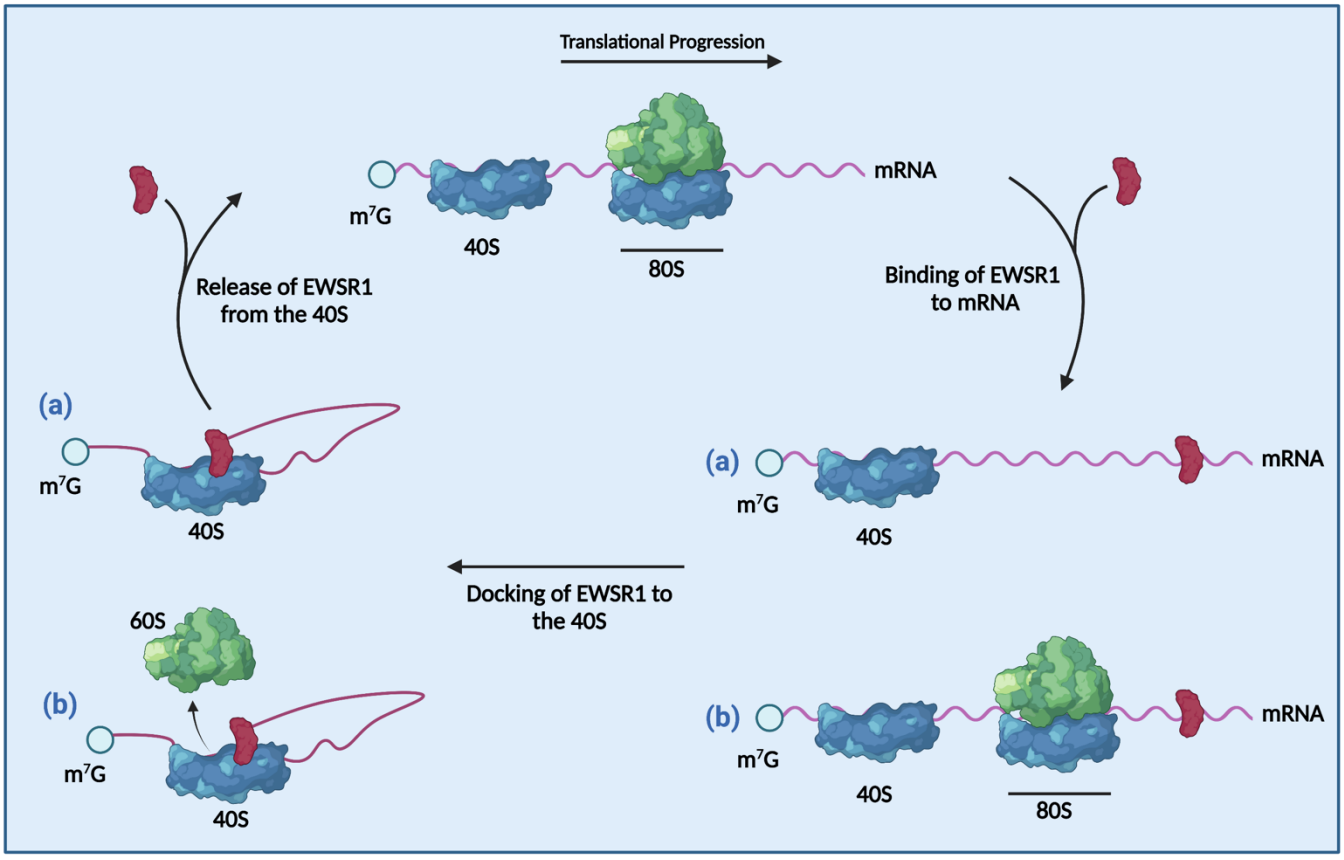


Figure 96: Model illustrating the new molecular function of EWSR1 in the repression of translation identified in this work. Created with BioRender.com**.** EWSR1 (red) may bind to mRNA presenting specific structural and sequence features (*e.g.,* high GC- and G-contents and high rG4 structures, EWSR1 RBNS and/or FUS bipartite motifs). Then, it docks on the 40S subunit (blue) and might either prevent the joining of the 60S (**a**) or dissociate elongating 80S ribosomes (**b**). Binding of EWSR1 to the mRNA and the ribosome synergistically inhibits translation, whereas its releasing from the 40S•mRNA complex activates the translation. m^7GTP: 7-methylguanylate structure.

4.8 EWSR1 localization and activity under stress conditions

We found that EWSR1 is present in the nucleus and the cytoplasm, which was consistent with previous reports[9] (**Figure 62**). A remaining question is whether this localization can be affected by stress conditions, as we found that EWSR1 is associated with activation of the UPR^{ER} under

stress conditions (*e.g.*, ER-stress inducing conditions) (**Figure 83**). Interestingly, EWSR1 has been known to be targeted to stress granules (SGs)[656].

First, it will be relevant to test the presence of EWSR1 in SGs under stress conditions and evaluate whether it might also affect SGs dynamics. Second, as we found that EWSR1 is present in the "free mRNA pool" and 40S fraction of the polysome profiling (**Figure 63c**), an interesting perspective is to check whether this distribution can also be affected. Finally, it might be worth to explore global protein synthesis in the presence/absence of EWSR1 under stress conditions using the SUnSET assay.

4.9 EWSR1 and PTMs

Post-translational modifications (PTMs) are covalent processing events changing protein properties by proteolytic cleavage and adding a modifying group, like acetyl, phosphoryl, glycosyl and methyl, to one or more amino acids. Therefore, they diversify and enlarge protein function beyond what is dictated by gene transcripts. There are more than 400 different types of PTMs allowing eukaryotic cells to dynamically regulate their signal integration and physiological states. As PTMs are involved in regulating almost all cellular events, disruption perturbates vital biological processes and leads to various diseases[678].

Interestingly, several PTMs affecting EWSR1 activity have been reported. For instance, EWSR1 phosphorylation status affects importantly its biological functions. Indeed, EWSR1 RNA binding activity is regulated by its phosphorylation within IQ domain by the PKC[54]. Moreover, the tyrosine of EWSR1 can be phosphorylated i) by c-Abl modulating EWSR1 self-association and DNA binding activity[679], or ii) by v-Src enhancing its transactivation ability[680]. EWSR1 is also phosphorylated by cytoplasmic protein tyrosine kinase, BTK, in cell cycle dependent manner[681], and at Thr(79) in response to either mitogens or DNA-damaging agents[79]. In addition, EWSR1 can undergo glycosylation at unknown sites[682]. Recently, a study revealed that UV-induced EWSR1 acetylation at K423, K432, K438, K640, and K643 acetylation sites is essential for its function in DNA damage response[683]. The list of EWSR1 PTMs also includes the asymmetric dimethylation of RGG boxes arginine residues by protein arginine methyltransferases (PRMTs), such as PRMT1[684], PRMT3[49] and PRMT8[12]. Of note, arginine methylation has been shown to affect protein-RNA interactions, as it can either sterically hinder binding with RNA or eliminate hydrogens that might participate in bonds with the RNA[685].

As PRMT1 recognizes most methylation sites of EWSR1 potential methylation sites and affects EWSR1 localization[11], we tested whether EWSR1 methylation by PRMT1 affects its translational repression activity. We found that PRMT1 KD did not impair the translational activity of the CTD, suggesting that this methylation has no effect on CTD activity in the context of translation (**Figure 97**). However, these results can also be explained by either our inability to achieve complete

loss of CTD methylation using siRNA against PRMT1, or the possibility that other PRMTs methylate EWSR1, such as PRMT3[49]. Furthermore, our preliminary results showed that FET proteins interacts with PRMT1 (**Figure 98**), which agreed with previous reports[684,686,687]. We also found that this interaction is mediated by the CTD of EWSR1 and is RNA-independent (**Figure 99**). To test whether EWSR1 binding to the 40S was affected by PRMT1, we co-immunoprecipitated endogenous EWSR1 with RPS6 and RPL26 in control and PRMT1 depleted conditions. We found that PRMT1 KD did not impair the binding of EWSR1 to ribosomal proteins (RPL26, RPS6) (**Figure 100**). However, we were not able to confirm the interaction between EWSR1 and PRMT1 at the endogenous level, therefore, it is crucial to repeat and optimize the conditions of this experiment to confirm the overexpression results.

An interesting perspective is to evaluate whether PRMT3 can affect the translational regulation by EWSR1, especially that this PRMT can affect EWSR1 binding to RNA. Indeed, it was shown that PRMT3-mediated methylation of specific arginine between amino acids 589 and 597 within the RGG3 is important for RGG3 binding to G-quadruplex DNA and RNA[49]. It might also be interesting to determine whether other PTMs, such as glycosylation and acetylation, can affect the translational activity, RNA and ribosomal binding of EWSR1 in the context of translation.

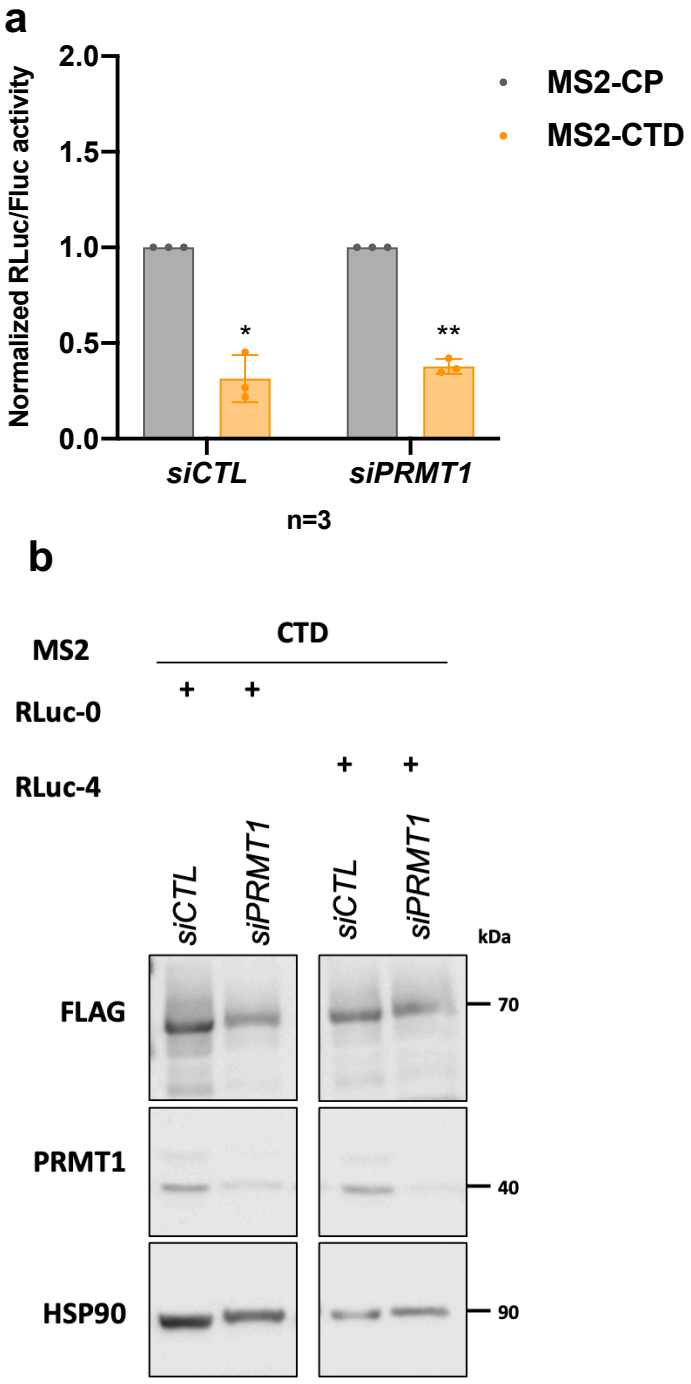


Figure 97: Repression of translation by CTD does not rely on PRMT1. (**a**) Luciferase tethering assays in HeLa cells transfected with MS2-CP or MS2-CP fused CTD constructs together with *RLuc-4* or *RLuc-0* reporters, in *siCTL* or *siPRMT1* conditions. Results are means of normalized *RLuc/FLuc* activities [(*RLuc-4/FLuc*)/(*RLuc-0/FLuc*)] ± s.d. (n = 3 independent experiments). **P* < 0.05, ***P* < 0.01 compared to the MS2-CP condition by one sample *t* test. (**b**) Western blot analysis of the levels of MS2-CTD with anti-FLAG antibody and PRMT1. HSP90 is used as a loading control. Samples are total cell lysates from HeLa cells transfected with MS2-CTD together with the *RLuc-0* or *RLuc-4* reporters constructs, in *siCTL* or *siPRMT1* conditions.

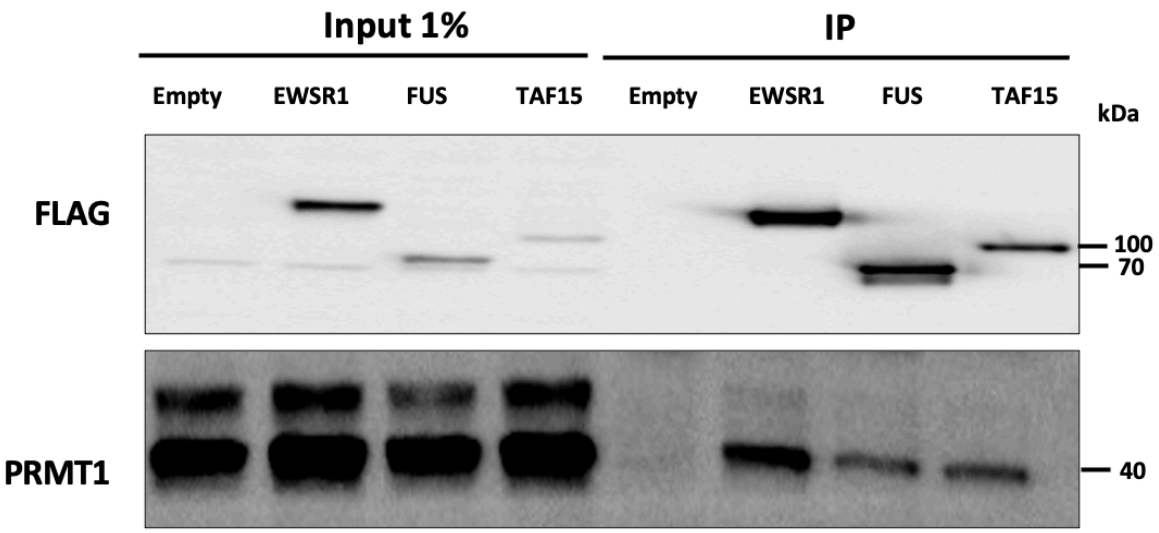


Figure 98: FET proteins bind to PRMT1. Immunoprecipitation (IP) of FLAG-tagged EWSR1, FUS and TAF15 and anti-FLAG and anti-PRMT1 western blotting. Samples are lysates from HeLa cells transfected with FLAG-EWSR1, FLAG-FUS, FLAG-TAF15 or control FLAG Empty-vector.

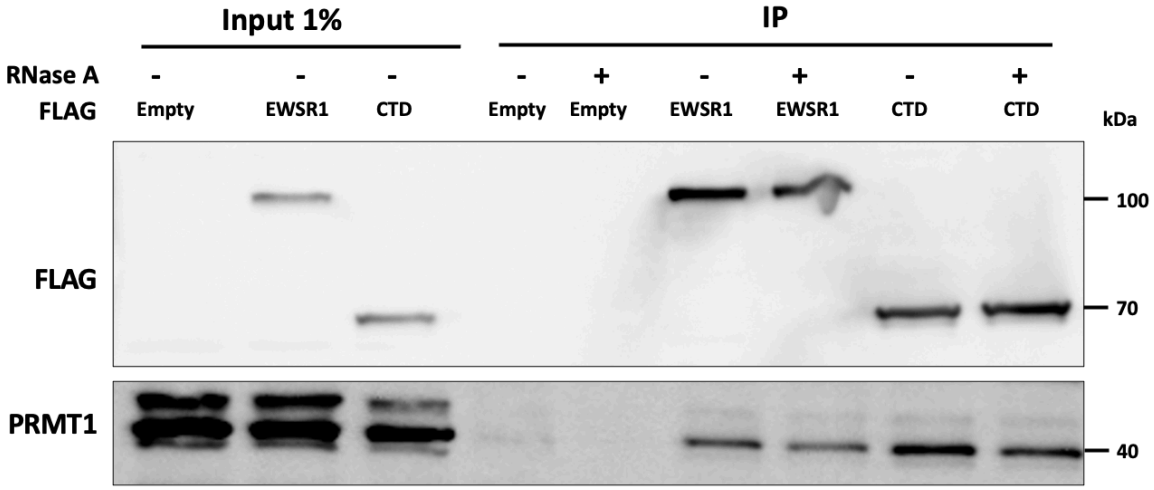


Figure 99: EWSR1 and CTD bind to PRMT1 through RNase-insensitive interaction. Immunoprecipitation (IP) of FLAG-tagged EWSR1 and CTD and anti-FLAG and anti-PRMT1 western blotting. Samples are nontreated (-) or RNase A treated (+) lysates from HeLa cells transfected with FLAG-EWSR1, FLAG-CTD or control FLAG Empty-vector.

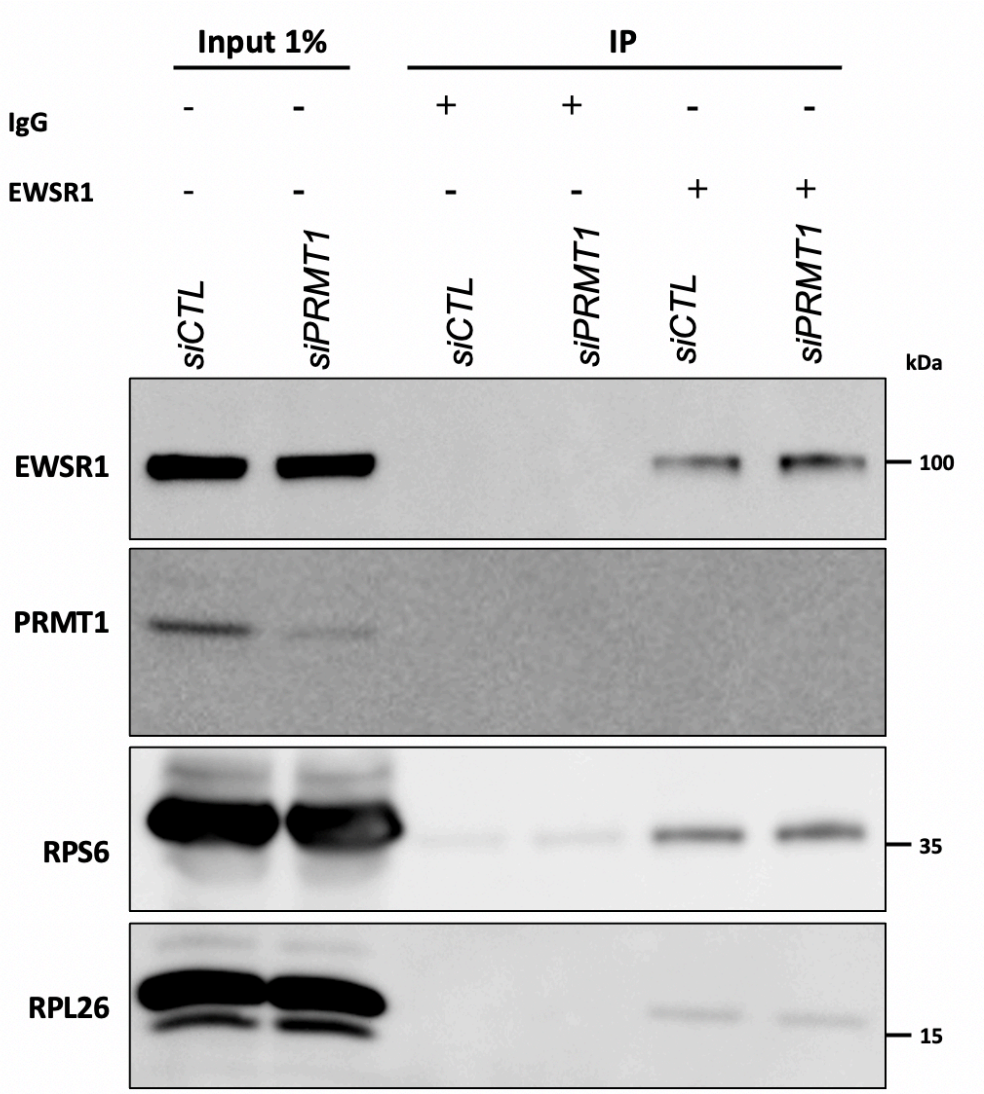


Figure 100: PRMT1 does not affect EWSR1 binding to the 40S. Immunoprecipitation (IP) of endogenous EWSR1 from HeLa cell lysates followed by western blotting for the indicated proteins.

4.10 EWSR1 and cellular homeostasis

In this study, we found that EWSR1-mediated translational regulation is associated with the regulation of lipid homeostasis, ER function and shape and UPR^{ER} activation. Indeed, EWSR1 loss led to lipid accumulation (**Figure 75**), leading to dilated ER morphology (**Figure 77**). This indicates that EWSR1 is a key player in maintaining normal ER shape and function by regulating lipid homeostasis. While no link was previously described between EWSR1 and ER, only few studies associated EWSR1 to lipid metabolism, thus it is an emerging area of study. Indeed, EWSR1 was reported to be essential for early brown[124] and white[688] fat lineage determination, pointing to its importance for mammalian energy homeostasis and metabolism. It was also

identified as an anabolic signaling hub linked to hepatocellular carcinoma (HCC) pathogenesis[689]. Recently, *Dong et al.*[690] demonstrated that RP11-728F11.4 lncRNA increases cholesterol uptake by monocytes-derived macrophages (MΦs) and promotes atherosclerosis, by binding to the RNA recognition domain of EWSR1, which relieved EWSR1 repressive activity on FXYD6 (FXYD domain containing ion transport regulator 6).

It is now evident that UPR^{ER} signaling pathways are co-opted for the maintenance of basal cellular homeostasis and can be enmeshed with cellular physiology events, such as differentiation of B-cells[691], muscle differentiation[692] and viral infections[693], in more complex and subtle ways[502] (**see Introduction; "1.3.7.7 Roles of UPR^{ER} in physiological processes"**). Interestingly, we showed that EWSR1 sustained the activation of UPR^{ER} under basal conditions (**Figures 78-82**), advocating its crucial role in the maintenance of basal cellular homeostasis. This was consistent with previous studies that revealed essential functions for EWSR1 in cell physiology maintenance, such as pre-B cell development (**see Introduction; "1.1.4 Physiological roles of EWSR1"**). Additionally, we found that EWSR1 is necessary to activate UPR^{ER} under ER-inducing stress conditions, however its depletion led to the impairment of activation of the three UPR^{ER} branches, no matter the applied ER-stress (**Figure 83**). Therefore, we concluded that absence of EWSR1 contributed to the failure of activation of ER stress sensors. We propose that this failure can be due to the accumulation of lipids (especially of membrane lipids, such as PC and PE) upon EWSR1 KD, which may cause lipid bilayer stress (LBS) (**see Introduction; "Lipid-dependent regulation of the UPR^{ER}"**). Accordingly, LBS upon EWSR1 KD can lead to changes in ER membrane properties and eventually to impaired functions of ER-resident transmembrane sensor proteins (PERK, ATF6 and IRE1). Indeed, it was reported that perturbation of ER membrane phospholipids can affect UPR^{ER} sensors and/or regulators which may impair their stability and/or oligomerization, as described recently for the transmembrane protein, Sbh1[694]. Taken together, these findings point towards a model where accumulation of lipids upon EWSR1 KD might cause LBS, which is responsible for the perturbation of ER properties, leading to impairment of UPR^{ER} sensors.

A remaining question is to determine whether EWSR1 might directly regulate, transcriptionally or post-transcriptionally, the expression of the three UPR^{ER} sensors (PERK, ATF6 and IRE1) or whether their impairment of activation upon EWSR1 KD is only a consequence of LBS. In our results section, we showed that ATF4 expression was regulated through uORF-mediated regulation (**Figure 79a**), downstream PERK, and that its four classical targets (*HERPUD1*, *PSAT1*, *TRIB3*, *MTHFD2)* (**Figure 79b**) were downregulated transcriptionally upon EWSR1 KD. Interestingly, we also found that PERK is a TE up gene pointing to the implication of EWSR1 in the transcriptional and translational regulation of the PERK axis. Furthermore, *ATF6B* and *ATF6* were downregulated transcriptionally upon EWSR1 KD (**data not shown**), suggesting that EWSR1 affects the ATF6 axis at the transcriptional level. However, we detected no regulation of IRE1 transcriptionally or translationally. Taken together, this suggests that EWSR1 might

regulate the UPR^{ER} transcriptionally and translationally. However, this does not exclude an indirect effect of EWSR1 on the UPR^{ER} due to the accumulation of lipids upon EWSR1 KD, enabling UPR^{ER} sensors to properly dimerize and to trigger the UPR^{ER}. It is also important to note that there are also intense crosstalks between the three UPR^{ER} branches[695]. This may propose a model whereby EWSR1 by affecting one pathway of the UPR^{ER} can affect the others.

Notably, cells depleted for EWSR1 were more sensitive to PA treatment. This can be explained by the twofold origin of lipids in EWSR1 KD cells (accumulation of lipids upon EWSR1 KD (**Figure 75**), accompanied with increased *de novo* biosynthesis of saturated phospholipids induced by palmitic acid), which can lead to deleterious effects on cells[696] and therefore results in more sensitivity to lipotoxicity in the absence of EWSR1 (**Figure 84c**).

In addition to canonical ER stress pathways directed by the three UPR^{ER} sensors (PERK, ATF6 and IRE1α), it has become clear recently that many other pathways, known as "noncanonical ER stress programs", converge on the UPR^{ER} or specific elements of the UPR^{ER} and might influence cell fate independently of UPR^{ER}. Such pathways mainly include the integrated stress response (ISR) (**see Introduction; "1.2.3.1 Regulation of translation initiation"**), translocation of proteins into the ER, extracellular-signal-regulated kinase reactivation, ERAD and ER-phagy[697] (**see Introduction; "1.3.8 ERAD" and "1.3.9 ER-phagy"**). Interestingly, besides its role in the activation of UPR^{ER}, our preliminary results also showed the implication of EWSR1 in the activation of GCN2 axis of the ISR under both basal and GCN2-inducing stress conditions, such as amino acid and glutamine deprivation stresses[698,699] (**Figure 101**). Then, we examined the ability of HeLa cells to cope with GCN2-inducing stress in the absence of EWSR1. Remarkably, although the effect of glutamine deprivation on apoptosis was not statistically different in control and EWSR1 KD cells, we found that EWSR1 KD cells were significantly less sensitive to AAD treatment (**Figure 102**). Analysis of cleavage of PARP by western blotting confirmed these observations (**Figure 101**). We primarily hypothesized that less cell apoptosis under AAD stress and upon EWSR1 KD might be explained by the accumulation of lipids, especially of TGs, upon EWSR1 KD (**Figure 75**). In this regard, it was reported that lipid droplets (LDs) made of TGs and under stress conditions, such as nutrient deprivation, present a pro-survival role and promote β-oxidation of fatty acids as source of energy[700]. Therefore, these findings suggest that accumulated TGs, which may be deposited subsequently in LDs, could be providers of metabolic fuel under AAD stress upon EWSR1 KD, which constitutes a pro-survival response and explains less sensitivity of cells in AAD condition.

Together, our observations underpin the importance of EWSR1 in controlling lipid homeostasis, ER function and shape, and activation of the UPR^{ER}, both under physiological and stress-induced conditions (**Figure 103**).

As EWSR1 affects the shape and function of the ER, it might be interesting to test whether it is localized in the ER, and particularly if it is present in ER sheets or tubules, using superresolution (SR) imaging, as described in[701]. Furthermore, it would be important to check whether ER changes upon EWSR1 KD are accompanied with ER structural rearrangements, for instance by evaluating tubules to sheets ratio using a confocal spinning disk microscope, as in[701]. If this is the case, it will be relevant to determine if the disequilibrium in the ratio is associated to: i) altered levels of ER-shaping proteins, such as CLIM63 and RTN4B, or ii) an effect on general proliferation of ER-resident proteins, such as PDI and CNX. Levels of these proteins can be verified by WB or immunofluorescence experiments[701]. It might also be important to assess the implication of EWSR1 in ER dynamics, by analyzing the dynamic motion of ER tubules and three-way junctions, as described in[702]. As we only assessed ER shape and function and lipidomic analysis under basal conditions, it might also be important to conduct these experiments under ER-inducing conditions. Another necessary experiment will be to perform rescue experiments by EWSR1 re-expression to check whether ER dilation and ER stress can be reverted.

Furthermore, to rule out cell line-specific effects, we are currently repeating Annexin V staining and flow cytometry analysis, in other cell types (MCF7, A2780A, U2OS) and under ER-inducing stress conditions, such as palmitic acid (PA). Additionally, to formally demonstrate that TGs are subsequently deposited in LDs, as they were highly expressed upon EWSR1 KD (**Figure 75**), we are evaluating by confocal microscopy, if there is increased LDs biogenesis upon EWSR1 KD.

Remarkably, we found that glycosylation was among the most enriched processes associated with TE up genes (**Figure 72b**). To reinforce these data, it might be relevant to perform glycome analysis by matrix-assisted laser desorption/ionization–time-of-flight mass spectrometry (MALDI-TOF MS), to enable absolute and relative quantification of glycoprotein N and O-glycan abundances. To particularly verify that protein O-glycan glycosylation is enriched upon EWSR1 KD, it is important to evaluate the amounts of N-acetylgalactosamine transferase 2 (GALNT2) and the glycosyltransferase activity in ER, as reported in[702].

To examine our idea that the effect of EWSR1 on the ER and UPR^{ER} is mainly due to increased lipid accumulation upon EWSR1 KD that causes LBS, and not to accumulation of misfolding proteins, it may be interesting to use chemical chaperones such as tauroursodeoxycholic acid (TUDCA) that reduces stress-induced aggregation of proteins[703]. Therefore, we can examine whether perturbated ER shape/function and impairment of UPR^{ER} can be reverted. As the three UPR^{ER} sensors were low expressed upon EWSR1 KD, it is interesting to study their topology using proteinase K (PK) digestion from isolated microsomes in the presence/absence of EWSR1, as described in[694]. It will also be important to assess the levels of BiP (the protein

binding to UPR^{ER} sensors when inactivated (**see Introduction; "1.3.7.3 Sensing proteotoxic ER stress"**) by WB to see if it is affected by EWSR1.
Finally, an interesting perspective is to assess the contribution of EWSR1 to fatty acid (FA) synthesis by determining the relative proportions of ^{13}C-labelled palmitic, oleic and stearic acids in the presence/absence of EWSR1, as described in[704]. This will help us to resolve the question about the origin of lipid accumulation upon EWSR1 KD.

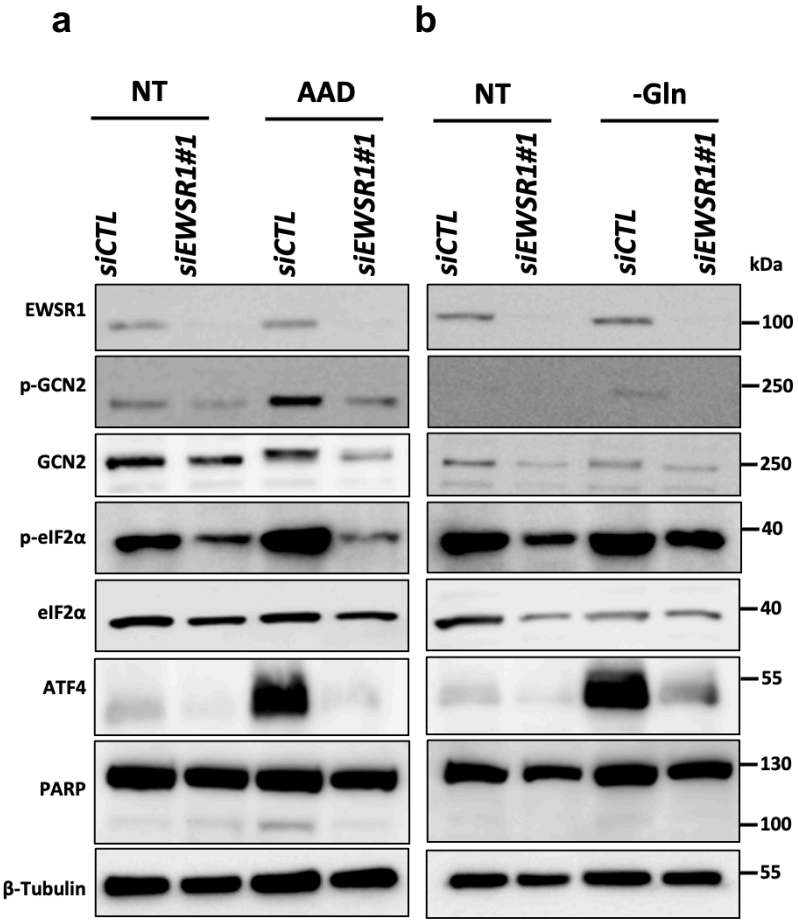


Figure 101: Knockdown of EWSR1 affects GCN2 axis of the ISR. (**a, b**) Western blot analysis of markers of the GCN2 axis and β-Tubulin (loading control) in cells transfected with *siEWSR1#1* or *siCTL*. Samples are total cell lysates from HeLa cells in nontreated (NT), amino acid deprivation (AAD; 8h) (**a**) or glutamine deprived conditions (-Gln; 8h) (**b**).

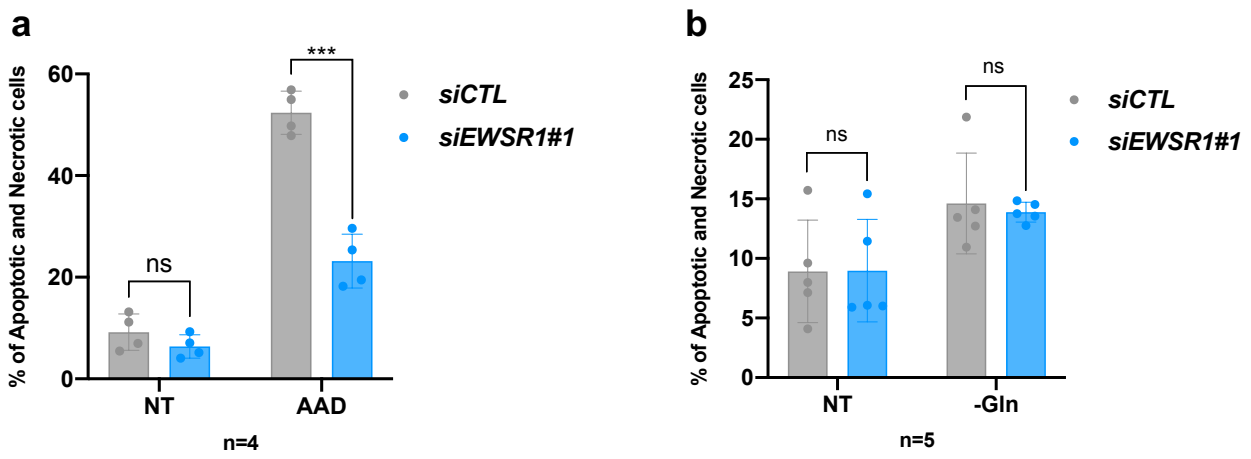


Figure 102: EWSR1 KD leads to decreased apoptosis in AAD treated condition. (**a**,**b**) Bar plots showing the percentage of apoptotic and necrotic cells. Samples are HeLa cells transfected with *siEWSR1#1* or *siCTL*, in nontreated (NT), amino acid deprivation (AAD; 18h) (**a**) or glutamine deprived conditions (-Gln; 48h) (**b**). Results are means ± sd (n ≥ 4 independent experiments). ***$P < 0.001$, ns = not significant compared to the *siCTL* condition by two-tailed unpaired Student's *t* test.

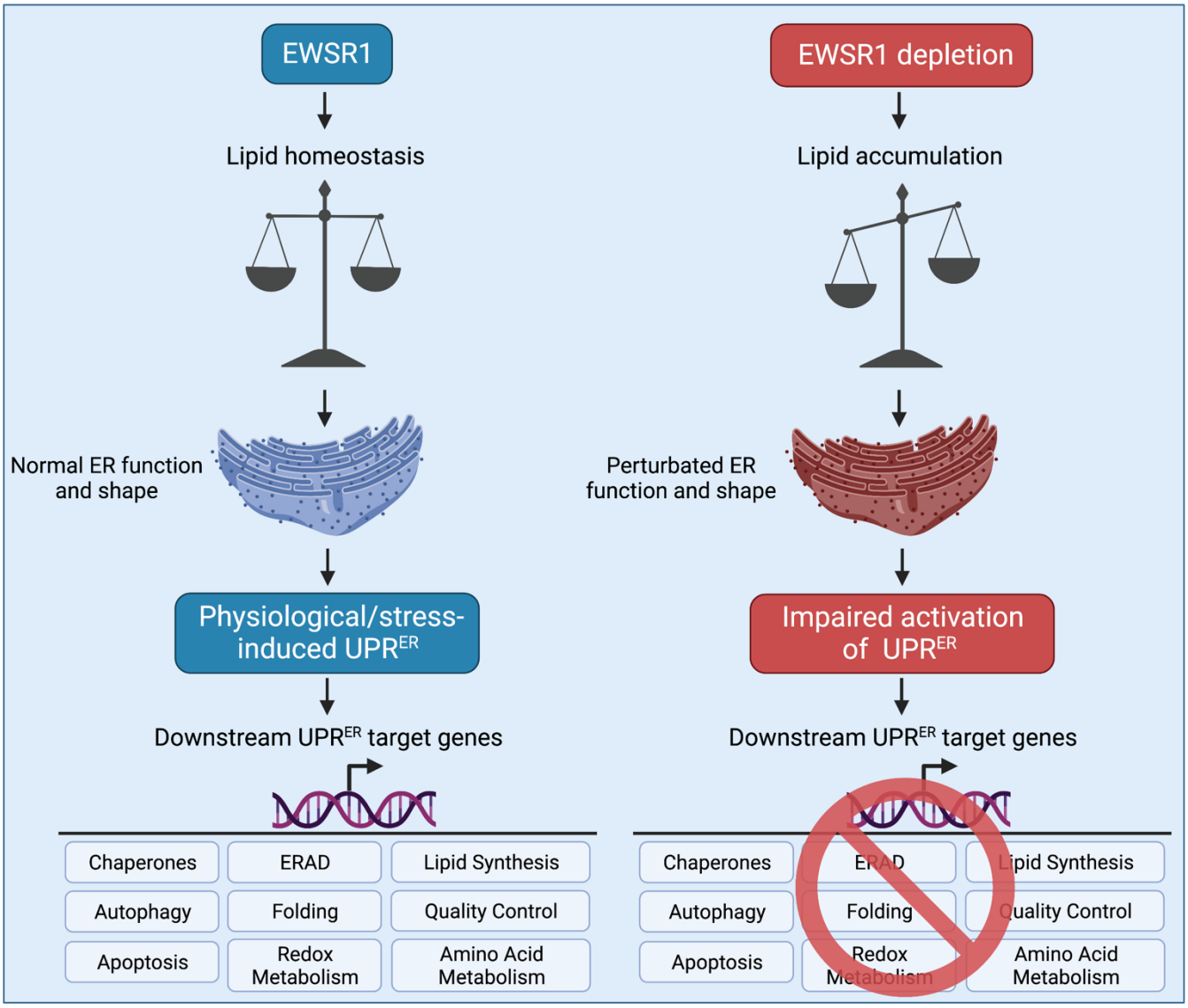


Figure 103: Model illustrating the new physiological function of EWSR1 in maintaining cellular homeostasis identified in this work. Created with Biorender.com. EWSR1 is important for the regulation of lipid homeostasis, which maintains normal ER function and shape and induces UPR^{ER} and its downstream effects (**left**). However, depletion of EWSR1 significantly leads to lipid accumulation, which perturbates ER function and shape, and impairs lipid-mediated activation of the UPR^{ER} (**right**). UPR^{ER}: unfolded protein response, ERAD: ER-associated degradation, ER: endoplasmic reticulum.

5 MATERIALS AND METHODS

5.1 Provenience of cell lines and cell culture conditions

HeLa, HEK293, A2780A and U2OS cells were purchased from American Type Culture Collection (ATCC). All cells were cultured in a 37°C, 5% CO_2 humidified incubator with medium supplemented with 10% FBS (Gibco) and 1% Penn/Strep (Biowest). All cell lines were maintained in DMEM high glucose (Biowest) and were routinely checked for mycoplasma by MycoAlertTM Mycoplasma Detection Kit (Lonza).

5.2 Cell treatments

48h of post-KD of EWSR1 with *siEWSR1#1* or *siEWSR1#2*, HeLa, A2780A and U2OS cells were treated with tunicamycin, 0.5 µg/ml-5 h; thapsigargin, 500 nM-24h, palmitic acid, 800 µM-24h; Amino acid deprivation, AAD-8h or Glutamine deprivation, -Glu, 8h. Cells were lysed in Laemmli buffer, ultrasonicated, boiled for 5 min at 100°C and analyzed by SDS-PAGE and western blotting.

5.3 Plasmids and cloning

Open reading frames (ORF) encoding human EWSR1 was obtained as pDONR223 from the human ORFeome v7.1 and v8.1 (The Center for Cancer Systems Biology, Dana-Farber Cancer Institute, CCSB-DFCI). N-terminal domain of EWSR1 (NTD), C-terminal domain of EWSR1 (CTD), CTD mutants and EWSR1 mutant were inserted into pDONR223 by BP cloning (Gateway recombination technology, Invitrogen) with specific primers flanked at the 5'site by the following AttB1 and AttB2 Gateway sites: 5'-GGGGACAACTTTGTACAAAAAAGTTGGC(ATG)-3' (AttB1) and 5'-GGGGACAACTTTGTACAAGAAAGTTGA-3' (AttB2). All constructs include an ATG initiation codon and lack a STOP codon. ΔRGG2, ΔZnF, ΔRGG3, ΔRGG2-ZnF, ΔZnF-RGG3, ΔRGG2-ZnF-RGG3 mutants were generated from pDONR223-CTD using the overlap extension PCR cloning procedure as described in[705]. EWSR1 ΔRGG2-ZnF-RGG3 mutant was generated from pDONR223-EWSR1 using the overlap extension PCR cloning procedure as described in[705]. Inserts from pDONR223 were subsequently transferred by LR cloning (Invitrogen) into different destination vectors: pDEST1899 (FLAG N-terminal tag), pGEX-6P-1 (Addgene). For the MS2-tethering assay, pDEST1899 Flag-tagged ORFs were subcloned in the pN-MS2-CP (MS2 N-terminal tag, described in[706]) following classical cloning procedure. To generate CrPv-*RLuc*-0 or CrPv-*RLuc*-8, the CrPv IRES sequence was amplified from CrPV-*luciferase*-MS2 reporter (a gift from Dr. Nicola K. Gray (MRC Human Genetics Unit, United Kingdom)) with 5'-GGTGGCGGCTAGCAAAGCAAAAATGTGATCTTGC-3' (Forward) and 5'-GGTGGCGGCTAGCCTATCTTGAAATGTAGCAGGTAA-3' (Reverse). The PCR products were digested with NheI and ligated into the NheI site of our bidirectional reporters *RLuc*-0 and *RLuc*-

8, respectively. To generate PV-*RLuc*-0 or PV-*RLuc*-8, the PV IRES sequence was amplified from PV-*luciferase*-MS2 reporter (a gift from Dr. Nicola K. Gray (MRC Human Genetics Unit, United Kingdom)) with 5'-GGTGGCGGCTAGCGGGAACAAAAGCTGGTACCG-3' (Forward) and 5'-GGTGGCGGCTAGCCGAGCTGAATCTCTATAATAATTAATGG-3' (Reverse). The PCR products were digested with NheI and ligated into the NheI site of our bidirectional reporters *RLuc*-0 and *RLuc*-8, respectively. All constructs were verified by Sanger sequencing at the GIGA-Genomics facility (University of Liège). All cloning and sequencing primers are listed in **Table 10**. Main plasmid maps are available in the **Appendix** section.

5.4 Plasmid DNA and siRNA transfection

For the MS2-tethering assay, HeLa cells were transfected with polyethyleneimine (PEI) using 1:2 DNA/PEI ratio. For co-immunoprecipitations, HeLa cells were transfected with PEI. For transient knockdown of EWSR1, AGO2 and PRMT1, siRNAs purchased from Eurogentec (Belgium) were used: *siEWSR1#1*, 5'-GAUGAAGACUCUGACAACAtt-3'; *siEWSR1#2*, 5'-GAUUAUUCCUUGUCUGUACUUtt-3', *siAGO2,* 5'-AAGGAUAUGCCUUCAAGCCUCtt-3', *siPRMT1*, 5'-CGUCAAAGCCAACAAGUUAtt-3' and control *siRNA* (SR-CL000-005). siRNAs were transfected in HeLa cells with JetPrime (Polyplus transfection) or Lipofectamine 2000 Transfection Reagent (Invitrogen) according to the manufacturer's instructions. Cells were processed at 48h after DNA or siRNA transfections.

5.5 RNA isolation and quantitative PCR

For MS2-tethering assays and RT-PCR of unspliced (u) and spliced (s) XBP1, total RNA was isolated using the Nucleospin RNA kit (Macherey-Nagel). For RNA-immunoprecipitations, total RNA was extracted from inputs or beads with TRIzol Reagent (Thermo Fisher Scientific). Total RNA was then reverse-transcribed with random primers using the RevertAid H Minus First Strand cDNA Synthesis Kit (Thermo Fisher Scientific). cDNA was diluted depending on the abundance of targets, amplified and quantified in triplicate using FastStart SYBR Green Master mix (Roche) on a LightCycler 480 instrument (Roche). In MS2-tethering studies, relative levels of *RLuc* mRNA were performed using primers specific for *Fluc* (5'-TGAGAACTTCAGGCTCCTGG-3' (Forward) and 5'-GCCTTATGCAGTTGCTCTCC-3' (Reverse)) and *Rluc* (5'-TCTTTTTCGCAACGGGTTT-3' (Forward) and 5'-GCCCAGTTTCTATTGGTCTCC-3' (Reverse)). Fold changes were calculated relatively to the MS2-CP condition with the ΔΔCt method. Primers used for RT-qPCR of unspliced (u) and spliced (s) XBP1 were from[707] and were 5'-CTGGAACAGCAAGTGGTAGA-3' (Forward) and 5'-CTGGGTCCTTCTGGGTAG-3' (Reverse).

5.6 Western blotting

Cells were washed with cold PBS, directly lysed in Laemmli buffer, ultrasonicated, and boiled for 5 min at 100°C. Protein extracts were separated by sodium dodecyl sulfate-polyacrylamide gel (SDS-PAGE) electrophoresis and transferred onto nitrocellulose membrane. Membranes were blocked for 1h with 5% non-fat milk and incubated in 4% BSA overnight at 4°C with primary antibodies followed by HRP-conjugated secondary antibodies. Proteins were detected by chemiluminescence. Images were acquired with ImageQuant LAS 4000 device (GE Healthcare) or Amersham ImageQuant 800 and quantified using ImageJ. Primary antibodies are listed in **Table 4**.

5.7 MS2-tethering assay

Luciferase MS2-tethering assays were performed in HeLa cells. Briefly, cells were transfected with control MS2-CP or various MS2-CP- or FLAG-tagged constructs, together with a bidirectional reporter encoding a control *Firefly* luciferase (*FLuc*) and a targeted *Renilla* luciferase (*RLuc*) carrying or lacking four repeats of the binding sequence for the MS2 coating peptide in its 3'UTR (*RLuc-4* and *RLuc-0*, respectively). For CrPv- or PV-*RLuc* reporters-tethering assays, cells were transfected with control MS2-CP or various MS2-CP-tagged constructs, together with a CrPV- or PV-*RLuc* bidirectional reporter encoding a control *Firefly* luciferase (*FLuc*) and a targeted *Renilla* luciferase (*RLuc*) carrying or lacking eight repeats of the binding sequence for the MS2 coating peptide in its 3'UTR (CrPv/PV-*RLuc-8* and CrPv/PV-*RLuc-0*, respectively). For TISU-*RLuc*-6xMS2-poly(A) or *RLuc*-6xMS2-poly(A) reporters-tethering assays, cells were transfected with control MS2-CP or various MS2-CP-tagged constructs, together with TISU-*RLuc*-6xMS2-poly(A) or *RLuc*-6xMS2-poly(A) reporters and pGL3-*Firefly* vector (Addgene) (a control *Firefly* luciferase vector). For measuring luciferase activities in all these assays, cell lysis and luciferase activity were performed in triplicate on a TriStar^2 S LB 942 luminometer (Berthold) with twinlite *Firefly* and *Renilla* Luciferase Reporter Gene Assay System (PerkinElmer).

5.8 SUnSET assay

HeLa cells were transfected with *siCTL* or *siEWSR1* (*siEWSR1#1* or *siEWSR1#2*). 48 hours post-transfection, cells were incubated with 1μM puromycin for 30 min, washed 2 x with PBS to remove residual puromycin, then lysed in Laemmli buffer and ultrasonicated. Lysates were finally boiled and subsequently analyzed by SDS-PAGE and western blotting.

5.9 Proliferation assay

48h of post-KD of EWSR1 with *siEWSR1#1* or *siEWSR1#2*, HeLa cells were seeded at a density of 7,000 cells per well in triplicate of a 24-well plate (Greiner). Cell growth was monitored and analyzed by time-lapse microscopy (IncuCyte S3, Essen Bioscience, Sartorius) every 4h for 72h.

5.10 Annexin V apoptosis assay

Cell death or apoptosis induced by EWSR1 KD in HeLa cells in nontreated and treated conditions (Tunicamycin, 0.5 µg/ml-48 h; Thapsigargin, 500 nM-48h, Palmitic Acid, 800 µM-24h; Amino acid deprivation, AAD-18h; Glutamine deprivation, -Glu, 48h) was assessed by flow cytometry analysis with dual labelling Annexin-V-FITC and Propidium Iodide. The latter two compounds were obtained from Becton and Dickinson (Franklin lakes, NJ, USA) and Sigma, respectively. A minimum of three independent experiments were performed.

5.11 FACS labeling and analysis

Cells trypsinized, rinsed 3 x with cold PBS, were fixed with ice-cold 70% ethanol for 30 min at 4°C, then used immediately. After fixation, cells were resuspended twice in 1 ml of PBS, then treated with 10 µg of RNase A at RT for 5 min. Finally, 400 µL Propidium Iodide (50 µg/mL in PBS) were added and data were acquired on CytoFlex Flow cytometer.

5.12 Co-immunoprecipitation (co-IP)

HeLa cells co-expressing tagged interaction partners were lysed in IPLS (immunoprecipitation low salt; Tris-HCl, pH 7.5, 50 mM, EDTA, pH 8, 0.5 mM, 0.5% NP-40, 10% glycerol, 120 mM NaCl) with 1 x cOmplete Protease Inhibitor (Roche) and 1 x Halt Phosphatase Inhibitors (Thermo Fisher Scientific). If needed, cleared cell lysates were incubated at 37°C for 30 min with or without 10 µg/ml RNase A (Thermo Fisher Scientific) or 2000 U RNase T1 (Thermo Fisher Scientific, catalog no. EN0541). Supernatants were then incubated with anti-FLAG M2 agarose beads (Sigma-Aldrich) for 2h at 4°C. Beads were washed 4 x with IPLS. Immunoprecipitates were finally boiled in 2 x SDS loading buffer and analyzed by SDS-PAGE and western blotting. For endogenous coimmunoprecipitations (co-IPs), HEK293 cell lysates were prepared in IPLS, precleared with Protein A/G magnetic beads (Millipore) for 1h at 4°C, and incubated overnight with 1.5 µg of relevant antibody or mouse (Santa Cruz) IgG. Then, samples were incubated with Protein A/G magnetic beads (Millipore) for 2h at 4°C. Beads were washed 4 x with IPLS buffer. Immunoprecipitates were finally boiled in 2 x SDS loading buffer and analyzed by SDS-PAGE and western blotting.

5.13 Oligo(dT) pulldowns

HeLa cells were washed twice in cold PBS and UV cross-linked (0.4 J/cm^2 of 365-nm UV light with a Stratalinker 2400). UV-cross-linked cells were divided into two fractions: 1) total and 2) cytoplasmic extracts. For total extract fraction, cells were lysed in oligo(dT) lysis buffer (Tris-HCl, pH 7.5, 20 mM, 500 mM LiCl, 0.5% SDS, 1 mM EDTA, 1 mM DTT, cOmplete Protease Inhibitor (Roche), Halt Phosphatase Inhibitors (Thermo Fisher Scientific) and Protector RNase Inhibitor (Roche)), then cell lysates were homogenized by four passages through a syringe with a 0.4-mm diameter needle. For cytoplasmic extract fraction, cells were resuspended in cytoplasmic extraction buffer (Tris-HCl, pH 7.5, 20 mM, 100 mM KCl, 5 mM $MgCl_2$, 0.3% Nonidet P-40, 1 mM DTT, cOmplete Protease Inhibitor (Roche), Halt Phosphatase Inhibitors (Thermo Fisher Scientific) and Protector RNase Inhibitor (Roche)). Then, after centrifugation, the supernatant was collected and lysed in 2 x lysis buffer (Tris-HCl, pH 7.5, 20 mM, 0.9 mM LiCl, 0.7% SDS, 2 mM EDTA, 1 mM DTT, cOmplete Protease Inhibitor (Roche), Halt Phosphatase Inhibitors (Thermo Fisher Scientific) and Protector RNase Inhibitor (Roche)). Homogenized cell lysates from total and cytoplasmic extracts were then incubated at 37°C for 30 min with or without RNase A (200 µg/ml Thermo Fisher Scientific) (neither DTT nor RNase protector was added in the RNase-treated samples used as negative controls for oligo(dT) pulldown). 1% of each cell lysate was kept for input, and the rest was incubated with oligo(dT) magnetic beads (New England BioLabs) for 2h at 4°C (25 µl of beads per 15-cm culture dish). Pulldowns were washed 1 x with wash buffer I (Tris-HCl, pH 7.5, 20 mM, 500 mM LiCl, 0.1% SDS, 1 mM EDTA and 1 mM DTT), 1 x with wash buffer II (Tris-HCl, pH 7.5, 20 mM, 500 mM LiCl, 1 mM EDTA and 1 mM DTT) and 1 x with wash buffer III (Tris-HCl, pH 7.5, 20 mM, 200 mM LiCl, 1 mM EDTA and 1 mM DTT). As cell lysates and successive washes remained viscous, bead homogenization was sometimes difficult, and bead separation (PureProteome Magnetic Stand, Millipore) took several minutes. Beads and inputs were resuspended in 2 x SDS loading buffer, sonicated for 5 min to reduce viscosity, boiled, and analyzed by SDS-PAGE and western blotting.

5.14 RNA-immunoprecipitation-qPCR analysis

Protein A/G magnetic beads (Millipore) were incubated with anti-EWSR1 (sc-398318, Santa Cruz) antibody or mouse IgG (Santa Cruz) at 4°C overnight with rotation. Before cell harvesting, RNA-protein complexes were crosslinked with 1% formaldehyde (incubation at RT for 10 min) and crosslink reaction was quenched using 125 mM glycine for 5 min at RT. Cells were washed 2 x with ice-cold PBS, then pelleted by centrifugation (4°C, 5 min, 1,600 x rpm). Cells were resuspended in 1 ml RIPA (Tris HCl, pH 8.0, 50 mM, 150 mM NaCl, 1.0% IGEPAL CA-630, 0.5% sodium deoxycholate, 0.1% SDS) with 1x cOmplete Protease Inhibitor (Roche) and lysed for 30 min at 4°C with rotation. Cell lysates were then sonicated on ice and cleared by centrifugation for 10 min at 4°C at 10,000 x rpm. Proteins were quantified with Pierce BCA protein assay kit (Thermo Fisher Scientific). An aliquot was used for RNA input and was treated with proteinase

K before RNA extraction with TRIzol Reagent (Thermo Fisher Scientific). For immunoprecipitation, 400 µg of protein were incubated with antibody-loaded Protein A/G magnetic beads overnight with rotation at 4°C. Supernatant was removed and beads were washed 5 x with 1 ml RIPA buffer. RNA-protein complexes were then eluted by incubation with 100 µL of elution buffer (Tris-HCl, pH 8, 100 mM, 10 mM Na2-EDTA, 1% SDS in H_2O) for 3 min at 90°C. Proteins were digested with proteinase K treatment and RNA from input or beads were extracted with TRIzol Reagent (Thermo Fisher Scientific) for RT-qPCR analysis. *GAPDH* mRNAs was used as control. All primers used for RIP experiments are listed in **Table 10**.

5.15 Polysome profiling

Cells were treated with CHX 100 µg/ml for 5 min to fix polysomes, and harvested in buffer containing 25 mM Hepes, 100 mM KCl, 5 mM $MgCl_2$, 0.5% Nonidet P-40, 2 µg/ml heparin, and 10 µg/ml cycloheximide. Cell extracts were cleared by centrifugation at 13,000 x rpm for 10 min at 4°C and loaded onto 15-50% linear sucrose gradients. Samples were ultracentrifuged for 2 hours at 39,000 x rpm in a SW41 rotor and subsequently fractionated with a gradient fractionator (Brandel). Absorbance was read at 260 nm and collected fractions are stored at -80°C for subsequent analysis by SDS-PAGE and western blotting.

5.16 Nuclear and cytoplasmic fractionation

HeLa cells were tripsinized, washed 2 x with cold PBS and lysed with CLB buffer (cytoplasmic lysis buffer; Tris-HCl, pH 7.9, 10 mM, 340 mM sucrose, 3 mM $CaCl_2$, 0.1 mM EDTA, 2 mM $MgCl_2$, 1 mM DTT, 0.5% NP40, cOmplete Protease Inhibitor Cocktail (Roche) and Halt Phosphatase Inhibitor Cocktail (Thermo Fisher Scientific)) on ice for 5 min. The cytoplasmic fraction was collected by a first centrifugation at 3,500 x g for 15 min at 4°C, then at maximum speed for 10 min. The pellet was washed several times with CLB wash buffer (CLB buffer without NP40) and lysed with NLB buffer (nuclear lysis buffer; HEPES, pH 7.9, 20 mM, 10% glycerol, 3 mM EDTA, 150 mM KOAc, 1.5 mM $MgCl_2$, 1 mM DTT, 0.1% NP40, cOmplete Protease Inhibitor Cocktail (Roche), Halt Phosphatase Inhibitor Cocktail and Protector RNase Inhibitor (Roche)). The nuclear fraction was collected by a first centrifugation at 15,000 × g for 30 min at 4°C, then at a maximum speed for 10 min at 4°C. Cell fractions were then analyzed by SDS-PAGE and western blotting.

5.17 Cap-association assay using m^7GTP-Agarose

To analyze the interaction of EWSR1 to cap-binding protein complex, HeLa cells were lysed as previously described in[351], using NET2 buffer (Tris-HCl, pH 7.4, 50 mM, 150 mM NaCl, 1 mM phenylmethanesulfonylfluoride, 2 mM benzamidine, 1% NP-40) and total cell extracts were incubated with m^7GTP-Agarose (Jena Bioscience, AC-155S) for 2 hours at 4°C. Then, the beads

were washed 3 x in NET2 buffer. The samples were analyzed by SDS-PAGE and western blotting.

5.18 Subcellular fractionation

The S30, S100, R and RSW fractions were obtained as previously described in[609] with little modifications. Briefly, HEK293 cells, grown to 90% confluence in 10-12 P100 dishes, were washed with ice cold PBS, and lysed in buffer 1 (Tris-HCl, pH 7.4, 15 mM, 80 mM KCl, 5 mM $MgCl_2$, 1% Triton-X-100, cOmplete Protease Inhibitor Cocktail (Roche)). Cell debris was discarded by spinning at 14,000 x g for 10 min at 4°C, twice. The supernatant of the second spinning is the S30 fraction. S30 centrifugation at 95,000 x rpm during 1.5 h using the TLA100.1 rotor yielded the S100 fraction (supernatant), and the ribosomes with associated factors (pellet-R fraction). To prepare the fraction containing ribosomes free of associated factors, the ribosomal pellet was resuspended in high salt buffer 2 (Tris-HCl, 15 mM, pH 7.4, 500 mM KCl, 5 mM $MgCl_2$, 2 mM DTT, 290 mM sucrose), loaded in a discontinuous sucrose gradient (1.5 ml buffer 40% (w/v) sucrose, Tris-HCl, pH 7.4, 15 mM, 500 mM KCl, 5 mM $MgCl_2$, 2 mM DTT (bottom layer) and 1 ml buffer 20% (w/v) sucrose, Tris-HCl, pH 7.4, 15 mM, 500 mM KCl, 5 mM $MgCl_2$, 2 mM DTT (top layer)), centrifuged at 4°C, 95,000 x rpm for 2 h using a TLA100.1 rotor. The pure ribosomes pellet (RSW fraction) was resuspended in buffer 1. S30, S100, R and RSW fractions were analyzed by SDS-PAGE and western blotting.

5.19 GST-fusion proteins purification and pulldowns

Fusion proteins were expressed in One Shot BL21 (DE3) Chemically Competent *E. coli* (Invitrogen) and purified according to the following protocol. Bacterial culture with appropriate antibiotics was incubated overnight (37°C, 220 x rpm). Next morning, the overnight starter culture was diluted into LB (1:10 dilution) supplemented with appropriate antibiotics and incubated for 2h (37°C, 220 x rpm). Then, IPTG, dioxane-free (Thermo Fisher Scientific) was added to a final concentration of 0.1 mM to induce protein expression and the culture was incubated for 4h (30°C, 220 x rpm). Cells were harvested by centrifugation at 10,000 x rpm for 10 min at 4°C, then lysed in lysis buffer (1 x PBS containing 1% Triton) followed by sonication on ice. The lysate was centrifuged at 10,000 x g for 10 min at 4°C and the supernatant was incubated with pre-equilibrated Glutathione Sepharose 4B Fast Flow (Cytiva) for 30 min at 4°C with rocking. Beads were collected by centrifugation at 1,000 x g at 4°C for 3 min, then washed 3 x in PBS. After centrifugation, beads can be used immediately for pulldown or stored at -80°C in PBS 10% glycerol.

For pulldown reactions, GST fusion proteins were incubated in IPLS with ribosomal fractions (40S or 60S/80S) collected from polysome profiling. After rocking for 30 min at 4°C, beads were

washed 3 x in IPLS buffer. Pulldown reactions were then analyzed by SDS-PAGE and western blotting.

5.20 Ribosome profiling (RIBO-seq)

Hela cell pellets, pre-treated with 100 mg/ml cycloheximide, were delivered at OHMX.bio facilities (Ghent, BE) in good condition. Frozen cell pellets were resuspended with 2 ml of lysis buffer supplemented with 100 mg/ml cycloheximide on ice. Efficient cell lysis was accomplished by passing the suspension through a 18G needle attached to a syringe. The cell suspension was incubated in lysis buffer for 10 min and centrifuged twice for complete removal of the cell debris. The clear lysate was aliquoted for total RNA purification with the RNA Clean and Concentrator-5 kit (Zymo Research) and subsequent footprinting. For footprinting and monosome purification, 10 µg of lysate was digested with RNase I (Thermo Fisher Scientific) at 21°C for 45 min. The reaction was stopped with the addition of Superase In RNase Inhibitor (Thermo Fisher Scientific), and the digested lysate was purified on a MicroSpin S-400 HR Sephacryl column (GE Healthcare). The flow through was collected and purified with the RNA Clean and Concentrator-5 (Zymo Research). The footprints were visualized on the Agilent Bioanalyzer with the RNA 6000 Nano kit (Agilent).

Ribosome protected fragments (RPFs) of 28-30 nt were selected by loading the footprints on a Novex TBE-Urea gel (15%) (Thermo Fisher Scientific) and performing gel electrophoresis. The gel parts corresponding to the fragments of interest were excised and let to diffuse in molecular grade water supplemented with ammonium acetate and SDS overnight at 4°C. The slurry was then transferred to CoStar filter tubes and centrifuged. The flow through was incubated at -80°C for 2 h with glycogen and isopropanol. Following centrifugation at 18,000 x g for 20 min, the supernatant was removed and the pellet was washed with ice-cold 80% EtOH. Pellets were left to dry for 10 min and were resuspended in 15 µl molecular grade water. The quality of the eluted RNA fragments was checked on the Agilent Bioanalyzer using the Small RNA kit.

For library preparation, the RPFs were 3’ dephosphorylated with T4 PNK (New England Biolabs) for 1h at 37°C to heal the 2’-3’ cyclic phosphate from the RNase I cleavage and were subsequently purified using the Oligo Clean and Concentrator kit (Zymo Research). The 3’ adapter was ligated to the RPFs and the ligated fragments were purified following the protocol of the Small RNA Library prep kit (Lexogen). After the 5’ phosphorylation of the RPFs and their purification with the Oligo Clean and Concentrator kit, 5’ adapter ligation took place based on the Small RNA Library prep kit. First strand cDNA synthesis was performed according to the Small RNA Library prep kit manual. Ribosomal RNA (rRNA) was depleted with the addition of LNA probes from the QIAseq FastSelect rRNA HMR kit (Qiagen) during the cDNA synthesis reaction. Prior to Indexing PCR amplification, a qPCR was performed in order to determine the optimal PCR cycle number. During PCR, index primers (Lexogen) were added at both ends of

the fragments to produce dual-indexed libraries. The PCR amplification protocol from the Small RNA Library prep kit was followed. Library products were purified with AMPure XP beads and their quality was determined on the Agilent Bioanalyzer with the DNA High Sensitivity kit (Agilent). For the efficient removal of adapter dimers and excess index primers, the library products run on a Novex TBE gel (8%) (Thermo Fisher Scientific) and the gel part corresponding to the library products was excised and purified. The quality of the final libraries was determined on the Agilent Bioanalyzer with the High Sensitivity DNA kit. The concentration of the libraries was estimated by performing qPCR using a dilution of PhiX Control v3 as standard and samples were pooled in an equimolar way. A test sequencing run (2 x 75 bp) on the Illumina MiSeq platform was performed with the MiSeq Reagent Nano kit v2 (Illumina) to check the quality of the prepared libraries prior to the in-depth sequencing on Novaseq.

For matching mRNA-seq, 100 ng of total RNA material was used. mRNA was first enriched using oligo(dT) beads and was fragmented randomly by adding fragmentation buffer. Then, cDNA was synthesized by using random primers, after which a second-strand synthesis buffer, dNTPs, RNase H and DNA polymerase I were added to initiate the second-strand synthesis. Following end repair and sequencing adaptor ligation, the double-stranded cDNA library was completed through size selection and PCR enrichment. The quality of the final libraries was determined on the Agilent Bioanalyzer with the High Sensitivity DNA kit. The concentration of the libraries was estimated by performing qPCR using a dilution of PhiX Control v3 as standard and samples were pooled in an equimolar way. Finally, high coverage sequencing was performed on the Novaseq 6000 system with the S4 flowcell and SE50 configuration.

For the quality control of the sequencing reads, the following computational steps were successfully applied: i) trimming sequence adapters (SeqPurge), ii) checking the presence of rRNA and removing it (Bowtie2) and iii) mapping the trimmed reads to reference transcriptome and quantification of the reads at transcript level (STAR). To visualize and explore ribosome profiling features, specific RIBO-seq quality control steps were performed (triplet periodicity, P-site offset, RPF length centered around 29-30 nt, checking that the majority of mRNA reads mapped to CDS) (**see RESULTS section**). The transcriptomics (RNA-seq) and translatomics (RIBO-seq) counts were summed at gene level using feature Counts and afterwards fed into the popular package DESeq2[708] for differential analysis.

5.21 Lipidomic analysis

An amount of cells containing 10 µg of DNA was homogenized in 700 µL of water with a handheld sonicator and was mixed with 800 µl HCl (1M):CH3OH 1:8 (v/v), 900 µl $CHCl_3$, 200 µg/ml of the antioxidant 2,6-di-tert-butyl-4-methylphenol (BHT; Sigma Aldrich) and 3 µl of SPLASH® LIPIDOMIX® Mass Spec Standard (#330707, Avanti Polar Lipids). After vortexing and centrifugation, the lower organic fraction was collected and evaporated using a Savant

Speedvac spd111v (Thermo Fisher Scientific) at room temperature and the remaining lipid pellet was stored at -20°C under argon.

Just before mass spectrometry analysis, lipid pellets were reconstituted in 100% ethanol. Lipid species were analyzed by liquid chromatography electrospray ionization tandem mass spectrometry (LC-ESI/MS/MS) on a Nexera X2 UHPLC system (Shimadzu) coupled with hybrid triple quadrupole/linear ion trap mass spectrometer (6500+ QTRAP system; AB SCIEX). Chromatographic separation was performed on a XBridge amide column (150 mm × 4.6 mm, 3.5 µm; Waters) maintained at 35°C using mobile phase A [1 mM ammonium acetate in water-acetonitrile 5:95 (v/v)] and mobile phase B [1 mM ammonium acetate in water-acetonitrile 50:50 (v/v)] in the following gradient: (0-6 min: 0% B → 6% B; 6-10 min: 6% B → 25% B; 10-11 min: 25% B → 98% B; 11-13 min: 98% B → 100% B; 13-19 min: 100% B; 19-24 min: 0% B) at a flow rate of 0.7 ml/min which was increased to 1.5 ml/min from 13 minutes onwards. SM, CE, CER, DCER, HCER, LCER were measured in positive ion mode with a precursor scan of 184.1, 369.4, 264.4, 266.4, 264.4 and 264.4 respectively. TAG, DAG and MAG were measured in positive ion mode with a neutral loss scan for one of the fatty acyl moieties. PC, LPC, PE, LPE, PG, PI and PS were measured in negative ion mode by fatty acyl fragment ions. Lipid quantification was performed by scheduled multiple reactions monitoring (MRM), the transitions being based on the neutral losses or the typical product ions as described above. The instrument parameters were as follows: Curtain Gas = 35 psi; Collision Gas = 8 a.u. (medium); IonSpray Voltage = 5500 V and -4,500 V; Temperature = 550°C; Ion Source Gas 1 = 50 psi; Ion Source Gas 2 = 60 psi; Declustering Potential = 60 V and −80 V; Entrance Potential = 10 V and -10 V; Collision Cell Exit Potential = 15 V and -15 V. The following fatty acyl moieties were taken into account for the lipidomic analysis: 14:0, 14:1, 16:0, 16:1, 16:2, 18:0, 18:1, 18:2, 18:3, 20:0, 20:1, 20:2, 20:3, 20:4, 20:5, 22:0, 22:1, 22:2, 22:4, 22:5 and 22:6 except for TGs which considered: 16:0, 16:1, 18:0, 18:1, 18:2, 18:3, 20:3, 20:4, 20:5, 22:2, 22:3, 22:4, 22:5, 22:6.

Peak integration was performed with the MultiQuantTM software version 3.0.3. Lipid species signals were corrected for isotopic contributions (calculated with Python Molmass 2019.1.1) and were quantified based on internal standard signals and adheres to the guidelines of the Lipidomics Standards Initiative (LSI) (level 2 type quantification as defined by the LSI). Unpaired T-test *P*-values and FDR corrected *P*-values (using the Benjamini/Hochberg procedure) were calculated in Python StatsModels version 0.10.1.

5.22 Transmission electron microscopy (TEM)

Cells were fixed for 60 minutes at 4°C with 2.5% glutaraldehyde in a Sörensen 0.1 M phosphate buffer (pH 7.4) and post-fixed for 30 min with 2% osmium tetroxide. After dehydration in graded ethanol, samples were embedded in Epon. Ultrathin sections obtained with a Reichert Ultracut S ultramicrotome were contrasted with uranyl acetate and lead citrate. Observations were made

with a Jeol JEM-1400 transmission electron microscope at 80 kV.

5.23 Statistics

Unless otherwise indicated, graph values are presented as mean ± sd, calculated for at least three independent experiments. Statistical tests were performed with GraphPad Prism 8 and are indicated in each figure legends. *P*-value thresholds are depicted as follows: **P* < 0.05; ***P* < 0.01; ****P* < 0.001; *****P* < 0.0001; ns: not significant.

5.24 Acknowledgements

We thank OHMX.bio facilities (Ottergemsesteenweg 460, 9000 Ghent, BE, contact@ohmx.bio - www.ohmx.bio) for performing RIBO-seq experiment. We thank Prof. Marc Thiry (laboratory of Cellular and Tissular biology) for the TEM experiment. We thank the laboratory of Lipid Metabolism and Cancer (KU, Leuven) for the lipidomic experiment. We thank Véronique Kruys and Romuald Soin (ULB) for polysome profiling experiments. We thank the GIGA-Viral Vectors, Imaging and Genomics facilities (University of Liège) for technical support and helpful assistance. This study was supported by the University of Liège (ULiège), the Fonds National de la Recherche Scientifique (FNRS), Télévie, Fondation contre le Cancer, and the Fonds Léon Frédéricq.

Antibodies	Source	Host species
EWSR1-M	SantaCruz (sc-398318)	Mouse
EWSR1-R	Cell signaling (#11910)	Rabbit
FLAG-M	Sigma (F3165)	Mouse
FLAG-R	Sigma (F7425)	Rabbit
GAPDH	SantaCruz (sc-166545)	Mouse
β-tubulin	Abcam (ab6046)	Rabbit
Histone H3	SantaCruz (sc-8654)	Goat
Hsp90	Cell signaling (#4877)	Rabbit
Puromycin	Millipore (MABE343-clone 12D10)	Mouse
RPS6	Cell signaling (#2217)	Rabbit
RPS23	SantaCruz (sc-100837)	Mouse
RPL26	Bethyl (A-300-686)	Rabbit
RPL22	SantaCruz (sc-373993)	Mouse
RPS5	SantaCruz (sc-390935)	Mouse
eIF2α	Cell signaling(#9722)	Rabbit
p-eIF2α	Cell signaling (#9721)	Rabbit
PERK	SantaCruz (sc-377400)	Mouse
ATF4	Cell signaling (#11815)	Rabbit
ATF6	Cell signaling (#65880)	Rabbit

GCN2	SantaCruz (sc-374609)	Mouse
p-GCN2 (phospho T899)	Abcam (ab75836)	Rabbit
CHOP	Cell signaling (#2895)	Mouse
PARP	Cell signalling (#9532)	Rabbit
$p21^{Cip1}$	SantaCruz (sc-397)	Rabbit
IRE1α	SantaCruz (sc-390960)	Mouse
p-IRE1α (phospho S724)	Abcam (ab124945)	Rabbit
XBP1s	Cell signaling (#12782)	Rabbit
eIF3B	Boster Immunoleader (PA2030)	Rabbit
eIF3η	SantaCruz (sc-137214)	Mouse
RBPMS	Proteintech (15187-1-AP)	Rabbit
RPL4	SantaCruz (sc-100838)	Mouse
RACK1	SantaCruz (sc-17754)	Mouse
eIF4E	Cell signaling (#9742)	Rabbit
eIF5	SantaCruz (sc-28309)	Mouse
eIF4G	SantaCruz (sc-133155)	Mouse
GST	SantaCruz (sc-138)	Mouse

Table 4: Antibodies used in this work with their source and host species.

Chemicals	Source	Identifier
Spectinomycin	Sigma-Aldrich	S4014
Ampicillin	Sigma-Aldrich	A0166
Polyethyleneimine (PEI)	Sigma-Aldrich	NA
RNAse A	Thermo Fisher Scientific	Cat#EN0531
JetPrime	Polyplus	101000027
FastDigest XhoI (10 U/µL)	Thermo Fisher Scientific	FD0694
T4 DNA ligase	Thermo Fisher Scientific	EL0011
cOmplete Protease Inhibitors	Roche	CO-RO
Anti-FLAG M2 agarose beads	Millipore	A2220
Pierce Protein A/G Magnetic beads	Thermo Fisher Scientific	11844554
TakyonTM No ROX SYBR 2X MasterMix blue dTTP	Eurogentec	UF-NSMT-B0701
Thapsigargin	Sigma-Aldrich	T9033
Cycloheximide 100 mg/ml	Sigma-Aldrich	C4859

Twinlite Firely and Renilla Luciferase Reporter Gene Assay System	Perkin Elmer	6066709
NucleoSpin Plasmid EasyPure	Macherey-Nagel	740727.50
Nucleospin RNA kit	Macherey-Nagel	740955.50
Palmitic acid	Sigma-Aldrich	SLCD8134
Tunicamycin	Sigma-Aldrich	T7765
DMEM w/o glutamine, w/o amino acids, with 1 g/L glucose	GENAXXON bioscience	C4150.0500
MetaPhor® Agarose	Lonza Bioscience catalog n° 50181	NA

Table 5: Chemicals used in this work with their source and identifier. NA = not available/applicable.

Cell lines	Source	Identifier
Human: HeLa	ATCC	CCL-2
Human: HEK293	ATCC	CRL-1573
Human: U2OS	ATCC	HTB-96
Human: A2780A	ATCC	CVCL_0134

Table 6: Cell lines used in this work with their source and identifier.

Plasmids	Source	Identifier
Human ORFeome v7.1 and v8.1	The Center for Cancer Systems Biology (CCSB- DFCI)	http://horfdb.dfci.harvard.edu/
pDONR223	[709]	NA
pDEST1899	[709]	NA
pN-MS2-CP	[709]	NA
RLuc-4MS2	[709]	NA
RLuc-8MS2	[709]	NA
RLuc-0MS2	[709]	NA
pGL3-promoter	Addgene	NA
pGEX-4T1	Addgene	NA
pGEX-6p-1	Addgene	NA
TISU-RLuc-6xMS2-poly(A)	[608]	NA
RLuc-6xMS2-poly(A)	[608]	NA

Table 7: Palsmids used in this work with their source and identifier. NA = not available/applicable.

Plasmids	Forward	Reverse
pDONR223	TCCCAGTCACGACGTTGTAA	GTAATACGACTCACTATAGG
pDEST1899	GACGGTGATTATAAAGATCATGACATCG	GGAGTGGCAACTTCCAGGG
pN-MS2	CTAAAGTGGCAACCCAGACTGTTGG	GGAGGGGCAAACAACAGATGGC
pGEX	CCGGGAGCTGCATGTGTCAGAGG	GGGCTGGCAAGCCACGTTTGGTG

Table 8: Plasmids sequencing primers.

Target (Cloning)	Forward	Reverse
RGG1-RRM	GGGGACAACTTTGTACAAAAAAGTTGGCATGGGAGGATTTTCCGGACCAGG	GGGGACAACTTTGTACAAGAAAGTTGAAAGGGAGACTTTAAGTTTGC
ΔRGG1-RRM	GGGGACAACTTTGTACAAAAAAGTTGGCATGCGGGGTGGTCTGCCACCCCG	GGGGACAACTTTGTACAAGAAAGTTGAGTAGGGCCGATCTCTGCG
ΔNLS	GGGGACAACTTTGTACAAAAAAGTTGGCATGGGAGGATTTTCCGGACCAGG	GGGGACAACTTTGTACAAGAAAGTTGATCCTCCACGTCCTCCTCTTCTTCC
ΔRGG3	GGGGACAACTTTGTACAAAAAAGTTGGCATGGGAGGATTTTCCGGACCAGG	GCCTTTATCCATTTTTCCAGGAGGCTTTGGGGCCTTACACTG
	CAGTGTAAGGCCCCAAAGCCTCCTGGAAAAATGGATAAAGGC	GGGGACAACTTTGTACAAGAAAGTTGAGTAGGGCCGATCTCTGCG
ΔZnF	GGGGACAACTTTGTACAAAAAAGTTGGCATGGGAGGATTTTCCGGACCAGG	GGGTGGCGGGAGGAAGCCTTCGTGCTGGACGTTTCCTCCTCC

	GGAGGAGGAAACGTCCAGCACGAAGGCTTCCTCCCGCCACCC	GGGGACAACTTTGTACAAGAAAGTTGAGTAGGGCCGATCTCTGCG
ΔZnF-RGG3	GGGGACAACTTTGTACAAAAAAGTTGGCATGGGAGGATTTTCCGGACCAGG	GCCTTTATCCATTTTTCCAGGGTGCTGGACGTTTCCTCCTCC
	GGAGGAGGAAACGTCCAGCACCCTGGAAAAATGGATAAAGGC	GGGGACAACTTTGTACAAGAAAGTTGAGTAGGGCCGATCTCTGCG
ΔRGG2	GGGGACAACTTTGTACAAAAAAGTTGGCATGGGAGGATTTTCCGGACCAGG	GGGACACTGCCAGTCTCCAGCTCGACTGTTCATTGGAGGCTTCTTCCG
	CGGAAGAAGCCTCCAATGAACAGTCGAGCTGGAGACTGGCAGTGTCCC	GGGGACAACTTTGTACAAGAAAGTTGAGTAGGGCCGATCTCTGCG
ΔRGG2-ZnF	GGGGACAACTTTGTACAAAAAAGTTGGCATGGGAGGATTTTCCGGACCAGG	GGGTGGCGGGAGGAAGCCTTCACTGTTCATTGGAGGCTTCTTCCG
	CGGAAGAAGCCTCCAATGAACAGTGAAGGCTTCCTCCCGCCACCC	GGGGACAACTTTGTACAAGAAAGTTGAGTAGGGCCGATCTCTGCG
ΔRGG2-ZnF-RGG3	GGGGACAACTTTGTACAAAAAAGTTGGCATGGGAGGATTTTCCGGACCAGG	GCCTTTATCCATTTTTCCAGGACTGTTCATTGGAGGCTTCTTCCG
	GGGGACAACTTTGTACAAAAAAGTTGGCATGGCGTCCACGGATTACAGTACC	GGGGACAACTTTGTACAAGAAAGTTGAGTAGGGCCGATCTCTGCG
EWSR1 ΔRGG2-ZnF-RGG3	GGGGACAACTTTGTACAAAAAAGTTGGCATGGCGTCCACGGATTACAGTACC	GCCTTTATCCATTTTTCCAGGACTGTTCATTGGAGGCTTCTTCCG
	CGGAAGAAGCCTCCAATGAACAGTCCTGGAAAAATGGATAAAGGC	GGGGACAACTTTGTACAAGAAAGTTGAGTAGGGCCGATCTCTGCG

Table 9: Cloning primers used in this work.

Target gene	Forward	Reverse	Ref.
FADS1	CAGCTTTGAGCCCACCAAGAA	AGCAAGATGTGCAGCAGGTACAG	https://www.nature.com/articles/s41598-019-46461-9
FADS2	TCATGACCATGATCGTCCATAAGAA	GCTCCCAGGATGCCG TAGAA	https://www.nature.com/articles/s41598-019-46461-9
ELOVL5	TGAGGCAGTGGTCAAACAGGTA	AGATATGTCATGAGTGGTTCCAAGA	https://www.nature.com/articles/s41598-019-46461-9
ELOVL6	GTGGTCGGCACCTAATG	AACCCTGGTCACAAACTG	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5393602/
LPCAT1	CGACTGAGCGCCCTGCAGAA	AAGGGCCAGGCCAGCAGCAT	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3421133/
LAMP1	AATGTCTGCAGCTCAAGGGC	CAGGATCACCCCGAATGTCA	NA
SLC39A1	CTGGCTGTGGGATCCTCTTC	CCATGCCCTCTAGCACAGAC	NA
HSD17B12	GGTCTTGAAATCGGCATCTTAGT	ATGTTCAGAATAGCCCCTTTGG	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7754038/#jcmm16026-sup-0001

SEC61G	AAAGGACTCCATTCGGCTGGTT	CAAAGAAGCCAATGAATCCC	https://www.nature.com/articles/s41419-021-03797-3
KDELR1	CTTCTACTCTGTGGCCCCAAT	AAAGATCTTGGACCCTGCCC	NA
ZNF664	AAAGCCACAATCATTGCCCG	GAGCTCCTCAAGGGGGAGTA	NA
RPL32	CATCTCCTTCTCGGCATCA	AACCCTGTTGTCAATGCCTC	NA
GAPDH	TTGCCATCAATGACCCCTTCA	CGCCCCACTTGATTTTGGA	NA
18S	AACCCGTTGAACCCCATT	CCATCCAATCGGTAGTAGCG	NA

Table 10: RIP-qPCR primers used in this work. NA = not available/applicable.

6 APPENDIX

6.1 Plasmids map

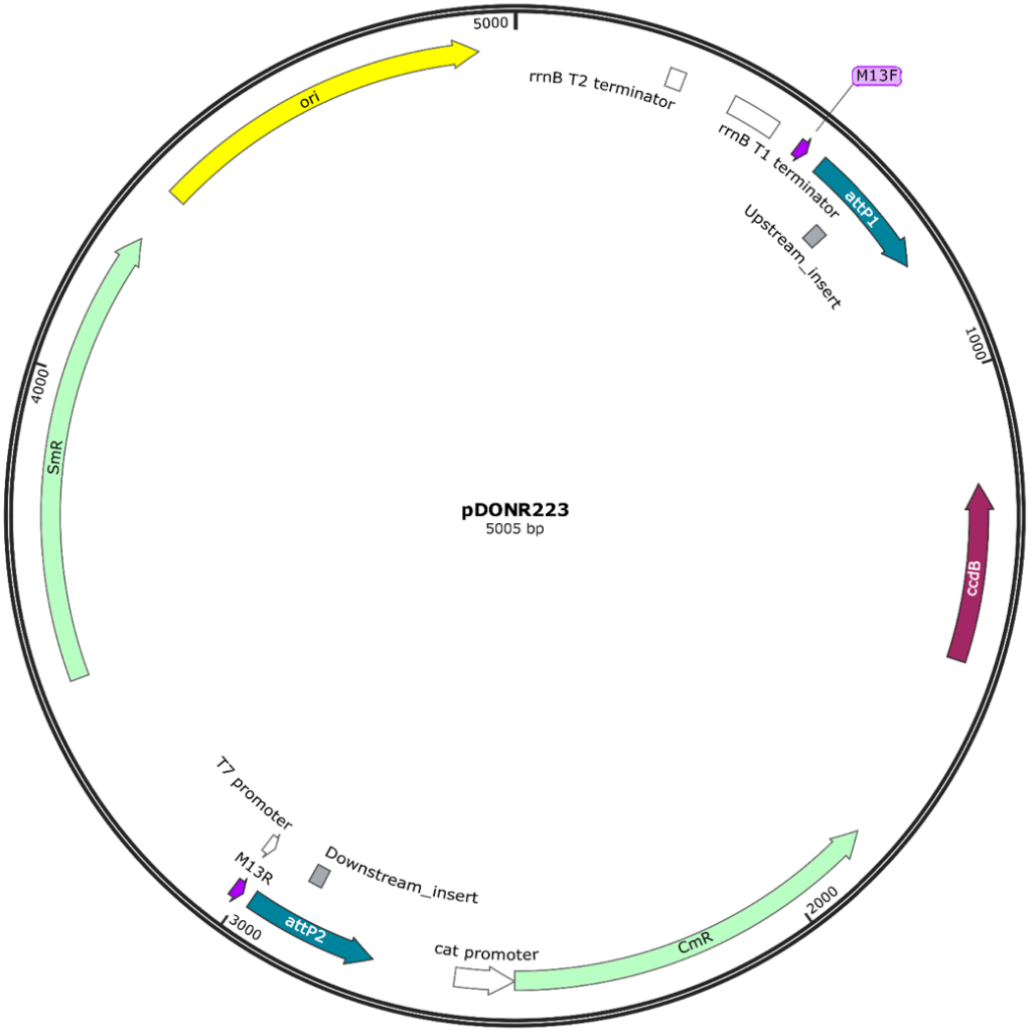


Figure 104: Annotated plasmid map of the pDONR223 vector. attP1 and attP2: Gateway sites for BP cloning. SmR: spectinomycin resistance gene. CmR: chloramphenicol resistance gene. Ori: bacterial origin of replication. From thesis of B.Galvan.

Figure 105: Annotated plasmid map of the pDEST1899 vector with 3 x FLAG tag. attR1 and attR2: Gateway sites for LR cloning. AmpR: ampicillin resistance gene. CmR: chloramphenicol resistance gene. Ori: bacterial origin of replication. From thesis of B.Galvan.

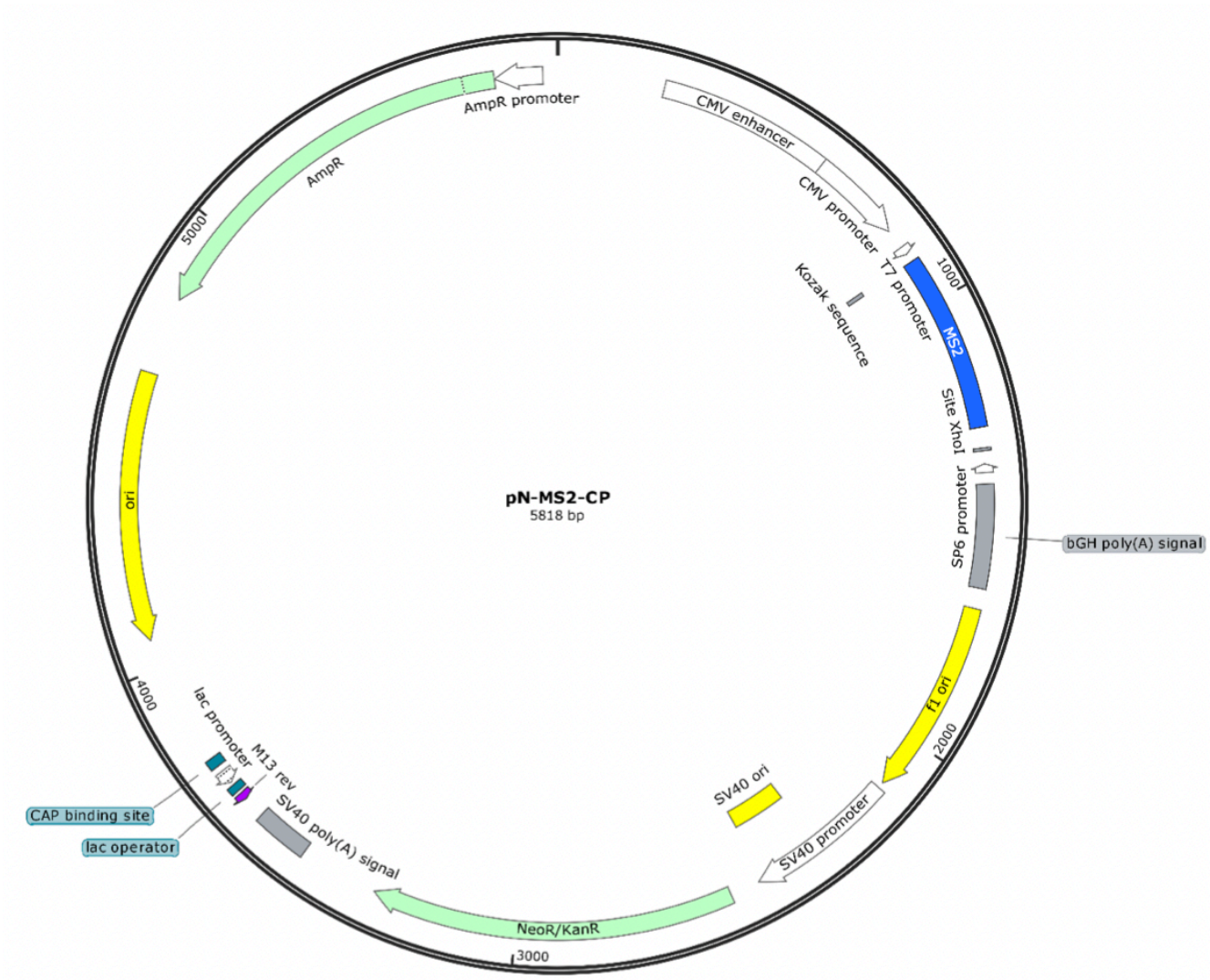


Figure 106: Annotated plasmid map of the pN-MS2-CP vector. AmpR: ampicillin resistance gene. NeoR/KanR neomycin/kanamycin resistance genes. Ori: bacterial origin of replication. From thesis of B.Galvan.

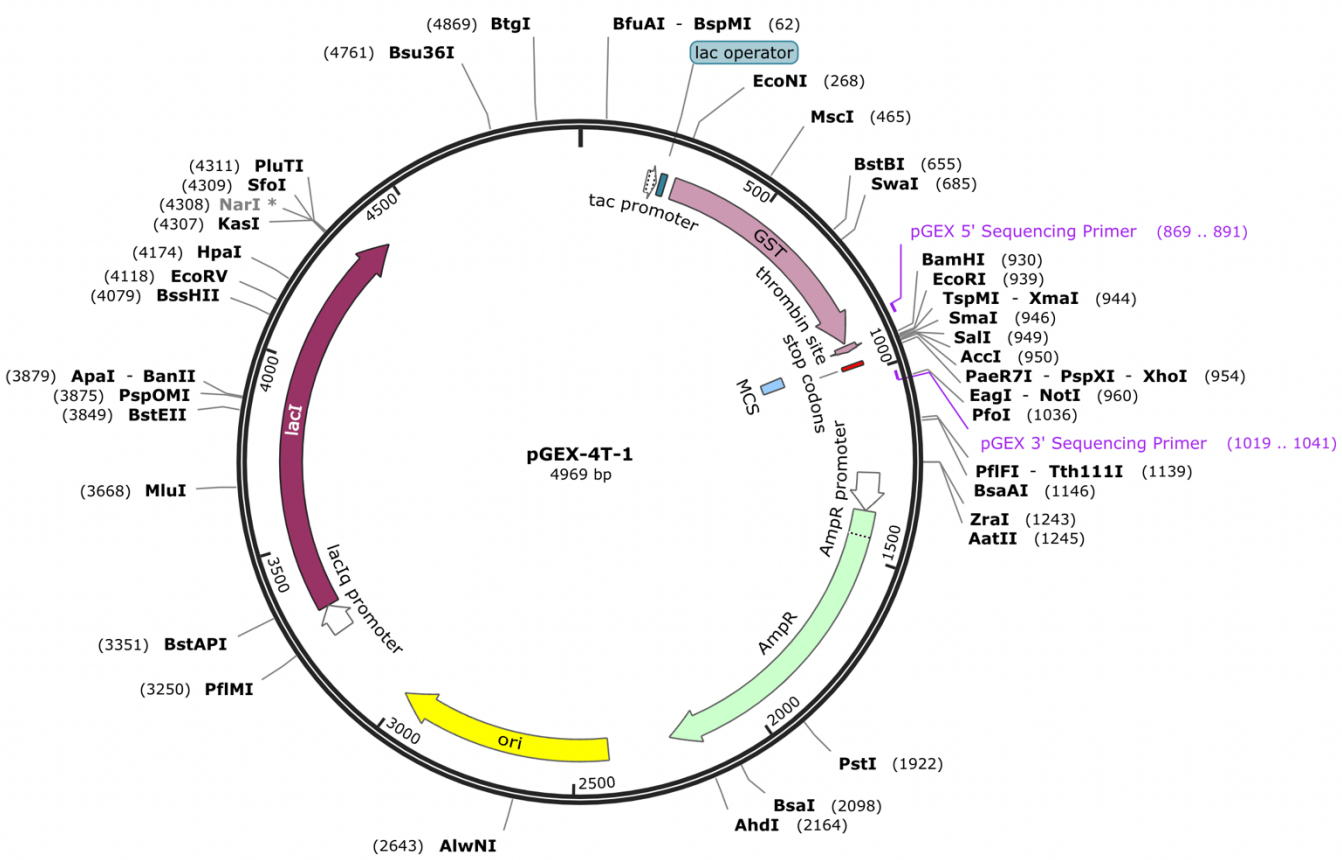


Figure 107: Annotated plasmid map of the pGEX-4T-1 vector. From SnapGene.com.

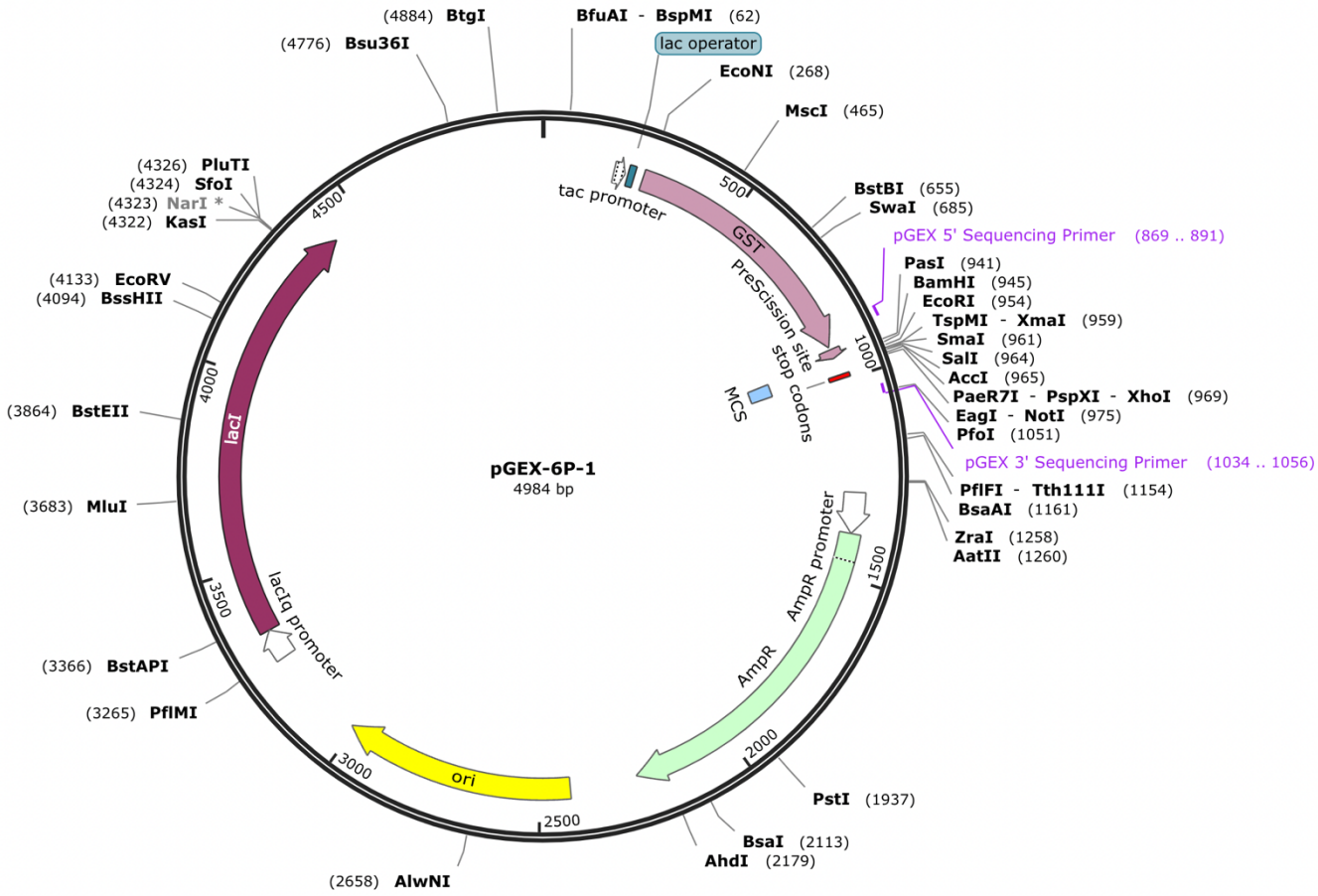


Figure 108: Annotated plasmid map of the pGEX-6P-1 vector. From SnapGene.com.

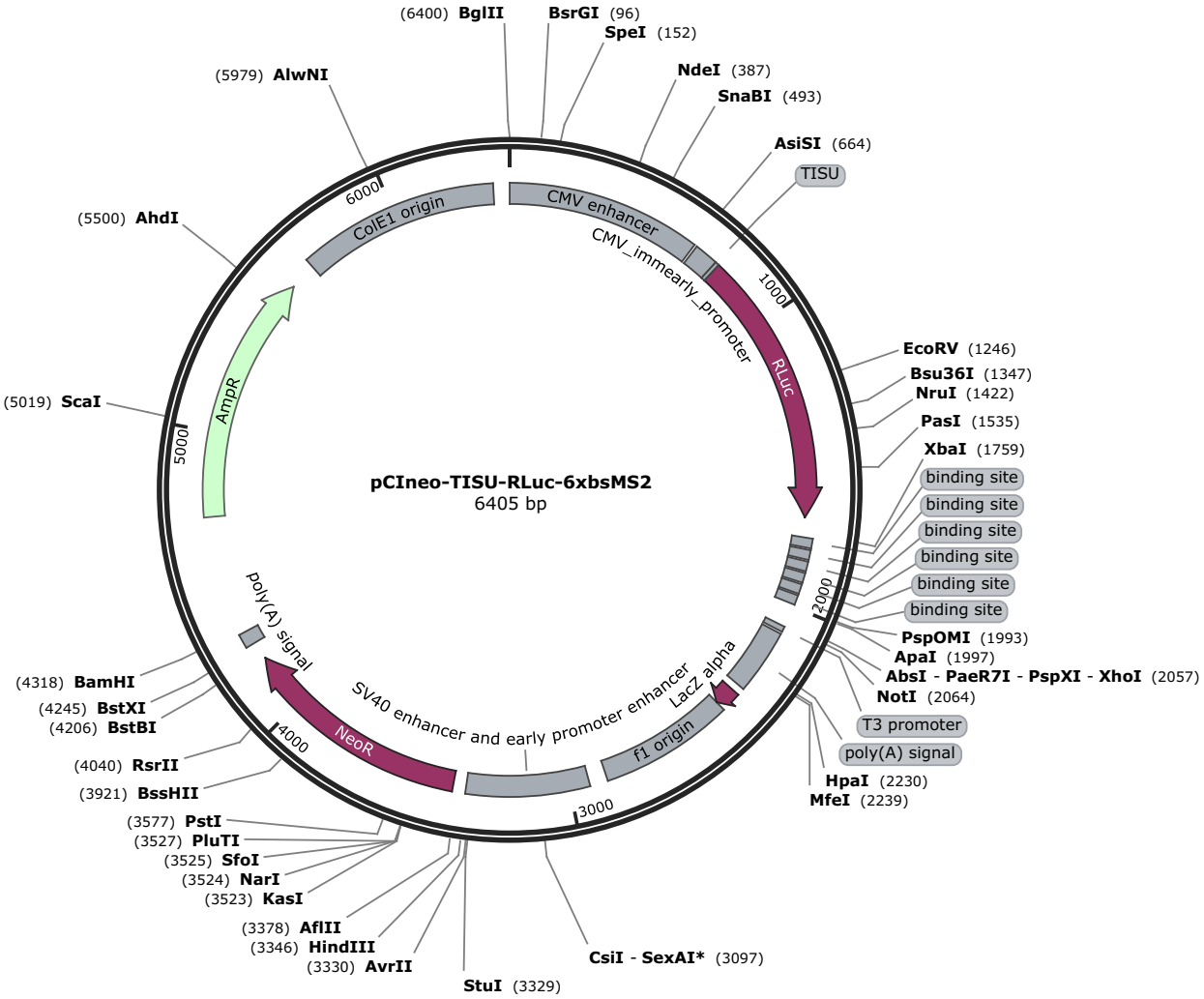


Figure 109: Annotated plasmid map of the pCIneo-TISU-RLuc-6xbsMS2 vector. From[608].

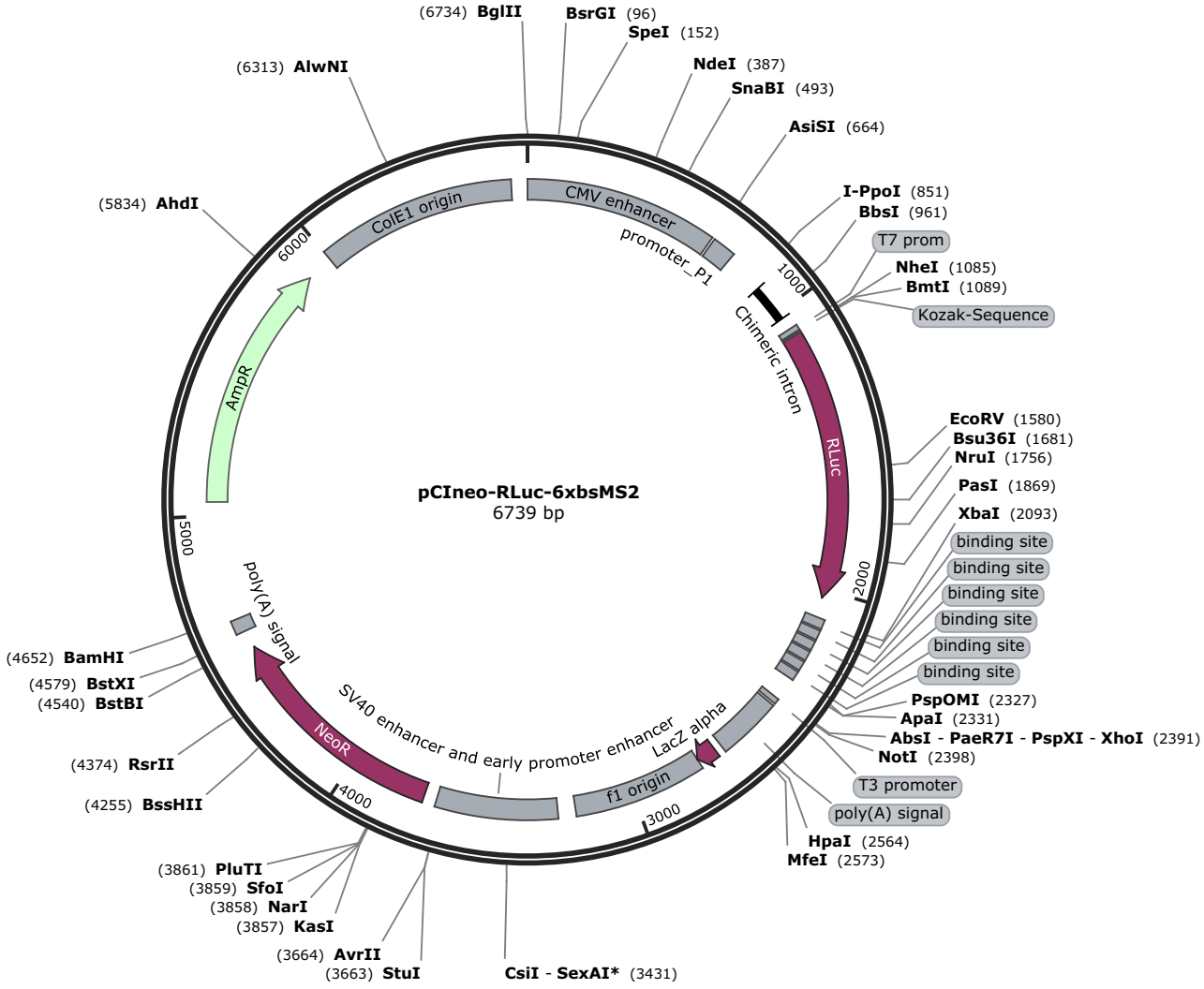


Figure 110: Annotated plasmid map of the pCIneo-RLuc-6xbsMS2 vector. From[608].

7 REFERENCES

1. Geuens, T., Bouhy, D. & Timmerman, V. The hnRNP family: insights into their role in health and disease. *Human Genetics* vol. 135 851–867 Preprint at https://doi.org/10.1007/s00439-016-1683-5 (2016).
2. Piñol-Roma, S. & Dreyfuss, G. Shuttling of pre-mRNA binding proteins between nucleus and cytoplasm. *Nature 1992 355:6362* **355**, 730–732 (1992).
3. Chaudhury, A., Chander, P. & Howe, P. H. Heterogeneous nuclear ribonucleoproteins (hnRNPs) in cellular processes: Focus on hnRNP E1's multifunctional regulatory roles. *RNA* **16**, 1449 (2010).
4. Bertolotti, A., Lutz, Y., Heard, D. J., Chambon, P. & Tora, L. hTAF(II)68, a novel RNA/ssDNA-binding protein with homology to the pro-oncoproteins TLS/FUS and EWS is associated with both TFIID and RNA polymerase II. *EMBO J* **15**, 5022–5031 (1996).
5. Tan, A. Y. & Manley, J. L. The TET family of proteins: functions and roles in disease. *J Mol Cell Biol* **1**, 82–92 (2009).
6. Andersson, M. K. *et al.* The multifunctional FUS, EWS and TAF15 proto-oncoproteins show cell type-specific expression patterns and involvement in cell spreading and stress response. *BMC Cell Biol* **9**, 37 (2008).
7. Lee, K. A. W. Molecular recognition by the EWS transcriptional activation domain. *Adv Exp Med Biol* **725**, 106–125 (2012).
8. Kovar, H. Dr. Jekyll and Mr. Hyde: The two faces of the FUS/EWS/TAF15 protein family. *Sarcoma* vol. 2011 13 Preprint at https://doi.org/10.1155/2011/837474 (2011).
9. Schwartz, J. C., Cech, T. R. & Parker, R. R. Biochemical Properties and Biological Functions of FET Proteins. *Annu Rev Biochem* **84**, 355–379 (2015).
10. Chau, B. L., Ng, K. P., Li, K. K. C. & Lee, K. A. W. RGG boxes within the TET/FET family of RNA-binding proteins are functionally distinct. *Transcription* **7**, 141–151 (2016).
11. Pahlich, S., Bschir, K., Chiavi, C., Belyanskaya, L. & Gehring, H. Different methylation characteristics of protein arginine methyltransferase 1 and 3 toward the Ewing Sarcoma protein and a peptide. *Proteins: Structure, Function and Genetics* **61**, 164–175 (2005).
12. Pahlich, S., Zakaryan, R. P. & Gehring, H. Identification of proteins interacting with protein arginine methyltransferase 8: The Ewing sarcoma (EWS) protein binds independent of its methylation state. *Proteins: Structure, Function and Genetics* **72**, 1125–1137 (2008).
13. Law, W. J., Cann, K. L. & Hicks, G. G. TLS, EWS and TAF15: A model for transcriptional integration of gene expression. *Brief Funct Genomic Proteomic* **5**, 8–14 (2006).
14. Rual, J. F. *et al.* Towards a proteome-scale map of the human protein-protein interaction network. *Nature* **437**, 1173–1178 (2005).
15. Bertolotti, A., Bell, B. & Tora, L. The N-terminal domain of human TAF(II)68 displays transactivation and oncogenic properties. *Oncogene* **18**, 8000–8010 (1999).

16. Gascoyne, D. M., Thomas, G. R. & Latchman, D. S. 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 (2004).
17. Jobert, L. *et al.* Human U1 snRNA forms a new chromatin-associated snRNP with TAF15. *EMBO Rep* **10**, 494 (2009).
18. Yamazaki, T. *et al.* FUS-SMN protein interactions link the motor neuron diseases ALS and SMA. *Cell Rep* **2**, 799 (2012).
19. Ballarino, M. *et al.* TAF15 is important for cellular proliferation and regulates the expression of a subset of cell cycle genes through miRNAs. *Oncogene 2013 32:39* **32**, 4646–4655 (2012).
20. Paronetto, M. P., Miñana, B. & Valcárcel, J. The Ewing Sarcoma Protein Regulates DNA Damage-Induced Alternative Splicing. *Mol Cell* **43**, 353–368 (2011).
21. Ishigaki, S. *et al.* Position-dependent FUS-RNA interactions regulate alternative splicing events and transcriptions. *Sci Rep* **2**, (2012).
22. Kim, K. Y. *et al.* A multifunctional protein EWS regulates the expression of Drosha and microRNAs. *Cell Death Differ* **21**, 136–145 (2014).
23. Gregory, R. I. *et al.* The Microprocessor complex mediates the genesis of microRNAs. *Nature* **432**, 235–240 (2004).
24. Morlando, M. *et al.* FUS stimulates microRNA biogenesis by facilitating co-transcriptional Drosha recruitment. *EMBO J* **31**, 4502 (2012).
25. Huang, L., Kuwahara, I. & Matsumoto, K. EWS represses cofilin 1 expression by inducing nuclear retention of cofilin 1 mRNA. *Oncogene* **33**, 2995–3003 (2014).
26. Velázquez-Cruz, A., Baños-Jaime, B., Díaz-Quintana, A., de la Rosa, M. A. & Díaz-Moreno, I. Post-translational Control of RNA-Binding Proteins and Disease-Related Dysregulation. *Front Mol Biosci* **8**, 234 (2021).
27. Gehring, H., Leemann-Zakaryan, R. P., Pahlich, S. & Grossenbacher, D. Tyrosine phosphorylation in the C-terminal nuclear localization and retention signal (C-NLS) of the EWS protein. *Sarcoma* **2011**, (2011).
28. Dormann, D. *et al.* Arginine methylation next to the PY-NLS modulates Transportin binding and nuclear import of FUS. *EMBO Journal* **31**, 4258–4275 (2012).
29. Sévigny, M. *et al.* FUS contributes to mTOR-dependent inhibition of translation. *Journal of Biological Chemistry* jbc.RA120.013801 (2020) doi:10.1074/jbc.ra120.013801.
30. Udagawa, T. *et al.* FUS regulates AMPA receptor function and FTLD/ALS-associated behaviour via GluA1 mRNA stabilization. *Nat Commun* **6**, (2015).
31. Kawaguchi, T. *et al.* Changes to the TDP-43 and FUS Interactomes Induced by DNA Damage. *J Proteome Res* **19**, 360–370 (2020).
32. Lee, S. *et al.* Ewing sarcoma protein promotes dissociation of poly(ADP-ribose) polymerase 1 from chromatin. *EMBO Rep* **21**, (2020).
33. Uversky, V. N. Intrinsically disordered proteins and their 'Mysterious' (meta)physics. *Frontiers in Physics* vol. 7 10 Preprint at https://doi.org/10.3389/fphy.2019.00010 (2019).

34. Uversky, V. N., Oldfield, C. J. & Dunker, A. K. Intrinsically disordered proteins in human diseases: Introducing the D 2 concept. *Annual Review of Biophysics* vol. 37 215–246 Preprint at https://doi.org/10.1146/annurev.biophys.37.032807.125924 (2008).
35. Garg, A., Dabburu, G. R., Singhal, N. & Kumar, M. Investigating the disordered regions (MoRFs, SLiMs and LCRs) and functions of mimicry proteins/peptides in silico. *PLoS One* **17**, e0265657 (2022).
36. Dobra, I., Pankivskyi, S., Samsonova, A., Pastre, D. & Hamon, L. Relation Between Stress Granules and Cytoplasmic Protein Aggregates Linked to Neurodegenerative Diseases. *Current Neurology and Neuroscience Reports* vol. 18 Preprint at https://doi.org/10.1007/s11910-018-0914-7 (2018).
37. Cascarina, S. M., Elder, M. R. & Ross, E. D. Atypical structural tendencies among low-complexity domains in the protein data bank proteome. *PLoS Comput Biol* **16**, (2020).
38. Ng, K. P. *et al.* Multiple aromatic side chains within a disordered structure are critical for transcription and transforming activity of EWS family oncoproteins. *Proc Natl Acad Sci U S A* **104**, 479–484 (2007).
39. Brocca, S., Grandori, R., Longhi, S. & Uversky, V. Liquid–Liquid Phase Separation by Intrinsically Disordered Protein Regions of Viruses: Roles in Viral Life Cycle and Control of Virus–Host Interactions. *Int J Mol Sci* **21**, 1–31 (2020).
40. Monahan, Z. *et al.* Phosphorylation of the FUS low-complexity domain disrupts phase separation, aggregation, and toxicity. *EMBO J* **36**, 2951–2967 (2017).
41. Nosella, M. L. *et al.* O-Linked- N-Acetylglucosaminylation of the RNA-Binding Protein EWS N-Terminal Low Complexity Region Reduces Phase Separation and Enhances Condensate Dynamics. *J Am Chem Soc* **143**, 11520–11534 (2021).
42. Hofweber, M. *et al.* Phase Separation of FUS Is Suppressed by Its Nuclear Import Receptor and Arginine Methylation. *Cell* **173**, 706-719.e13 (2018).
43. Tanikawa, C. *et al.* Citrullination of RGG Motifs in FET Proteins by PAD4 Regulates Protein Aggregation and ALS Susceptibility. *Cell Rep* **22**, 1473–1483 (2018).
44. Kwon, I. *et al.* Phosphorylation-Regulated Binding of RNA Polymerase II to Fibrous Polymers of Low-Complexity Domains. (2013) doi:10.1016/j.cell.2013.10.033.
45. Yasuhara, T. *et al.* Condensates induced by transcription inhibition localize active chromatin to nucleoli. *Mol Cell* **82**, 2738-2753.e6 (2022).
46. Franzmann, T. M. & Alberti, S. Prion-like low-complexity sequences: Key regulators of protein solubility and phase behavior. *Journal of Biological Chemistry* vol. 294 7128–7136 Preprint at https://doi.org/10.1074/jbc.TM118.001190 (2019).
47. Plougastel, B., Zucman, J., Peter, M., Thomas, G. & Delattre, O. Genomic structure of the EWS gene and its relationship to EWSR1, a site of tumor-associated chromosome translocation. *Genomics* **18**, 609–15 (1993).
48. Kovar, H. *et al.* The EWS protein is dispensable for Ewing tumor growth. *undefined* (2001).

49. Takahama, K., Kino, K., Arai, S., Kurokawa, R. & Oyoshi, T. Identification of Ewing's sarcoma protein as a G-quadruplex DNA- and RNA-binding protein. *FEBS Journal* **278**, 988–998 (2011).
50. Luo, Y. *et al.* EWS and FUS bind a subset of transcribed genes encoding proteins enriched in RNA regulatory functions. *BMC Genomics* **16**, (2015).
51. Hoell, J. I. *et al.* RNA targets of wild-type and mutant FET family proteins. *Nat Struct Mol Biol* **18**, 1428–1431 (2011).
52. Paronetto, M. P. *et al.* Regulation of FAS exon definition and apoptosis by the ewing sarcoma protein. *Cell Rep* **7**, 1211–1226 (2014).
53. Ouyang, H. *et al.* The RNA binding protein EWS is broadly involved in the regulation of pri-miRNA processing in mammalian cells. *Nucleic Acids Res* **45**, 12481–12495 (2017).
54. Deloulme, J. C., Prichard, L., Delattre, O. & Storm, D. R. The prooncoprotein EWS binds calmodulin and is phosphorylated by protein kinase C through an IQ domain. *Journal of Biological Chemistry* **272**, 27369–27377 (1997).
55. Zhang, D., Paley, A. J. & Childs, G. *The Transcriptional Repressor ZFM1 Interacts with and Modulates the Ability of EWS to Activate Transcription**. http://www.jbc.org/ (1998).
56. Song, J., Ng, S. C., Tompa, P., Lee, K. A. W. & Chan, H. S. Polycation-π Interactions Are a Driving Force for Molecular Recognition by an Intrinsically Disordered Oncoprotein Family. *PLoS Comput Biol* **9**, e1003239 (2013).
57. Petermann, R. *et al.* Oncogenic EWS-Fli1 interacts with hsRPB7, a subunit of human RNA polymerase II. *Oncogene* **17**, 603–610 (1998).
58. Bertolotti, A. *et al.* 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 hTAF II 68, and Subunits of TFIID and RNA Polymerase II Complexes . *Mol Cell Biol* **18**, 1489–1497 (1998).
59. Fujimura, Y., Siddique, H., Lee, L., Rao, V. N. & Reddy, E. S. P. EWS-ATF-1 chimeric protein in soft tissue clear cell sarcoma associates with CREB-binding protein and interferes with p53-mediated trans-activation function. *Oncogene* **20**, 6653–6659 (2001).
60. Alex, D. & Lee, K. A. W. W. RGG-boxes of the EWS oncoprotein repress a range of transcriptional activation domains. *Nucleic Acids Res* **33**, 1323–1331 (2005).
61. Järvelin, A. I., Noerenberg, M., Davis, I. & Castello, A. The new (dis)order in RNA regulation. *Cell Communication and Signaling* vol. 14 1–22 Preprint at https://doi.org/10.1186/s12964-016-0132-3 (2016).
62. Thandapani, P., O'Connor, T. R., Bailey, T. L. & Richard, S. Defining the RGG/RG Motif. *Molecular Cell* vol. 50 613–623 Preprint at https://doi.org/10.1016/j.molcel.2013.05.021 (2013).
63. Shaw, D. J. *et al.* Identification of a self-association domain in the Ewing's sarcoma protein: a novel function for arginine-glycine-glycine rich motifs? *J Biochem* **147**, 885–93 (2010).

64. Belyanskaya, L. L., Gehrig, P. M. & Gehring, H. Exposure on Cell Surface and Extensive Arginine Methylation of Ewing Sarcoma (EWS) Protein. *Journal of Biological Chemistry* **276**, 18681–18687 (2001).
65. Kim, J.-D., Kako, K., Kakiuchi, M., Park, G. G. & Fukamizu, A. EWS is a substrate of type I protein arginine methyltransferase, PRMT8. *Int J Mol Med* **22**, 309–15 (2008).
66. Ozdilek, B. A. *et al.* Intrinsically disordered RGG/RG domains mediate degenerate specificity in RNA binding. *Nucleic Acids Res* **45**, 7984–7996 (2017).
67. Tantin, D. Oct transcription factors in development and stem cells: Insights and mechanisms. *Development (Cambridge)* **140**, 2857–2866 (2013).
68. Lee, J., Rhee, B. K., Bae, G.-Y., Han, Y.-M. & Kim, J. Stimulation of Oct-4 Activity by Ewing's Sarcoma Protein. *Stem Cells* **23**, 738–751 (2005).
69. Thomas, G. R. & Latchman, D. S. The pro-oncoprotein EWS (Ewing's sarcoma protein) interacts with the Brn-3a POU transcription factor and inhibits its ability to activate transcription. *Cancer Biol Ther* **1**, 428–432 (2002).
70. Boulay, G. *et al.* Cancer-Specific Retargeting of BAF Complexes by a Prion-like Domain. *Cell* **171**, 163-178.e19 (2017).
71. Dancy, B. M. & Cole, P. A. Protein lysine acetylation by p300/CBP. *Chemical Reviews* vol. 115 2419–2452 Preprint at https://doi.org/10.1021/cr500452k (2015).
72. Rossow, K. L. & Janknecht, R. The Ewing's sarcoma gene product functions as a transcriptional activator. *Cancer Res* **61**, 2690–2695 (2001).
73. Araya, N. *et al.* Cooperative interaction of EWS with CREB-binding protein selectively activates hepatocyte nuclear factor 4-mediated transcription. *Journal of Biological Chemistry* **278**, 5427–5432 (2003).
74. Chatterjee, N. & Walker, G. C. Mechanisms of DNA damage, repair, and mutagenesis. *Environmental and Molecular Mutagenesis* vol. 58 235–263 Preprint at https://doi.org/10.1002/em.22087 (2017).
75. Naro, C., Bielli, P., Pagliarini, V. & Sette, C. The interplay between DNA damage response and RNA processing: The unexpected role of splicing factors as gatekeepers of genome stability. *Front Genet* **6**, 142 (2015).
76. Li, H. *et al.* Ewing sarcoma gene EWS is essential for meiosis and B lymphocyte development. *Journal of Clinical Investigation* **117**, 1314–1323 (2007).
77. Hurov, K. E., Cotta-Ramusino, C. & Elledge, S. J. A genetic screen identifies the Triple T complex required for DNA damage signaling and ATM and ATR stability. *Genes Dev* **24**, 1939–1950 (2010).
78. O'Connell, B. C. *et al.* A Genome-wide Camptothecin Sensitivity Screen Identifies a Mammalian MMS22L-NFKBIL2 Complex Required for Genomic Stability. *Mol Cell* **40**, 645–657 (2010).
79. Klevernic, I. v., Morton, S., Davis, R. J. & Cohen, P. Phosphorylation of Ewing's sarcoma protein (EWS) and EWS-Fli1 in response to DNA damage. *Biochemical Journal* **418**, 625–634 (2009).

80. Dutertre, M. *et al.* Cotranscriptional exon skipping in the genotoxic stress response. *Nat Struct Mol Biol* **17**, 1358–1366 (2010).
81. Dominguez-Brauer, C. *et al.* Targeting Mitosis in Cancer: Emerging Strategies. *Molecular Cell* vol. 60 524–536 Preprint at https://doi.org/10.1016/j.molcel.2015.11.006 (2015).
82. Azuma, M., Embree, L. J., Sabaawy, H. & Hickstein, D. D. Ewing Sarcoma Protein Ewsr1 Maintains Mitotic Integrity and Proneural Cell Survival in the Zebrafish Embryo. *PLoS One* **2**, e979 (2007).
83. Spahn, L. *et al.* Interaction of the EWS NH2 terminus with BARD1 links the Ewing's sarcoma gene to a common tumor suppressor pathway. *Cancer Res* **62**, 4583–7 (2002).
84. Tarsounas, M. & Sung, P. The antitumorigenic roles of BRCA1–BARD1 in DNA repair and replication. *Nature Reviews Molecular Cell Biology* 1–16 Preprint at https://doi.org/10.1038/s41580-020-0218-z (2020).
85. Park, H. *et al.* Ewing sarcoma EWS protein regulates midzone formation by recruiting Aurora B kinase to the midzone. *Cell Cycle* **13**, 2391–2399 (2014).
86. Wang, Y. L. *et al.* EWSR1 regulates mitosis by dynamically influencing microtubule acetylation. *Cell Cycle* **15**, 2202–2215 (2016).
87. Kishore, S. *et al.* 3′-UTR Poly(T/U) tract deletions and altered expression of EWSR1 are a hallmark of mismatch repair-deficient cancers. *Cancer Res* **74**, 224–234 (2014).
88. Shi, Y. The Spliceosome: A Protein-Directed Metalloribozyme. *Journal of Molecular Biology* vol. 429 2640–2653 Preprint at https://doi.org/10.1016/j.jmb.2017.07.010 (2017).
89. Rappsilber, J., Ryder, U., Lamond, A. I. & Mann, M. Large-scale proteomic analysis of the human spliceosome. *Genome Res* **12**, 1231–1245 (2002).
90. Knoop, L. L. & Baker, S. J. The splicing factor U1C represses EWS/FLI-mediated transactivation. *Journal of Biological Chemistry* **275**, 24865–24871 (2000).
91. Wang, H. J. *et al.* Using the yeast two-hybrid assay to discover protein partners for the leucine-rich amelogenin peptide and for tuftelin-interacting protein 11. in *European Journal of Oral Sciences* vol. 114 276–279 (Eur J Oral Sci, 2006).
92. Wen, X. *et al.* Structural organization and cellular localization of tuftelin-interacting protein 11 (TFIP11). *Cellular and Molecular Life Sciences* **62**, 1038–1046 (2005).
93. Chansky, H. A., Hu, M., Hickstein, D. D. & Yang, L. Oncogenic TLS/ERG and EWS/Fli-1 fusion proteins inhibit RNA splicing mediated by YB-1 protein. *undefined* (2001).
94. Yang, L., Chansky, H. A. & Hickstein, D. D. EWS·Fli-1 fusion protein interacts with hyperphosphorylated RNA polymerase II and interferes with serine-arginine protein-mediated RNA splicing. *Journal of Biological Chemistry* **275**, 37612–37618 (2000).
95. Tannukit, S., Wen, X., Wang, H. J. & Paine, M. L. TFIP11, CCNL1 and EWSR1 protein-protein interactions, and their nuclear localization. *Int J Mol Sci* **9**, 1504–1514 (2008).
96. Young, P. J. *et al.* The Ewing's sarcoma protein interacts with the Tudor domain of the survival motor neuron protein. *Molecular Brain Research* **119**, 37–49 (2003).

97. Saldi, T., Cortazar, M. A., Sheridan, R. M. & Bentley, D. L. Coupling of RNA Polymerase II Transcription Elongation with Pre-mRNA Splicing. *Journal of Molecular Biology* vol. 428 2623–2635 Preprint at https://doi.org/10.1016/j.jmb.2016.04.017 (2016).
98. Fong, N. *et al.* Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. *Genes Dev* **28**, 2663 (2014).
99. Sanchez, G., Delattre, O., Auboeuf, D. & Dutertre, M. Coupled alteration of transcription and splicing by a single oncogene: Boosting the effect on cyclin D1 activity. *Cell Cycle* vol. 7 2299–2305 Preprint at https://doi.org/10.4161/cc.6445 (2008).
100. Zhang, P., Wu, W., Chen, Q. & Chen, M. Non-Coding RNAs and their Integrated Networks. *Journal of integrative bioinformatics* vol. 16 Preprint at https://doi.org/10.1515/jib-2019-0027 (2019).
101. Han, C., Sun, L. Y., Wang, W. T., Sun, Y. M. & Chen, Y. Q. Non-coding RNAs in cancers with chromosomal rearrangements: The signatures, causes, functions and implications. *J Mol Cell Biol* **11**, 886–898 (2019).
102. O'Brien, J., Hayder, H., Zayed, Y. & Peng, C. Overview of microRNA biogenesis, mechanisms of actions, and circulation. *Frontiers in Endocrinology* vol. 9 402 Preprint at https://doi.org/10.3389/fendo.2018.00402 (2018).
103. Kim, Y. *et al.* Uvrag targeting by Mir125a and Mir351 modulates autophagy associated with Ewsr1 deficiency. *Autophagy* **11**, 796–811 (2015).
104. Reinhart, B. J. *et al.* The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature* **403**, 901–906 (2000).
105. Sohn, E. J., Park, J., Kang, S. I. & Wu, Y. P. Accumulation of pre-let-7g and downregulation of mature let-7g with the depletion of EWS. *Biochem Biophys Res Commun* **426**, 89–93 (2012).
106. Ouyang, H. *et al.* The RNA binding protein EWS is broadly involved in the regulation of pri-miRNA processing in mammalian cells. *Nucleic Acids Res* **45**, 12481–12495 (2017).
107. Lee, C.-J. *et al.* Integrated analysis of omics data using microRNA-target mRNA network and PPI network reveals regulation of Gnai1 function in the spinal cord of Ews/Ewsr1 KO mice. *BMC Med Genomics* **9**, 33 (2016).
108. Wu, J. & Hann, S. S. Cellular Physiology and Biochemistry Cellular Physiology and Biochemistry Functions and Roles of Long-Non-Coding RNAs in Human Nasopharyngeal Carcinoma. *Cell Physiol Biochem* **45**, 1191–1204 (2018).
109. Sun, X., Haider Ali, M. S. S. & Moran, M. The role of interactions of long non-coding RNAs and heterogeneous nuclear ribonucleoproteins in regulating cellular functions. *Biochemical Journal* vol. 474 2925–2935 Preprint at https://doi.org/10.1042/BCJ20170280 (2017).
110. Oyoshi, T. & Kurokawa, R. Structure of noncoding RNA is a determinant of function of RNA binding proteins in transcriptional regulation. *Cell and Bioscience* vol. 2 1 Preprint at https://doi.org/10.1186/2045-3701-2-1 (2012).

111. Gawronski, A. R. *et al.* MechRNA: Prediction of lncRNA mechanisms from RNA–RNA and RNA–protein interactions. *Bioinformatics* **34**, 3101–3110 (2018).
112. Ernst, E. H., Nielsen, J., Ipsen, M. B., Villesen, P. & Lykke-Hartmann, K. Transcriptome analysis of long non-coding RNAs and genes encoding paraspeckle proteins during human ovarian follicle development. *Front Cell Dev Biol* **6**, 78 (2018).
113. Felsch, J. S., Lane, W. S. & Peralta, E. G. Tyrosine kinase Pyk2 mediates G-protein-coupled receptor regulation of the Ewing sarcoma RNA-binding protein EWS. *Current Biology* **9**, 485–490 (1999).
114. Andersen, J. S. *et al.* Nucleolar proteome dynamics. *Nature 2004 433:7021* **433**, 77–83 (2005).
115. Dubois, M. L. & Boisvert, F. M. The nucleolus: Structure and function. in *The Functional Nucleus* 29–49 (Springer International Publishing, 2016). doi:10.1007/978-3-319-38882-3_2.
116. Umegaki-Arao, N., Tamai, K., Nimura, K., Serada, S. & Naka, T. Karyopherin Alpha2 Is Essential for rRNA Transcription and Protein Synthesis in Proliferative Keratinocytes. *PLoS One* **8**, 76416 (2013).
117. Abraham, K. J. *et al.* Nucleolar RNA polymerase II drives ribosome biogenesis. *Nature* **585**, 298–302 (2020).
118. Wang, F., Li, J., Fan, S., Jin, Z. & Huang, C. Targeting stress granules: A novel therapeutic strategy for human diseases. *Pharmacological Research* vol. 161 105143 Preprint at https://doi.org/10.1016/j.phrs.2020.105143 (2020).
119. Svetoni, F., Frisone, P. & Paronetto, M. P. Role of FET proteins in neurodegenerative disorders. *RNA Biology* vol. 13 1089–1102 Preprint at https://doi.org/10.1080/15476286.2016.1211225 (2016).
120. Huang, L., Nakai, Y., Kuwahara, I. & Matsumoto, K. PRAS40 is a functionally critical target for EWS repression in Ewing sarcoma. *Cancer Res* **72**, 1260–1269 (2012).
121. López-Erauskin, J. *et al.* ALS/FTD-Linked Mutation in FUS Suppresses Intra-axonal Protein Synthesis and Drives Disease Without Nuclear Loss-of-Function of FUS. *Neuron* **100**, 816-830.e7 (2018).
122. Kamelgarn, M. *et al.* ALS mutations of FUS suppress protein translation and disrupt the regulation of nonsense-mediated decay. *Proc Natl Acad Sci U S A* **115**, E11904–E11913 (2018).
123. Lee, J. *et al.* EWSR1, a multifunctional protein, regulates cellular function and aging via genetic and epigenetic pathways. *Biochimica et Biophysica Acta - Molecular Basis of Disease* vol. 1865 1938–1945 Preprint at https://doi.org/10.1016/j.bbadis.2018.10.042 (2019).
124. Park, J. H. *et al.* A Multifunctional protein, EWS, is essential for early brown fat lineage determination. *Dev Cell* **26**, 393–404 (2013).
125. Park, J. H. *et al.* Inactivation of EWS reduces PGC-1α protein stability and mitochondrial homeostasis. *Proc Natl Acad Sci U S A* **112**, 6074–6079 (2015).

126. Cho, J. *et al.* Ewing sarcoma gene Ews regulates hematopoietic stem cell senescence. *Blood* **117**, 1156–1166 (2011).
127. Kim, Y. *et al.* Uvrag targeting by Mir125a and Mir351 modulates autophagy associated with Ewsr1 deficiency. *Autophagy* **11**, 796–811 (2015).
128. Ewing Sarcomas Phenocopy BRCA1-Deficient Tumors. *Cancer Discov* **8**, 527.2-527 (2018).
129. Anumanthan, G., Halder, S. K., Friedman, D. B. & Datta, P. K. Oncogenic serine-threonine kinase receptor-associated protein modulates the function of ewing sarcoma protein through a novel mechanism. *Cancer Res* **66**, 10824–10832 (2006).
130. Dugger, B. N. & Dickson, D. W. Pathology of neurodegenerative diseases. *Cold Spring Harbor Perspectives in Biology* vol. 9 Preprint at https://doi.org/10.1101/cshperspect.a028035 (2017).
131. Shang, Y. & Huang, E. J. Mechanisms of FUS mutations in familial amyotrophic lateral sclerosis. *Brain Research* vol. 1647 65–78 Preprint at https://doi.org/10.1016/j.brainres.2016.03.036 (2016).
132. Bitetto, G. & Di Fonzo, A. Nucleo-cytoplasmic transport defects and protein aggregates in neurodegeneration. *Translational Neurodegeneration* vol. 9 1–16 Preprint at https://doi.org/10.1186/s40035-020-00205-2 (2020).
133. Galzitskaya, O. V. Influence of Repeats in the Protein Chain on its Aggregation Capacity for ALS-Associated Proteins. in *Update on Amyotrophic Lateral Sclerosis* (InTech, 2016). doi:10.5772/63104.
134. Bose, S. & Cho, J. Targeting chaperones, heat shock factor-1, and unfolded protein response: Promising therapeutic approaches for neurodegenerative disorders. *Ageing Research Reviews* vol. 35 155–175 Preprint at https://doi.org/10.1016/j.arr.2016.09.004 (2017).
135. Zarei, S. *et al.* A comprehensive review of amyotrophic lateral sclerosis. *Surgical Neurology International* vol. 6 Preprint at https://doi.org/10.4103/2152-7806.169561 (2015).
136. Mejzini, R. *et al.* ALS Genetics, Mechanisms, and Therapeutics: Where Are We Now? *Frontiers in Neuroscience* vol. 13 1310 Preprint at https://doi.org/10.3389/fnins.2019.01310 (2019).
137. Hardiman, O. *et al.* Amyotrophic lateral sclerosis. *Nature Reviews Disease Primers* vol. 3 Preprint at https://doi.org/10.1038/nrdp.2017.71 (2017).
138. Lattante, S., Rouleau, G. A. & Kabashi, E. TARDBP and FUS Mutations Associated with Amyotrophic Lateral Sclerosis: Summary and Update. *Hum Mutat* **34**, 812–826 (2013).
139. Chen, C., Ding, X., Akram, N., Xue, S. & Luo, S. Z. Fused in sarcoma: Properties, self-assembly and correlation with neurodegenerative diseases. *Molecules* vol. 24 Preprint at https://doi.org/10.3390/molecules24081622 (2019).
140. Couthouis, J. *et al.* Evaluating the role of the FUS/TLS-related gene EWSR1 in amyotrophic lateral sclerosis. *Hum Mol Genet* **21**, 2899 (2012).

141. Moore, K. M. *et al.* Age at symptom onset and death and disease duration in genetic frontotemporal dementia: an international retrospective cohort study. *Lancet Neurol* **19**, 145–156 (2020).
142. Convery, R., Mead, S. & Rohrer, J. D. Review: Clinical, genetic and neuroimaging features of frontotemporal dementia. *Neuropathology and Applied Neurobiology* vol. 45 6–18 Preprint at https://doi.org/10.1111/nan.12535 (2019).
143. McKhann, G. M. *et al.* Clinical and Pathological Diagnosis of Frontotemporal Dementia: Report of the Work Group on Frontotemporal Dementia and Pick's Disease. *Arch Neurol* **58**, 1803–1809 (2001).
144. Van Mossevelde, S., Engelborghs, S., Van Der Zee, J. & Van Broeckhoven, C. Genotype-phenotype links in frontotemporal lobar degeneration. *Nature Reviews Neurology* vol. 14 363–378 Preprint at https://doi.org/10.1038/s41582-018-0009-8 (2018).
145. Chornenka, K. *et al.* Expanding the Phenotype of Frontotemporal Lobar Degeneration With FUS-Positive Pathology (FTLD-FUS). *J Neuropathol Exp Neurol* **79**, 809–812 (2020).
146. Mackenzie, I. R. A. & Neumann, M. Fused in sarcoma neuropathology in neurodegenerative disease. *Cold Spring Harb Perspect Med* **7**, 16 (2017).
147. Abramzon, Y. A., Fratta, P., Traynor, B. J. & Chia, R. The Overlapping Genetics of Amyotrophic Lateral Sclerosis and Frontotemporal Dementia. *Frontiers in Neuroscience* vol. 14 42 Preprint at https://doi.org/10.3389/fnins.2020.00042 (2020).
148. Chunling, W. & Zheng, X. Review on clinical update of essential tremor. *Neurological Sciences* vol. 37 495–502 Preprint at https://doi.org/10.1007/s10072-015-2380-1 (2016).
149. Labbé, C. *et al.* Investigating FUS variation in Parkinson's disease. *Parkinsonism Relat Disord* **20**, S147 (2014).
150. Yoon, Y. *et al.* Genetic ablation of EWS RNA binding protein 1 (EWSR1) leads to neuroanatomical changes and motor dysfunction in mice. *Exp Neurobiol* **27**, 103–111 (2018).
151. Perry, J. A., Seong, B. K. A. & Stegmaier, K. Biology and therapy of dominant fusion oncoproteins involving transcription factor and chromatin regulators in sarcomas. *Annu Rev Cancer Biol* **3**, 299–321 (2019).
152. Burningham, Z., Hashibe, M., Spector, L. & Schiffman, J. D. The Epidemiology of Sarcoma. *Clin Sarcoma Res* **2**, 14 (2012).
153. Gounder, M. M. *et al.* Clinical genomic profiling in the management of patients with soft tissue and bone sarcoma. *Nature Communications 2022 13:1* **13**, 1–15 (2022).
154. Malkin, D. *et al.* Germ Line p53 Mutations in a Familial Syndrome of Breast Cancer, Sarcomas, and Other Neoplasms. *Science (1979)* **250**, 1233–1238 (1990).
155. Hansen, M. F. *et al.* Osteosarcoma and retinoblastoma: a shared chromosomal mechanism revealing recessive predisposition. *Proc Natl Acad Sci U S A* **82**, 6216 (1985).
156. Ewing, J. Diffuse Endothelioma of Bone. *CA Cancer J Clin* **22**, 95–98 (1972).
157. Grünewald, T. G. P. *et al.* Ewing sarcoma. *Nature Reviews Disease Primers* vol. 4 Preprint at https://doi.org/10.1038/s41572-018-0003-x (2018).

158. K., N. *et al.* Childhood cancer incidence and survival in Japan and England: A population-based study (1993-2010). *Cancer Sci* **109**, 422–434 (2018).
159. Vidya Rani, P. S., Shyamala, K., Girish, H. C. & Murgod, S. Pathogenesis of Ewing sarcoma: A review. *Journal of Advanced Clinical & Research Insights* • **2**, 164–168 (2015).
160. Riggi, N., Suvà, M. L. & Stamenkovic, I. Ewing's Sarcoma. *New England Journal of Medicine* **384**, 154–164 (2021).
161. Bernstein, M. *et al.* Ewing's Sarcoma Family of Tumors: Current Management. *Oncologist* **11**, 503–519 (2006).
162. Hayden, R. & Leggas, M. Novel Drug Treatments for Ewing Sarcoma. *Curr Mol Biol Rep* **5**, 153–166 (2019).
163. Marina, N. M. *et al.* Longitudinal follow-up of adult survivors of Ewing sarcoma: A report from the Childhood Cancer Survivor Study. *Cancer* **123**, 2551–2560 (2017).
164. Town, J. *et al.* Exploring the surfaceome of Ewing sarcoma identifies a new and unique therapeutic target. *Proc Natl Acad Sci U S A* **113**, 3603–3608 (2016).
165. Tu, J. *et al.* The histogenesis of Ewing Sarcoma. *Cancer Rep Rev* **1**, (2017).
166. Lawlor, E. R. & Sorensen, P. H. Twenty years on: What do we really know about ewing sarcoma and what is the path forward? *Crit Rev Oncog* **20**, 155–171 (2015).
167. Riggi, N. *et al.* EWS-FLI-1 modulates miRNA145 and SOX2 expression to initiate mesenchymal stem cell reprogramming toward Ewing sarcoma cancer stem cells. *Genes Dev* **24**, 916–932 (2010).
168. von Levetzow, C. *et al.* Modeling initiation of ewing sarcoma in human neural crest cells. *PLoS One* **6**, (2011).
169. Castillero-Trejo, Y., Eliazer, S., Xiang, L., Richardson, J. A. & Ilaria, R. L. Expression of the EWS/FLI-1 oncogene in murine primary bone-derived cells results in EWS/FLI-1-dependent, Ewing sarcoma-like tumors. *Cancer Res* **65**, 8698–8705 (2005).
170. Tanaka, M. *et al.* Ewing's sarcoma precursors are highly enriched in embryonic osteochondrogenic progenitors. *Journal of Clinical Investigation* **124**, 3061–3074 (2014).
171. Jeon, I. S. *et al.* A variant Ewing's sarcoma translocation (7;22) fuses the EWS gene to the ETS gene ETV1. *Oncogene* **10**, 1229–34 (1995).
172. Kaneko, Y. *et al.* Fusion of an ETS-family gene, EIAF, to EWS by t(17;22)(q12;q12) chromosome translocation in an undifferentiated sarcoma of infancy. *Genes Chromosomes Cancer* **15**, 115–21 (1996).
173. Peter, M. *et al.* A new member of the ETS family fused to EWS in Ewing tumors. *Oncogene* **14**, 1159–1164 (1997).
174. Urano, F., Umezawa, A., Hong, W., Kikuchi, H. & Hata, J. I. A novel chimera gene between EWS and E1A-F, encoding the adenovirus E1A enhancer-binding protein, in extraosseous Ewing's sarcoma. *Biochem Biophys Res Commun* **219**, 608–612 (1996).
175. Oikawa, T. & Yamada, T. Molecular biology of the Ets family of transcription factors. *Gene* vol. 303 11–34 Preprint at https://doi.org/10.1016/S0378-1119(02)01156-3 (2003).

176. Jo, V. Y. EWSR1 fusions: Ewing sarcoma and beyond. *Cancer Cytopathol* (2020) doi:10.1002/cncy.22239.
177. Anderson, N. D. *et al.* Rearrangement bursts generate canonical gene fusions in bone and soft tissue tumors. *Science* **361**, (2018).
178. Pellestor, F. Chromoanagenesis: cataclysms behind complex chromosomal rearrangements. *Molecular Cytogenetics 2019 12:1* **12**, 1–12 (2019).
179. Delattre, O. *et al.* Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours. *Nature* **359**, 162–165 (1992).
180. González, I., Vicent, S., De Alava, E. & Lecanda, F. EWS/FLI-1 oncoprotein subtypes impose different requirements for transformation and metastatic activity in a murine model. *J Mol Med* **85**, 1015–1029 (2007).
181. Ng, T. L. *et al.* Ewing Sarcoma with Novel Translocation t(2;16) Producing an In-Frame Fusion of FUS and FEV. *J Mol Diagn* **9**, 459 (2007).
182. Shing, D. *et al.* FUS/ERG gene fusions in Ewing's tumors. *undefined* (2003).
183. Sorensen, P. H. B. *et al.* A second Ewing's sarcoma translocation, t(21;22), fuses the EWS gene to another ETS–family transcription factor, ERG. *Nature Genetics 1994 6:2* **6**, 146–151 (1994).
184. Zucman, J. *et al.* Combinatorial generation of variable fusion proteins in the Ewing family of tumours. *EMBO J* **12**, 4481 (1993).
185. Boone, M. A. *et al.* Identification of a novel FUS/ETV4 fusion and comparative analysis with other Ewing sarcoma fusion proteins. *Mol Cancer Res* **19**, 1795 (2021).
186. Picard, C. *et al.* Identification of a novel translocation producing an in-frame fusion of TAF15 and ETV4 in a case of extraosseous Ewing sarcoma revealed in the prenatal period. *Virchows Archiv 2022 481:4* **481**, 665–669 (2022).
187. Cantile, M. *et al.* Molecular detection and targeting of EWSR1 fusion transcripts in soft tissue tumors. *Medical Oncology* vol. 30 Preprint at https://doi.org/10.1007/s12032-012-0412-8 (2013).
188. Sbaraglia, M., Righi, A., Gambarotti, M. & Dei Tos, A. P. Ewing sarcoma and Ewing-like tumors. *Virchows Archiv 2019 476:1* **476**, 109–119 (2019).
189. Renzi, S., Anderson, N. D., Light, N. & Gupta, A. Ewing-like sarcoma: An emerging family of round cell sarcomas. *J Cell Physiol* **234**, 7999–8007 (2019).
190. Zucman-Rossi, J., Legoix, P., Victor, J. M., Lopez, B. & Thomas, G. Chromosome translocation based on illegitimate recombination in human tumors. *Proc Natl Acad Sci U S A* **95**, 11786–11791 (1998).
191. Zucman, J. *et al.* Cloning and characterization of the Ewing's sarcoma and peripheral neuroepithelioma t(11;22) translocation breakpoints. *Genes Chromosomes Cancer* **5**, 271–277 (1992).
192. Sankar, S. & Lessnick, S. L. Promiscuous partnerships in Ewing's sarcoma. *Cancer Genetics* vol. 204 351–365 Preprint at https://doi.org/10.1016/j.cancergen.2011.07.008 (2011).

193. Grohar, P. J. *et al.* Functional Genomic Screening Reveals Splicing of the EWS-FLI1 Fusion Transcript as a Vulnerability in Ewing Sarcoma. *Cell Rep* **14**, 598–610 (2016).
194. Huijbers, E. J. M. *et al.* Targeting tumor vascular CD99 inhibits tumor growth. *Front Immunol* **10**, 651 (2019).
195. Louati, S., Senhaji, N., Chbani, L. & Bennis, S. EWSR1 rearrangement and CD99 expression as diagnostic biomarkers for Ewing/PNET sarcomas in a moroccan population. *Dis Markers* **2018**, (2018).
196. Orth, M. F. *et al.* High Specificity of BCL11B and GLG1 for EWSR1-FLI1 and EWSR1-ERG Positive Ewing Sarcoma. *Cancers (Basel)* **12**, 644 (2020).
197. dos Santos, D. C., Evangelista, L. da C. & Kerche-Silva, L. E. Genetic alterations and diagnosis in Ewing sarcoma: A review. *Integr Cancer Sci Ther* **4**, (2017).
198. Sánchez-Molina, S. *et al.* RING1B recruits EWSR1-FLI1 and cooperates in the remodeling of chromatin necessary for Ewing sarcoma tumorigenesis. *Sci Adv* **6**, (2020).
199. García-Domínguez, D. J. *et al.* An inducible ectopic expression system of EWSR1-FLI1 as a tool for understanding Ewing sarcoma oncogenesis. *PLoS One* **15**, (2020).
200. May, W. A. *et al.* Ewing sarcoma 11;22 translocation produces a chimeric transcription factor that requires the DNA-binding domain encoded by FLI1 for transformation. *Proc Natl Acad Sci U S A* **90**, 5752–5756 (1993).
201. Riggi, N. *et al.* EWS-FLI-1 expression triggers a ewing's sarcoma initiation program in primary human mesenchymal stem cells. *Cancer Res* **68**, 2176–2185 (2008).
202. Cidre-Aranaz, F. & Alonso, J. EWS/FLI1 target genes and therapeutic opportunities in Ewing sarcoma. *Frontiers in Oncology* vol. 5 162 Preprint at https://doi.org/10.3389/fonc.2015.00162 (2015).
203. Kauer, M. *et al.* A molecular function map of Ewing's sarcoma. *PLoS One* **4**, (2009).
204. Boulay, G. *et al.* Epigenome editing of microsatellite repeats defines tumor-specific enhancer functions and dependencies. *Genes Dev* **32**, 1008–1019 (2018).
205. Aynaud, M. M. *et al.* Transcriptional Programs Define Intratumoral Heterogeneity of Ewing Sarcoma at Single-Cell Resolution. *Cell Rep* **30**, 1767-1779.e6 (2020).
206. Montoya, C. *et al.* Epigenetic control of the EWS-FLI1 promoter in Ewing's sarcoma. *Oncol Rep* **43**, 1199–1207 (2020).
207. Lawlor, E. R. & Sorensen, P. H. Twenty years on: What do we really know about ewing sarcoma and what is the path forward? *Crit Rev Oncog* **20**, 155–171 (2015).
208. Riggi, N. *et al.* EWS-FLI1Utilizes Divergent Chromatin Remodeling Mechanisms to Directly Activate or Repress Enhancer Elements in Ewing Sarcoma. *Cancer Cell* **26**, 668–681 (2014).
209. Nacev, B. A. *et al.* The epigenomics of sarcoma. *Nat Rev Cancer* **20**, 608 (2020).
210. May, W. A. *et al.* The Ewing's sarcoma EWS/FLI-1 fusion gene encodes a more potent transcriptional activator and is a more powerful transforming gene than FLI-1. *Mol Cell Biol* **13**, 7393–7398 (1993).

211. Douglas, D. *et al.* BMI-1 promotes Ewing sarcoma tumorigenicity independent of CDKN2A repression. *Cancer Res* **68**, 6507–6515 (2008).
212. Shimizu, R. *et al.* EWS-FLI1 regulates a transcriptional program in cooperation with Foxq1 in mouse Ewing sarcoma. *Cancer Sci* **109**, 2907–2918 (2018).
213. Scotlandi, K. *et al.* Effectiveness of insulin-like growth factor I receptor antisense strategy against Ewing's sarcoma cells. *Cancer Gene Ther* **9**, 296–307 (2002).
214. Rocchi, A. *et al.* CD99 inhibits neural differentiation of human Ewing sarcoma cells and thereby contributes to oncogenesis. *Journal of Clinical Investigation* **120**, 668–680 (2010).
215. Wei, G. *et al.* Prognostic impact of INK4A deletion in Ewing sarcoma. *Cancer* **89**, 793–9 (2000).
216. de Alava, E. *et al.* Prognostic impact of P53 status in Ewing sarcoma. *Cancer* **89**, 783–92 (2000).
217. Park, H., Galbraith, R., Turner, T., Mehojah, J. & Azuma, M. Loss of Ewing sarcoma EWS allele promotes tumorigenesis by inducing chromosomal instability in zebrafish. *Scientific Reports 2016 6:1* **6**, 1–8 (2016).
218. Ramakrishnan, R. *et al.* Role of protein-protein interactions in the antiapoptotic function of EWS-Fli-1. *Oncogene* **23**, 7087–7094 (2004).
219. Toretsky, J. A. *et al.* Oncoprotein EWS-FLI1 activity is enhanced by RNA helicase A. *Cancer Res* **66**, 5574–5581 (2006).
220. Knoop, L. L. & Baker, S. J. EWS/FLI Alters 5′-Splice Site Selection. *Journal of Biological Chemistry* **276**, 22317–22322 (2001).
221. Selvanathan, S. P. *et al.* Oncogenic fusion protein EWS-FLI1 is a network hub that regulates alternative splicing. *Proc Natl Acad Sci U S A* **112**, E1307–E1316 (2015).
222. Sanchez, G. *et al.* Alteration of cyclin D1 transcript elongation by a mutated transcription factor up-regulates the oncogenic D1b splice isoform in cancer. *Proc Natl Acad Sci U S A* **105**, 6004–6009 (2008).
223. Selvanathan, S. P. *et al.* EWS-FLI1 modulated alternative splicing of ARID1A reveals novel oncogenic function through the BAF complex. *Nucleic Acids Res* **47**, 9619–9636 (2019).
224. Li, Z., Yu, X., Shen, J., Wu, W. K. K. & Chan, M. T. V. MicroRNA expression and its clinical implications in Ewing's sarcoma. *Cell Prolif* **48**, 1–6 (2015).
225. Howarth, M. M. *et al.* Long noncoding RNA EWSAT1-mediated gene repression facilitates Ewing sarcoma oncogenesis. *Journal of Clinical Investigation* **124**, 5275–5290 (2014).
226. Yang, H. *et al.* Long non-coding RNA EWSAT1 contributes to the proliferation and invasion of glioma by sponging miR-152-3p. *Oncol Lett* **20**, 1846–1854 (2020).
227. Palombo, R. *et al.* The promoter-associated noncoding RNA PNCCCND1_B assembles a protein–RNA complex to regulate cyclin D1 transcription in Ewing sarcoma. *Cancer Res* **79**, 3570–3582 (2019).

228. Mercatelli, N., Fortini, D., Palombo, R. & Paronetto, M. P. Small molecule inhibition of Ewing sarcoma cell growth via targeting the long non coding RNA HULC. *Cancer Lett* **469**, 111–123 (2020).
229. Gorthi, A. & Bishop, A. J. R. Ewing sarcoma fusion oncogene: At the crossroads of transcription and DNA damage response. *Mol Cell Oncol* **5**, 1465014 (2018).
230. Halbkat, H. & Ghosal, G. Investigating EWS-FLI1 Induced Replication Stress in Ewing Sarcoma Pathogenesis. *The FASEB Journal* **34**, 1–1 (2020).
231. Gorthi, A. *et al.* EWS-FLI1 increases transcription to cause R-Loops and block BRCA1 repair in Ewing sarcoma. *Nature* **555**, 387–391 (2018).
232. McManus, J., Cheng, Z. & Vogel, C. Next-generation analysis of gene expression regulation – comparing the roles of synthesis and degradation. *Mol Biosyst* **11**, 2680 (2015).
233. van Nostrand, E. L. *et al.* A large-scale binding and functional map of human RNA-binding proteins. *Nature 2020 583:7818* **583**, 711–719 (2020).
234. Plotnikova, O., Baranova, A. & Skoblov, M. Comprehensive Analysis of Human microRNA–mRNA Interactome. *Front Genet* **10**, 933 (2019).
235. Dahan, O., Gingold, H. & Pilpel, Y. Regulatory mechanisms and networks couple the different phases of gene expression. *Trends in Genetics* **27**, 316–322 (2011).
236. Buccitelli, C. & Selbach, M. mRNAs, proteins and the emerging principles of gene expression control. *Nat Rev Genet* **21**, 630–644 (2020).
237. Sonenberg, N. & Hinnebusch, A. G. Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets. *Cell* **136**, 731 (2009).
238. Mailliot, J. & Martin, F. Viral internal ribosomal entry sites: four classes for one goal. *Wiley Interdiscip Rev RNA* **9**, e1458 (2018).
239. Paek, K. Y. *et al.* Translation initiation mediated by RNA looping. *Proc Natl Acad Sci U S A* **112**, 1041–1046 (2015).
240. Adivarahan, S. *et al.* Spatial Organization of Single mRNPs at Different Stages of the Gene Expression Pathway. *Mol Cell* **72**, 727-738.e5 (2018).
241. Kozak, M. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**, 283–292 (1986).
242. Jackson, R. J., Hellen, C. U. T. & Pestova, T. v. THE MECHANISM OF EUKARYOTIC TRANSLATION INITIATION AND PRINCIPLES OF ITS REGULATION. *Nat Rev Mol Cell Biol* **11**, 113 (2010).
243. Valle, M. *et al.* Incorporation of aminoacyl-tRNA into the ribosome as seen by cryo-electron microscopy. *Nat Struct Biol* **10**, 899–906 (2003).
244. Lareau, L. F., Hite, D. H., Hogan, G. J. & Brown, P. O. Distinct stages of the translation elongation cycle revealed by sequencing ribosome-protected mRNA fragments. *Elife* **2014**, (2014).
245. Dever, T. E. & Green, R. The Elongation, Termination, and Recycling Phases of Translation in Eukaryotes. *Cold Spring Harb Perspect Biol* **4**, 1–16 (2012).

246. Zhang, W., Dunkle, J. A. & Cate, J. H. D. Structures of the ribosome in intermediate states of ratcheting. *Science* **325**, 1014 (2009).
247. Bertram, G., Bell, H. A., Ritchie, D. W., Fullerton, G. & Stansfield, I. Terminating eukaryote translation: Domain 1 of release factor eRF1 functions in stop codon recognition. *RNA* **6**, 1236–1247 (2000).
248. Frolova, L. Y. *et al.* Mutations in the highly conserved GGQ motif of class 1 polypeptide release factors abolish ability of human eRF1 to trigger peptidyl-tRNA hydrolysis. *RNA* **5**, 1014 (1999).
249. Mantsyzov, A. B. *et al.* NMR solution structure and function of the C-terminal domain of eukaryotic class 1 polypeptide chain release factor. *FEBS J* **277**, 2611 (2010).
250. Hellen, C. U. T. Translation Termination and Ribosome Recycling in Eukaryotes. *Cold Spring Harb Perspect Biol* **10**, a032656 (2018).
251. Cheng, Z. *et al.* Structural insights into eRF3 and stop codon recognition by eRF1. *Genes Dev* **23**, 1106 (2009).
252. Chen, H. H. & Tarn, W. Y. uORF-mediated translational control: recently elucidated mechanisms and implications in cancer. *RNA Biol* **16**, 1327 (2019).
253. Calvo, S. E., Pagliarini, D. J. & Mootha, V. K. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc Natl Acad Sci U S A* **106**, 7507–7512 (2009).
254. Barbosa, C., Peixeiro, I. & Romão, L. Gene Expression Regulation by Upstream Open Reading Frames and Human Disease. *PLoS Genet* **9**, e1003529 (2013).
255. Lu, P. D., Harding, H. P. & Ron, D. Translation reinitiation at alternative open reading frames regulates gene expression in an integrated stress response. *J Cell Biol* **167**, 27–33 (2004).
256. Szamecz, B. *et al.* eIF3a cooperates with sequences 5′ of uORF1 to promote resumption of scanning by post-termination ribosomes for reinitiation on GCN4 mRNA. *Genes Dev* **22**, 2414 (2008).
257. Young, S. K. & Wek, R. C. Upstream Open Reading Frames Differentially Regulate Gene-specific Translation in the Integrated Stress Response. *J Biol Chem* **291**, 16927 (2016).
258. Pakos-Zebrucka, K. *et al.* The integrated stress response. *EMBO Rep* **17**, 1374 (2016).
259. Amorim, I. S., Lach, G. & Gkogkas, C. G. The Role of the Eukaryotic Translation Initiation Factor 4E (eIF4E) in Neuropsychiatric Disorders. *Front Genet* **9**, 561 (2018).
260. Leppek, K., Das, R. & Barna, M. Functional 5′ UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nature Reviews Molecular Cell Biology 2017 19:3* **19**, 158–174 (2017).
261. J., M. Distinctive features of the 5'-terminal sequences of the human mitochondrial mRNAs. *Nature* **290**, 465–470 (1981).
262. Elfakess, R. *et al.* Unique translation initiation of mRNAs-containing TISU element. *Nucleic Acids Res* **39**, 7598 (2011).

263. Muckenthaler, M. U., Rivella, S., Hentze, M. W. & Galy, B. A Red Carpet for Iron Metabolism. *Cell* **168**, 344 (2017).
264. Gray, N. K. & Hentze, M. W. Iron regulatory protein prevents binding of the 43S translation pre-initiation complex to ferritin and eALAS mRNAs. *EMBO J* **13**, 3882 (1994).
265. Muckenthaler, M., Gray, N. K. & Hentze, M. W. IRP-1 binding to ferritin mRNA prevents the recruitment of the small ribosomal subunit by the cap-binding complex eIF4F. *Mol Cell* **2**, 383–388 (1998).
266. Özeş, A. R., Feoktistova, K., Avanzino, B. C. & Fraser, C. S. Duplex unwinding and ATPase activities of the DEAD-box helicase eIF4A are coupled by eIF4G and eIF4B. *J Mol Biol* **412**, 674 (2011).
267. Rubio, C. A. *et al.* Transcriptome-wide characterization of the eIF4A signature highlights plasticity in translation regulation. *Genome Biol* **15**, 1–19 (2014).
268. Sen, N. D., Zhou, F., Harris, M. S., Ingolia, N. T. & Hinnebusch, A. G. eIF4B stimulates translation of long mRNAs with structured 5′ UTRs and low closed-loop potential but weak dependence on eIF4G. *Proc Natl Acad Sci U S A* **113**, 10464–10472 (2016).
269. Kumari, S., Bugaut, A. & Balasubramanian, S. Position and Stability Are Determining Factors for Translation Repression by an RNA G-Quadruplex-Forming Sequence within the 5′ UTR of the NRAS Proto-oncogene. *Biochemistry* **47**, 12664 (2008).
270. Castets, M. *et al.* FMRP interferes with the Rac1 pathway and controls actin cytoskeleton dynamics in murine fibroblasts. *Hum Mol Genet* **14**, 835–844 (2005).
271. Benhalevy, D. *et al.* The human CCHC-type Zinc Finger Nucleic Acid Binding Protein binds G-rich elements in target mRNA coding sequences and promotes translation. *Cell Rep* **18**, 2979 (2017).
272. Brierley, I., Pennell, S. & Gilbert, R. J. C. Viral RNA pseudoknots: versatile motifs in gene expression and replication. *Nature Reviews Microbiology 2007 5:8* **5**, 598–610 (2007).
273. Ben-Asouli, Y., Banai, Y., Pel-Or, Y., Shir, A. & Kaempfer, R. Human interferon-γ mRNA autoregulates its translation through a pseudoknot that activates the interferon-inducible protein kinase PKR. *Cell* **108**, 221–232 (2002).
274. Carrieri, C. *et al.* Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature 2012 491:7424* **491**, 454–457 (2012).
275. Leppek, K., Das, R. & Barna, M. Functional 5′ UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nat Rev Mol Cell Biol* **19**, 158 (2018).
276. Harvey, R. F. *et al.* Trans-acting translational regulatory RNA binding proteins. *Wiley Interdiscip Rev RNA* **9**, e1465 (2018).
277. Auweter, S. D., Oberstrass, F. C. & Allain, F. H. T. Sequence-specific binding of single-stranded RNA: is there a code for recognition? *Nucleic Acids Res* **34**, 4943 (2006).
278. Nag, S., Goswami, B., das Mandal, S. & Ray, P. S. Cooperation and competition by RNA-binding proteins in cancer. *Semin Cancer Biol* (2022) doi:10.1016/J.SEMCANCER.2022.02.023.

279. Lahr, R. M. *et al.* La-related protein 1 (LARP1) binds the mRNA cap, blocking eIF4F assembly on TOP mRNAs. *Elife* **6**, (2017).
280. Fonseca, B. D. *et al.* La-related Protein 1 (LARP1) Represses Terminal Oligopyrimidine (TOP) mRNA Translation Downstream of mTOR Complex 1 (mTORC1). *J Biol Chem* **290**, 15996–16020 (2015).
281. Bernstein, P., Peltz, S. W. & Ross, J. The poly(A)-poly(A)-binding protein complex is a major determinant of mRNA stability in vitro. *Mol Cell Biol* **9**, 659–670 (1989).
282. Wei, C. C., Balasta, M. L., Ren, J. & Goss, D. J. Wheat Germ Poly(A) Binding Protein Enhances the Binding Affinity of Eukaryotic Initiation Factor 4F and (iso)4F for Cap Analogues†. *Biochemistry* **37**, 1910–1916 (1998).
283. Rajkowitsch, L., Vilela, C., Berthelot, K., Ramirez, C. V. & McCarthy, J. E. G. Reinitiation and Recycling are Distinct Processes Occurring Downstream of Translation Termination in Yeast. *J Mol Biol* **335**, 71–85 (2004).
284. Oliveto, S., Mancino, M., Manfrini, N. & Biffo, S. Role of microRNAs in translation regulation and cancer. *World J Biol Chem* **8**, 45 (2017).
285. Kiriakidou, M. *et al.* An mRNA m7G Cap Binding-like Motif within Human Ago2 Represses Translation. *Cell* **129**, 1141–1151 (2007).
286. Zekri, L., Kuzuoğlu-Öztürk, D. & Izaurralde, E. GW182 proteins cause PABP dissociation from silenced miRNA targets in the absence of deadenylation. *EMBO J* **32**, 1052 (2013).
287. Ricci, E. P. *et al.* miRNA repression of translation in vitro takes place during 43S ribosomal scanning. *Nucleic Acids Res* **41**, 586 (2013).
288. Lodish, H. F. Model for the regulation of mRNA translation applied to haemoglobin synthesis. *Nature 1974 251:5474* **251**, 385–388 (1974).
289. Mills, E. W. & Green, R. Ribosomopathies: There's strength in numbers. *Science (1979)* **358**, (2017).
290. Ferretti, M. B. & Karbstein, K. Does functional specialization of ribosomes really exist? *RNA* **25**, 521 (2019).
291. Farley-Barnes, K. I., Ogawa, L. M. & Baserga, S. J. Ribosomopathies: old concepts, new controversies. *Trends Genet* **35**, 754 (2019).
292. Avcilar-Kucukgoze, I. & Kashina, A. Hijacking tRNAs From Translation: Regulatory Functions of tRNAs in Mammalian Cell Physiology. *Front Mol Biosci* **7**, 388 (2020).
293. Neelagandan, N., Lamberti, I., Carvalho, H. J. F., Gobet, C. & Naef, F. What determines eukaryotic translation elongation: recent molecular and quantitative analyses of protein synthesis. *Open Biol* **10**, 200292 (2020).
294. Dittmar, K. A., Goodenbour, J. M. & Pan, T. Tissue-Specific Differences in Human Transfer RNA Expression. *PLoS Genet* **2**, e221 (2006).
295. Gingold, H. *et al.* A dual program for translation regulation in cellular proliferation and differentiation. *Cell* **158**, 1281–1292 (2014).
296. Guimaraes, J. C. *et al.* A rare codon-based translational program of cell proliferation. *Genome Biol* **21**, 1–20 (2020).

297. Krutyhołowa, R., Zakrzewski, K. & Glatt, S. Charging the code — tRNA modification complexes. *Curr Opin Struct Biol* **55**, 138–146 (2019).
298. Nedialkova, D. D. & Leidel, S. A. Optimization of Codon Translation Rates via tRNA Modifications Maintains Proteome Integrity. *Cell* **161**, 1606 (2015).
299. Tuorto, F. *et al.* The tRNA methyltransferase Dnmt2 is required for accurate polypeptide synthesis during haematopoiesis. *EMBO J* **34**, 2350 (2015).
300. Liu, Y., Yang, Q. & Zhao, F. Synonymous but Not Silent: The Codon Usage Code for Gene Expression and Protein Folding. *Annu Rev Biochem* **90**, 375–401 (2021).
301. Dana, A. & Tuller, T. The effect of tRNA levels on decoding times of mRNA codons. *Nucleic Acids Res* **42**, 9171 (2014).
302. Fang, H. *et al.* Scikit-ribo Enables Accurate Estimation and Robust Modeling of Translation Dynamics at Codon Resolution. *Cell Syst* **6**, 180-191.e4 (2018).
303. O'Connor, P. B. F., Andreev, D. E. & Baranov, P. v. Comparative survey of the relative impact of mRNA features on local ribosome profiling read density. *Nature Communications 2016 7:1* **7**, 1–12 (2016).
304. Gobet, C. *et al.* Robust landscapes of ribosome dwell times and aminoacyl-tRNAs in response to nutrient stress in liver. *Proc Natl Acad Sci U S A* **117**, 9630–9641 (2020).
305. Riba, A. *et al.* Protein synthesis rates and ribosome occupancies reveal determinants of translation elongation rates. *Proc Natl Acad Sci U S A* **116**, 15023–15032 (2019).
306. Dittmar, K. A., Sørensen, M. A., Elf, J., Ehrenberg, M. & Pan, T. Selective charging of tRNA isoacceptors induced by amino-acid starvation. *EMBO Rep* **6**, 151–157 (2005).
307. Saikia, M. *et al.* Codon optimality controls differential mRNA translation during amino acid starvation. *RNA* **22**, 1719–1727 (2016).
308. Hernandez-Alias, X., Benisty, H., Schaefer, M. H. & Serrano, L. Translational efficiency across healthy and tumor tissues is proliferation-related. *Mol Syst Biol* **16**, (2020).
309. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S. & Weissman, J. S. Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. *Science* **324**, 218 (2009).
310. Chandrasekaran, V. *et al.* Mechanism of ribosome stalling during translation of a poly(A) tail. *Nat Struct Mol Biol* **26**, 1132 (2019).
311. Darnell, A. M., Subramaniam, A. R. & O'Shea, E. K. Translational control through differential ribosome pausing during amino acid limitation in mammalian cells. *Mol Cell* **71**, 229 (2018).
312. Han, P. *et al.* Genome-wide Survey of Ribosome Collision. *Cell Rep* **31**, 107610 (2020).
313. Arpat, A. B. *et al.* Transcriptome-wide sites of collided ribosomes reveal principles of translational pausing. *Genome Res* **30**, 985–999 (2020).
314. Dever, T. E., Gutierrez, E. & Shin, B. S. The hypusine-containing translation factor eIF5A. *Crit Rev Biochem Mol Biol* **49**, 413 (2014).
315. Huter, P. *et al.* Structural Basis for Polyproline-Mediated Ribosome Stalling and Rescue by the Translation Elongation Factor EF-P. *Mol Cell* **68**, 515-527.e6 (2017).

316. Gutierrez, E. *et al.* eIF5A promotes translation of polyproline motifs. *Mol Cell* **51**, 35–45 (2013).
317. Manjunath, H. *et al.* Suppression of Ribosomal Pausing by eIF5A Is Necessary to Maintain the Fidelity of Start Codon Selection. *Cell Rep* **29**, 3134-3146.e6 (2019).
318. Brandman, O. & Hegde, R. S. Ribosome-associated protein quality control. *Nat Struct Mol Biol* **23**, 7 (2016).
319. van Hoof, A. & Wagner, E. J. A brief survey of mRNA surveillance. *Trends Biochem Sci* **36**, 585 (2011).
320. LaRiviere, F. J., Cole, S. E., Ferullo, D. J. & Moore, M. J. A late-acting quality control process for mature eukaryotic rRNAs. *Mol Cell* **24**, 619–626 (2006).
321. Yadavalli, S. S. & Ibba, M. Quality control in aminoacyl-tRNA synthesis its role in translational fidelity. *Adv Protein Chem Struct Biol* **86**, 1–43 (2012).
322. Hopefield, J. J. Kinetic Proofreading: A New Mechanism for Reducing Errors in Biosynthetic Processes Requiring High Specificity. *Proceedings of the National Academy of Sciences* **71**, 4135–4139 (1974).
323. Wolff, S., Weissman, J. S. & Dillin, A. Differential Scales of Protein Quality Control. *Cell* **157**, 52–64 (2014).
324. Joazeiro, C. A. P. Mechanisms and functions of ribosome-associated protein quality control. *Nat Rev Mol Cell Biol* **20**, 368 (2019).
325. Nickless, A., Bailis, J. M. & You, Z. Control of gene expression through the nonsense-mediated RNA decay pathway. *Cell & Bioscience 2017 7:1* **7**, 1–12 (2017).
326. Serdar, L. D., Whiteside, D. J. L. & Baker, K. E. ATP hydrolysis by UPF1 is required for efficient translation termination at premature stop codons. *Nat Commun* **7**, (2016).
327. Neu-Yilik, G. *et al.* Dual function of UPF3B in early and late translation termination. *EMBO J* **36**, 2968 (2017).
328. Pelechano, V. & Alepuz, P. eIF5A facilitates translation termination globally and promotes the elongation of many non polyproline-specific tripeptide sequences. *Nucleic Acids Res* **45**, 7326 (2017).
329. Tang, X. *et al.* Structural basis of suppression of host translation termination by Moloney Murine Leukemia Virus. *Nat Commun* **7**, (2016).
330. Li, C. & Zhang, J. Stop-codon read-through arises largely from molecular errors and is generally nonadaptive. *PLoS Genet* **15**, (2019).
331. Namy, O. & Rousset, J.-P. Specification of Standard Amino Acids by Stop Codons. 79–100 (2010) doi:10.1007/978-0-387-89382-2_4.
332. Houck-Loomis, B. *et al.* An equilibrium-dependent retroviral mRNA switch regulates translational recoding. *Nature* **480**, 561–564 (2011).
333. Jungreis, I. *et al.* Evolutionary Dynamics of Abundant Stop Codon Readthrough. *Mol Biol Evol* **33**, 3108 (2016).
334. Steneberg, P. & Samakovlis, C. A novel stop codon readthrough mechanism produces functional Headcase protein in Drosophila trachea. *EMBO Rep* **2**, 593 (2001).

335. Eswarappa, S. M. *et al.* Programmed Translational Readthrough Generates Anti-Angiogenic VEGF-Ax. *Cell* **157**, 1605 (2014).
336. Feng, T. *et al.* Optimal Translational Termination Requires C4 Lysyl Hydroxylation of eRF1. *Mol Cell* **53**, 645 (2014).
337. James, C. C. & Smyth, J. W. Alternative mechanisms of translation initiation: an emerging dynamic regulator of the proteome in health and disease. *Life Sci* **212**, 138 (2018).
338. Martin, F. *et al.* Cap-Assisted Internal Initiation of Translation of Histone H4. *Mol Cell* **41**, 197–209 (2011).
339. Marzluff, W. F., Wagner, E. J. & Duronio, R. J. Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. *Nat Rev Genet* **9**, 843 (2008).
340. Martin, F. *et al.* Ribosomal 18S rRNA base pairs with mRNA during eukaryotic translation initiation. *Nat Commun* **7**, (2016).
341. Truniger, V., Miras, M. & Aranda, M. A. Structural and functional diversity of plant virus 3′-cap-independent translation enhancers (3′-CITEs). *Front Plant Sci* **8**, 2047 (2017).
342. Treder, K. *et al.* The 3' cap-independent translation element of Barley yellow dwarf virus binds eIF4F via the eIF4G subunit to initiate translation. *RNA* **14**, 134–147 (2008).
343. Wang, Z., Treder, K. & Miller, W. A. Structure of a viral cap-independent translation element that functions via high affinity binding to the eIF4E subunit of eIF4F. *J Biol Chem* **284**, 14189–14202 (2009).
344. Wang, S. *et al.* Dynamic regulation and functions of mRNA m6A modification. *Cancer Cell Int* **22**, 1–12 (2022).
345. Wang, X. *et al.* N6-methyladenosine Modulates Messenger RNA Translation Efficiency. *Cell* **161**, 1388 (2015).
346. Meyer, K. D. *et al.* 5′ UTR m6A Promotes Cap-Independent Translation. *Cell* **163**, 999 (2015).
347. Choi, J. *et al.* N6-methyladenosine in mRNA disrupts tRNA selection and translation elongation dynamics. *Nat Struct Mol Biol* **23**, 110 (2016).
348. Gomes-Duarte, A., Lacerda, R., Menezes, J. & Romão, L. eIF3: a factor for human health and disease. *RNA Biol* **15**, 26 (2018).
349. Valášek, L. S. *et al.* Embraced by eIF3: structural and functional insights into the roles of eIF3 across the translation cycle. *Nucleic Acids Res* **45**, 10948 (2017).
350. Shi, H. *et al.* YTHDF3 facilitates translation and decay of N6-methyladenosine-modified RNA. *Cell Research 2017 27:3* **27**, 315–328 (2017).
351. Choe, J. *et al.* mRNA circularization by METTL3-eIF3h enhances translation and promotes oncogenesis. *Nature* **561**, 556 (2018).
352. Sinha, T., Panigrahi, C., Das, D. & Chandra Panda, A. Circular RNA translation, a path to hidden proteome. *Wiley Interdiscip Rev RNA* **13**, e1685 (2022).
353. Legnini, I. *et al.* Circ-ZNF609 Is a Circular RNA that Can Be Translated and Functions in Myogenesis. *Mol Cell* **66**, 22-37.e9 (2017).

354. Yang, Y. *et al.* Extensive translation of circular RNAs driven by N6-methyladenosine. *Cell Research 2017 27:5* **27**, 626–641 (2017).
355. Johannes, G., Carter, M. S., Eisen, M. B., Brown, P. O. & Sarnow, P. Identification of eukaryotic mRNAs that are translated at reduced cap binding complex eIF4F concentrations using a cDNA microarray. *Proc Natl Acad Sci U S A* **96**, 13118–13123 (1999).
356. Komar, A. A. & Hatzoglou, M. Cellular IRES-mediated translation: The war of ITAFs in pathophysiological states. *Cell Cycle* **10**, 229 (2011).
357. Sweeney, T. R., Abaeva, I. S., Pestova, T. v. & Hellen, C. U. T. The mechanism of translation initiation on Type 1 picornavirus IRESs. *EMBO J* **33**, 76 (2014).
358. Toyoda, H., Franco, D., Fujita, K., Paul, A. v. & Wimmer, E. Replication of poliovirus requires binding of the poly(rC) binding protein to the cloverleaf as well as to the adjacent C-rich spacer sequence between the cloverleaf and the internal ribosomal entry site. *J Virol* **81**, 10017–10028 (2007).
359. Lozano, G. & Martínez-Salas, E. Structural insights into viral IRES-dependent translation mechanisms. *Curr Opin Virol* **12**, 113–120 (2015).
360. Lozano, G., Fernandez, N. & Martinez-Salas, E. Modeling Three-Dimensional Structural Motifs of Viral IRES. *J Mol Biol* **428**, 767–776 (2016).
361. Ló Pez De Quinto And Encarnació N Martínez-Salas, S. Involvement of the Aphthovirus RNA Region Located between the Two Functional AUGs in Start Codon Selection. (1999).
362. Belsham, G. J. Divergent picornavirus IRES elements. *Virus Res* **139**, 183–192 (2009).
363. Honda, M., Brown, E. A. & Lemon, S. M. Stability of a stem-loop involving the initiator AUG controls the efficiency of internal initiation of translation on hepatitis C virus RNA. *RNA* **2**, 955 (1996).
364. Jopling, C. L., Yi, M. K., Lancaster, A. M., Lemon, S. M. & Sarnow, P. Molecular biology: Modulation of hepatitis C virus RNA abundance by a liver-specific MicroRNA. *Science (1979)* **309**, 1577–1581 (2005).
365. Lukavsky, P. J., Kim, I., Otto, G. A. & Puglisi, J. D. Structure of HCV IRES domain II determined by NMR. *Nature Structural & Molecular Biology 2003 10:12* **10**, 1033–1038 (2003).
366. Cate, J. H. D. Human eIF3: from 'blobology' to biological insight. *Philosophical Transactions of the Royal Society B: Biological Sciences* **372**, (2017).
367. Pestova, T. v. & Hellen, C. U. T. Translation elongation after assembly of ribosomes on the Cricket paralysis virus internal ribosomal entry site without initiation factors or initiator tRNA. *Genes Dev* **17**, 181 (2003).
368. Wilson, J. E., Pestova, T. v., Hellen, C. U. T. & Sarnow, P. Initiation of protein synthesis from the A site of the ribosome. *Cell* **102**, 511–520 (2000).
369. Jan, E. & Sarnow, P. Factorless Ribosome Assembly on the Internal Ribosome Entry Site of Cricket Paralysis Virus. *J Mol Biol* **324**, 889–902 (2002).

370. Ford, D. Ribosomal heterogeneity – A new inroad for pharmacological innovation. *Biochem Pharmacol* **175**, 113874 (2020).
371. Kressler, D., Hurt, E. & Baßler, J. A Puzzle of Life: Crafting Ribosomal Subunits. *Trends Biochem Sci* **42**, 640–654 (2017).
372. Zhao, P. Y. *et al.* Eukaryotic ribosome quality control system: a potential therapeutic target for human diseases. *Int J Biol Sci* **18**, 2497–2514 (2022).
373. de La Cruz, J., Karbstein, K. & Woolford, J. L. Functions of Ribosomal Proteins in Assembly of Eukaryotic Ribosomes In Vivo. *Annu Rev Biochem* **84**, 93 (2015).
374. Yusupova, G. & Yusupov, M. Crystal structure of eukaryotic ribosome and its complexes with inhibitors. *Philosophical Transactions of the Royal Society B: Biological Sciences* **372**, (2017).
375. Londei, P. & Ferreira-Cerca, S. Ribosome Biogenesis in Archaea. *Front Microbiol* **12**, 1476 (2021).
376. Melnikov, S. *et al.* One core, two shells: bacterial and eukaryotic ribosomes. *Nat Struct Mol Biol* **19**, 560–567 (2012).
377. Timsit, Y., Sergeant-Perthuis, G. & Bennequin, D. Evolution of ribosomal protein network architectures. *Scientific Reports 2021 11:1* **11**, 1–13 (2021).
378. Weisser, M. & Ban, N. Extensions, Extra Factors, and Extreme Complexity: Ribosomal Structures Provide Insights into Eukaryotic Translation. *Cold Spring Harb Perspect Biol* **11**, a032367 (2019).
379. Kressler, D., Hurt, E. & Baßler, J. Driving ribosome assembly. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1803**, 673–683 (2010).
380. Guo, H. Specialized ribosomes and the control of translation. *Biochem Soc Trans* **46**, 855–869 (2018).
381. Jagannathan, S., Reid, D. W., Cox, A. H. & Nicchitta, C. v. De novo translation initiation on membrane-bound ribosomes as a mechanism for localization of cytosolic protein mRNAs to the endoplasmic reticulum. *RNA* **20**, 1489 (2014).
382. Scaltsoyiannes, V., Corre, N., Waltz, F. & Giegé, P. Types and Functions of Mitoribosome-Specific Ribosomal Proteins across Eukaryotes. *Int J Mol Sci* **23**, (2022).
383. de Silva, D., Tu, Y. T., Amunts, A., Fontanesi, F. & Barrientos, A. Mitochondrial ribosome assembly in health and disease. *Cell Cycle* **14**, 2226 (2015).
384. Greber, B. J. & Ban, N. Structure and Function of the Mitochondrial Ribosome. *http://dx.doi.org/10.1146/annurev-biochem-060815-014343* **85**, 103–132 (2016).
385. Liu, T. Y. *et al.* Time-Resolved Proteomics Extends Ribosome Profiling-Based Measurements of Protein Synthesis Dynamics. *Cell Syst* **4**, 636 (2017).
386. Fujii, K., Shi, Z., Zhulyn, O., Denans, N. & Barna, M. Pervasive translational regulation of the cell signalling circuitry underlies mammalian development. *Nature Communications 2017 8:1* **8**, 1–13 (2017).
387. Norris, K., Hopes, T. & Aspden, J. L. Ribosome heterogeneity and specialization in development. *Wiley Interdiscip Rev RNA* **12**, (2021).

388. Mauro, V. P. & Edelman, G. M. The ribosome filter hypothesis. *Proc Natl Acad Sci U S A* **99**, 12031 (2002).
389. Komili, S., Farny, N. G., Roth, F. P. & Silver, P. A. Functional specificity among ribosomal proteins regulates gene expression. *Cell* **131**, 557–571 (2007).
390. Li, J. *et al.* Regulation and trafficking of three distinct 18 S ribosomal RNAs during development of the malaria parasite. *J Mol Biol* **269**, 203–213 (1997).
391. Gay, D. M., Lund, A. H. & Jansson, M. D. Translational control through ribosome heterogeneity and functional specialization. *Trends Biochem Sci* **47**, 66–81 (2022).
392. Anger, A. M. *et al.* Structures of the human and Drosophila 80S ribosome. *Nature 2013 497:7447* **497**, 80–85 (2013).
393. Parks, M. M. *et al.* Variant ribosomal RNA alleles are conserved and exhibit tissue-specific expression. *Sci Adv* **4**, (2018).
394. Locati, M. D. *et al.* Expression of distinct maternal and somatic 5.8S, 18S, and 28S rRNA types during zebrafish development. *RNA* **23**, 1188–1199 (2017).
395. Parker, M. S., Balasubramaniam, A., Sallee, F. R. & Parker, S. L. The expansion segments of 28S Ribosomal RNA extensively match human messenger RNAs. *Front Genet* **9**, 66 (2018).
396. Leppek, K. *et al.* Gene- and species-specific Hox mRNA translation by ribosome expansion segments. *Mol Cell* **80**, 980 (2020).
397. Taoka, M. *et al.* Landscape of the complete RNA chemical modifications in the human 80S ribosome. *Nucleic Acids Res* **46**, 9289 (2018).
398. Sloan, K. E. *et al.* Tuning the ribosome: The influence of rRNA modification on eukaryotic ribosome biogenesis and function. *RNA Biol* **14**, 1138 (2017).
399. Decatur, W. A. & Fournier, M. J. rRNA modifications and ribosome function. *Trends Biochem Sci* **27**, 344–351 (2002).
400. Jansson, M. D. *et al.* Regulation of translation by site-specific ribosomal RNA methylation. *Nature Structural & Molecular Biology 2021 28:11* **28**, 889–899 (2021).
401. Jack, K. *et al.* rRNA Pseudouridylation Defects Affect Ribosomal Ligand Binding and Translational Fidelity from Yeast to Human Cells. *Mol Cell* **44**, 660 (2011).
402. Elhamamsy, A. R., Metge, B. J., Alsheikh, H. A., Shevde, L. A. & Samant, R. S. Ribosome Biogenesis: A Central Player in Cancer Metastasis and Therapeutic Resistance. *Cancer Res* **82**, 2344–2353 (2022).
403. Bretones, G. *et al.* Altered patterns of global protein synthesis and translational fidelity in RPS15-mutated chronic lymphocytic leukemia. *Blood* **132**, 2375 (2018).
404. Babaian, A. *et al.* Loss of m1acp3Ψ Ribosomal RNA Modification Is a Major Feature of Cancer. *Cell Rep* **31**, (2020).
405. Slavov, N., Semrau, S., Airoldi, E., Budnik, B. & van Oudenaarden, A. Differential Stoichiometry among Core Ribosomal Proteins. *Cell Rep* **13**, 865 (2015).
406. Xue, S. *et al.* RNA regulons in Hox 5′UTRs confer ribosome specificity to gene regulation. *Nature* **517**, 33 (2015).

407. Warner, J. R. & McIntosh, K. B. How Common are Extra-ribosomal Functions of Ribosomal Proteins? *Mol Cell* **34**, 3 (2009).
408. Landry, D. M., Hertz, M. I. & Thompson, S. R. RPS25 is essential for translation initiation by the Dicistroviridae and hepatitis C viral IRESs. *Genes Dev* **23**, 2753 (2009).
409. Hertz, M. I., Landry, D. M., Willis, A. E., Luo, G. & Thompson, S. R. Ribosomal Protein S25 Dependency Reveals a Common Mechanism for Diverse Internal Ribosome Entry Sites and Ribosome Shunting. *Mol Cell Biol* **33**, 1016 (2013).
410. Segev, N. & Gerst, J. E. Specialized ribosomes and specific ribosomal protein paralogs control translation of mitochondrial proteins. *J Cell Biol* **217**, 117 (2018).
411. O'Leary, M. N. *et al.* The Ribosomal Protein Rpl22 Controls Ribosome Composition by Directly Repressing Expression of Its Own Paralog, Rpl22l1. *PLoS Genet* **9**, 1003708 (2013).
412. Anderson, S. J. *et al.* Ablation of Ribosomal Protein L22 Selectively Impairs αβ T Cell Development by Activation of a p53-Dependent Checkpoint. *Immunity* **26**, 759–772 (2007).
413. Zhang, Y. *et al.* Ribosomal proteins Rpl22 and Rpl22l1 control morphogenesis by regulating pre-mRNA splicing. *Cell Rep* **18**, 545 (2017).
414. Genuth, N. R. & Barna, M. Heterogeneity and specialized functions of translation machinery: from genes to organisms. *Nat Rev Genet* **19**, 431 (2018).
415. Naganuma, T., Shiogama, K. & Uchiumi, T. The N-terminal regions of eukaryotic acidic phosphoproteins P1 and P2 are crucial for heterodimerization and assembly into the ribosomal GTPase-associated center. *Genes Cells* **12**, 501–510 (2007).
416. Artero-Castro, A. *et al.* Disruption of the ribosomal P complex leads to stress-induced autophagy. *Autophagy* **11**, 1499 (2015).
417. Meyuhas, O. Ribosomal Protein S6 Phosphorylation: Four Decades of Research. *Int Rev Cell Mol Biol* **320**, 41–73 (2015).
418. Martin, I. *et al.* Ribosomal protein s15 phosphorylation mediates LRRK2 neurodegeneration in Parkinson's disease. *Cell* **157**, 472 (2014).
419. Spence, J. *et al.* Cell cycle-regulated modification of the ribosome by a variant multiubiquitin chain. *Cell* **102**, 67–76 (2000).
420. Chen, E., Sharma, M. R., Shi, X., Agrawal, R. K. & Joseph, S. Fragile X Mental Retardation Protein Regulates Translation by Binding Directly to the Ribosome. *Mol Cell* **54**, 407 (2014).
421. Oliver, E. R., Saunders, T. L., Tarlé, S. A. & Glaser, T. Ribosomal protein L24 defect in belly spot and tail (Bst), a mouse Minute. *Development* **131**, 3907–3920 (2004).
422. Orgebin, E. *et al.* Ribosomopathies: New Therapeutic Perspectives. *Cells* **9**, (2020).
423. Horos, R. *et al.* Ribosomal deficiencies in Diamond-Blackfan anemia impair translation of transcripts essential for differentiation of murine and human erythroblasts. *Blood* **119**, 262–272 (2012).

424. Draptchinskaia, N. *et al.* The gene encoding ribosomal protein S19 is mutated in Diamond-Blackfan anaemia. *Nature Genetics 1999 21:2* **21**, 169–175 (1999).
425. Fadus, M. C., Rush, E. T. & Lettieri, C. K. Syndrome of progressive bone marrow failure and pancreatic insufficiency remains cryptic despite whole exome sequencing: variant of Shwachman-Diamond syndrome or new condition? *Clin Case Rep* **5**, 748 (2017).
426. Trainor, P. A., Dixon, J. & Dixon, M. J. Treacher Collins syndrome: etiology, pathogenesis and prevention. *European Journal of Human Genetics* **17**, 275 (2009).
427. Ebert, B. L. *et al.* Identification of RPS14 as a 5q- syndrome gene by RNA interference screen. *Nature* **451**, 335 (2008).
428. Chagnon, P. *et al.* A Missense Mutation (R565W) in Cirhin (FLJ14728) in North American Indian Childhood Cirrhosis. *Am J Hum Genet* **71**, 1443 (2002).
429. Bolze, A. *et al.* Ribosomal protein SA haploinsufficiency in humans with isolated congenital asplenia. *Science* **340**, 976 (2013).
430. King, H. A. & Gerber, A. P. Translatome profiling: methods for genome-scale analysis of mRNA translation. *Brief Funct Genomics* **15**, 22–31 (2016).
431. Zhao, J., Qin, B., Nikolay, R., Spahn, C. M. T. & Zhang, G. Translatomics: The Global View of Translation. *Int J Mol Sci* **20**, (2019).
432. Kuhn, K. M., DeRisi, J. L., Brown, P. O. & Sarnow, P. Global and Specific Translational Regulation in the Genomic Response of Saccharomyces cerevisiae to a Rapid Transfer from a Fermentable to a Nonfermentable Carbon Source. *Mol Cell Biol* **21**, 916 (2001).
433. Bushell, M. *et al.* Polypyrimidine tract binding protein regulates IRES-mediated gene expression during apoptosis. *Mol Cell* **23**, 401–412 (2006).
434. Kawai, T., Fan, J., Mazan-Mamczarz, K. & Gorospe, M. Global mRNA Stabilization Preferentially Linked to Translational Repression during the Endoplasmic Reticulum Stress Response. *Mol Cell Biol* **24**, 6773 (2004).
435. Thomas, J. D. & Johannes, G. J. Identification of mRNAs that continue to associate with polysomes during hypoxia. *RNA* **13**, 1116 (2007).
436. Johannes, G., Carter, M. S., Eisen, M. B., Brown, P. O. & Sarnow, P. Identification of eukaryotic mRNAs that are translated at reduced cap binding complex eIF4F concentrations using a cDNA microarray. *Proc Natl Acad Sci U S A* **96**, 13118 (1999).
437. Larsson, O. *et al.* Distinct perturbation of the translatome by the antidiabetic drug metformin. *Proc Natl Acad Sci U S A* **109**, 8977–8982 (2012).
438. Wang, T. *et al.* Translating mRNAs strongly correlate to proteins in a multivariate manner and their translation ratios are phenotype specific. *Nucleic Acids Res* **41**, 4743 (2013).
439. Heiman, M., Kulicke, R., Fenster, R. J., Greengard, P. & Heintz, N. Cell type-specific mRNA purification by translating ribosome affinity purification (TRAP). *Nat Protoc* **9**, 1282–1291 (2014).
440. Salussolia, C. L., Winden, K. D. & Sahin, M. Translating Ribosome Affinity Purification (TRAP) of Cell Type-specific mRNA from Mouse Brain Lysates. *Bio Protoc* **12**, (2022).

441. Brar, G. A. & Weissman, J. S. Ribosome profiling reveals the what, when, where and how of protein synthesis. (2015) doi:10.1038/nrm4069.
442. Chothani, S. *et al.* deltaTE: Detection of Translationally Regulated Genes by Integrative Analysis of Ribo-seq and RNA-seq Data. *Curr Protoc Mol Biol* **129**, e108 (2019).
443. Archer, S. K., Shirokikh, N. E., Beilharz, T. H. & Preiss, T. Dynamics of ribosome scanning and recycling revealed by translation complex profiling. *Nature* **535**, 570–574 (2016).
444. Giess, A. *et al.* Profiling of Small Ribosomal Subunits Reveals Modes and Regulation of Translation Initiation. *Cell Rep* **31**, 107534 (2020).
445. Mazzarello, P., Calligaro, A., Vannini, V. & Muscatello, U. The sarcoplasmic reticulum: Its discovery and rediscovery. *Nat Rev Mol Cell Biol* **4**, 69–74 (2003).
446. Friedman, J. R. & Voeltz, G. K. The ER in 3-D: a multifunctional dynamic membrane network. *Trends Cell Biol* **21**, 709 (2011).
447. Reid, D. W. & Nicchitta, C. v. Diversity and selectivity in mRNA translation on the endoplasmic reticulum. *Nat Rev Mol Cell Biol* **16**, 221 (2015).
448. Zhou, C. *et al.* Organelle-Based Aggregation and Retention of Damaged Proteins in Asymmetrically Dividing Cells. *Cell* **159**, 530 (2014).
449. Walter, P. & Blobel, G. Translocation of proteins across the endoplasmic reticulum. II. Signal recognition protein (SRP) mediates the selective binding to microsomal membranes of in-vitro-assembled polysomes synthesizing secretory protein. *J Cell Biol* **91**, 551–556 (1981).
450. Voorhees, R. M., Fernández, I. S., Scheres, S. H. W. & Hegde, R. S. Structure of the Mammalian Ribosome-Sec61 Complex to 3.4 Å Resolution. *Cell* **157**, 1632 (2014).
451. Savitz, A. J. & Meyer, D. I. 180-kD ribosome receptor is essential for both ribosome binding and protein translocation. *J Cell Biol* **120**, 853 (1993).
452. Schwarz, D. S. & Blower, M. D. The endoplasmic reticulum: structure, function and response to cellular signaling. *Cellular and Molecular Life Sciences* **73**, 79 (2016).
453. Stevens, F. J. & Argon, Y. Protein folding in the ER. *Semin Cell Dev Biol* **10**, 443–454 (1999).
454. McCaffrey, K. & Braakman, I. Protein quality control at the endoplasmic reticulum. *Essays Biochem* **60**, 227–235 (2016).
455. Stalder, L. *et al.* The rough endoplasmatic reticulum is a central nucleation site of siRNA-mediated RNA silencing. *EMBO J* **32**, 1115 (2013).
456. Jacquemyn, J., Cascalho, A. & Goodchild, R. E. The ins and outs of endoplasmic reticulum-controlled lipid biosynthesis. *EMBO Rep* **18**, 1905 (2017).
457. Brown, M. S. & Goldstein, J. L. A proteolytic pathway that controls the cholesterol content of membranes, cells, and blood. *Proc Natl Acad Sci U S A* **96**, 11041–11048 (1999).
458. Clapham, D. E. Calcium Signaling. *Cell* **131**, 1047–1058 (2007).
459. Meldolesi, J. & Pozzan, T. The endoplasmic reticulum Ca2+ store: A view from the lumen. *Trends Biochem Sci* **23**, 10–14 (1998).

460. Arruda, A. P. & Hotamisligil, G. S. Calcium Homeostasis and Organelle Function in the Pathogenesis of Obesity and Diabetes. *Cell Metab* **22**, 381–397 (2015).
461. Hetzer, M. W., Walther, T. C. & Mattaj, I. W. PUSHING THE ENVELOPE: Structure, Function, and Dynamics of the Nuclear Periphery. *http://dx.doi.org/10.1146/annurev.cellbio.21.090704.151152* **21**, 347–380 (2005).
462. Tzur, Y. B., Wilson, K. L. & Gruenbaum, Y. SUN-domain proteins: 'Velcro' that links the nucleoskeleton to the cytoskeleton. *Nature Reviews Molecular Cell Biology 2006 7:10* **7**, 782–788 (2006).
463. Westrate, L. M., Lee, J. E., Prinz, W. A. & Voeltz, G. K. Form follows function: the importance of endoplasmic reticulum shape. *Annu Rev Biochem* **84**, 791–811 (2015).
464. Voeltz, G. K., Rolls, M. M. & Rapoport, T. A. Structural organization of the endoplasmic reticulum. *EMBO Rep* **3**, 944 (2002).
465. Shibata, Y., Voeltz, G. K. & Rapoport, T. A. Rough sheets and smooth tubules. *Cell* **126**, 435–439 (2006).
466. West, M., Zurek, N., Hoenger, A. & Voeltz, G. K. A 3D analysis of yeast ER structure reveals how ER domains are organized by membrane curvature. *J Cell Biol* **193**, 333 (2011).
467. Block, B. A., Imagawa, T., Campbell, K. P. & Franzini-Armstrong, C. Structural evidence for direct interaction between the molecular components of the transverse tubule/sarcoplasmic reticulum junction in skeletal muscle. *J Cell Biol* **107**, 2587 (1988).
468. Nixon-Abell, J. *et al.* Increased spatiotemporal resolution reveals highly dynamic dense tubular matrices in the peripheral ER. *Science* **354**, (2016).
469. Schroeder, L. K. *et al.* Dynamic nanoscale morphology of the ER surveyed by STED microscopy. *Journal of Cell Biology* **218**, 83–96 (2019).
470. Weigel, A. v. *et al.* ER-to-Golgi protein delivery through an interwoven, tubular network extending from ER. *Cell* **184**, 2412-2429.e16 (2021).
471. Zucker, B. & Kozlov, M. M. Mechanism of shaping membrane nanostructures of endoplasmic reticulum. *Proc Natl Acad Sci U S A* **119**, e2116142119 (2022).
472. Voeltz, G. K., Prinz, W. A., Shibata, Y., Rist, J. M. & Rapoport, T. A. A class of membrane proteins shaping the tubular endoplasmic reticulum. *Cell* **124**, 573–586 (2006).
473. Shibata, Y. *et al.* The Reticulon and Dp1/Yop1p Proteins Form Immobile Oligomers in the Tubular Endoplasmic Reticulum. *J Biol Chem* **283**, 18892 (2008).
474. Yamamoto, Y., Yoshida, A., Miyazaki, N., Iwasaki, K. & Sakisaka, T. Arl6IP1 has the ability to shape the mammalian ER membrane in a reticulon-like fashion. *Biochemical Journal* **458**, 69–79 (2014).
475. English, A. R. & Voeltz, G. K. Rab10 GTPase regulates ER dynamics and morphology. *Nat Cell Biol* **15**, 169 (2013).
476. Hu, J. *et al.* A class of dynamin-like GTPases involved in the generation of the tubular ER network. *Cell* **138**, 549 (2009).

477. Brady, J. P., Claridge, J. K., Smith, P. G., Schnell, J. R. & DeGrado, W. F. A conserved amphipathic helix is required for membrane tubule formation by Yop1p. *Proc Natl Acad Sci U S A* **112**, E639–E648 (2015).
478. Shibata, Y. *et al.* Mechanisms determining the morphology of the peripheral ER. *Cell* **143**, 774 (2010).
479. Klopfenstein, D. R. *et al.* Subdomain-Specific Localization of Climp-63 (P63) in the Endoplasmic Reticulum Is Mediated by Its Luminal α-Helical Segment. *J Cell Biol* **153**, 1287 (2001).
480. Zhang, H. & Hu, J. Shaping the Endoplasmic Reticulum into a Social Network. *Trends Cell Biol* **26**, 934–943 (2016).
481. Scorrano, L. *et al.* Coming together to define membrane contact sites. *Nature Communications 2019 10:1* **10**, 1–11 (2019).
482. Eisenberg-Bord, M., Shai, N., Schuldiner, M. & Bohnert, M. A Tether Is a Tether Is a Tether: Tethering at Membrane Contact Sites. *Dev Cell* **39**, 395–409 (2016).
483. Raiborg, C., Wenzel, E. M. & Stenmark, H. ER-endosome contact sites: molecular compositions and functions. *EMBO J* **34**, 1848–1858 (2015).
484. Almeida, C. & Amaral, M. D. A central role of the endoplasmic reticulum in the cell emerges from its functional contact sites with multiple organelles. *Cellular and Molecular Life Sciences 2020 77:23* **77**, 4729–4745 (2020).
485. Vance, J. E. Phospholipid synthesis in a membrane fraction associated with mitochondria. *Journal of Biological Chemistry* **265**, 7248–7256 (1990).
486. Aoyama-Ishiwatari, S. & Hirabayashi, Y. Endoplasmic Reticulum–Mitochondria Contact Sites—Emerging Intracellular Signaling Hubs. *Front Cell Dev Biol* **9**, 1040 (2021).
487. Petrungaro, C. & Kornmann, B. Lipid exchange at ER-mitochondria contact sites: a puzzle falling into place with quite a few pieces missing. *Curr Opin Cell Biol* **57**, 71–76 (2019).
488. Abrisch, R. G., Gumbin, S. C., Wisniewski, B. T., Lackner, L. L. & Voeltz, G. K. Fission and fusion machineries converge at ER contact sites to regulate mitochondrial morphology. *Journal of Cell Biology* **219**, (2020).
489. Gomez-Suaga, P. *et al.* The ER-Mitochondria Tethering Complex VAPB-PTPIP51 Regulates Autophagy. *Curr Biol* **27**, 371–385 (2017).
490. Mesmin, B. *et al.* A four-step cycle driven by PI(4)P hydrolysis directs sterol/PI(4)P exchange by the ER-Golgi tether OSBP. *Cell* **155**, 830 (2013).
491. Hanada, K. *et al.* Molecular machinery for non-vesicular trafficking of ceramide. (2003).
492. Raiborg, C., Wenzel, E. M. & Stenmark, H. ER–endosome contact sites: molecular compositions and functions. *EMBO J* **34**, 1848 (2015).
493. van der Kant, R. & Neefjes, J. Small regulators, major consequences - Ca^{2+} and cholesterol at the endosome-ER interface. *J Cell Sci* **127**, 929–938 (2014).
494. Hugenroth, M. & Bohnert, M. Come a little bit closer! Lipid droplet-ER contact sites are getting crowded. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1867**, 118603 (2020).

495. Li, H. & Sun, S. Protein Aggregation in the ER: Calm behind the Storm. *Cells* **10**, (2021).
496. Marzec, M., Eletto, D. & Argon, Y. GRP94: an HSP90-like protein specialized for protein folding and quality control in the Endoplasmic Reticulum. *Biochim Biophys Acta* **1823**, 774 (2012).
497. Schoberer, J., Shin, Y. J., Vavra, U., Veit, C. & Strasser, R. Protein glycosylation in the ER. *Methods Mol Biol* **1691**, 205 (2018).
498. Kozlov, G. & Gehring, K. Calnexin cycle – structural features of the ER chaperone system. *FEBS J* **287**, 4322–4340 (2020).
499. Caramelo, J. J. & Parodi, A. J. A sweet code for glycoprotein folding. *FEBS Lett* **589**, 3379–3387 (2015).
500. Read, A. & Schröder, M. The Unfolded Protein Response: An Overview. *Biology (Basel)* **10**, (2021).
501. Sidrauski, C., Chapman, R. & Walter, P. The unfolded protein response: An intracellular signalling pathway with many surprising features. *Trends Cell Biol* **8**, 245–249 (1998).
502. Rutkowski, D. T. & Hegde, R. S. Regulation of basal cellular physiology by the homeostatic unfolded protein response. *J Cell Biol* **189**, 783 (2010).
503. Imagawa, Y., Hosoda, A., Sasaka, S. ichi, Tsuru, A. & Kohno, K. RNase domains determine the functional difference between IRE1α and IRE1β. *FEBS Lett* **582**, 656–660 (2008).
504. Almanza, A. *et al.* Endoplasmic reticulum stress signalling – from basic mechanisms to clinical applications. *FEBS J* **286**, 241 (2019).
505. Park, S. M., Kang, T. il & So, J. S. Roles of XBP1s in Transcriptional Regulation of Target Genes. *Biomedicines* **9**, (2021).
506. Hollien, J. & Weissman, J. S. Decay of endoplasmic reticulum-localized mRNAs during the unfolded protein response. *Science (1979)* **313**, 104–107 (2006).
507. Moncan, M. *et al.* Regulation of lipid metabolism by the unfolded protein response. *J Cell Mol Med* **25**, 1359 (2021).
508. Walter, P. & Ron, D. The Unfolded Protein Response: From Stress Pathway to Homeostatic Regulation. *Science (1979)* **334**, 1081–1086 (2011).
509. Marasco, O. N. J. M., Roussel, M. R. & Thakor, N. Probabilistic models of uORF-mediated ATF4 translation control. *Math Biosci* **343**, 108762 (2022).
510. Hetz, C., Chevet, E. & Harding, H. P. Targeting the unfolded protein response in disease. *Nature Reviews Drug Discovery 2013 12:9* **12**, 703–719 (2013).
511. Hillary, R. F. & Fitzgerald, U. A lifetime of stress: ATF6 in development and homeostasis. *J Biomed Sci* **25**, 1–10 (2018).
512. Ye, J. *et al.* ER stress induces cleavage of membrane-bound ATF6 by the same proteases that process SREBPs. *Mol Cell* **6**, 1355–1364 (2000).
513. Tabas, I. & Ron, D. Integrating the mechanisms of apoptosis induced by endoplasmic reticulum stress. *Nat Cell Biol* **13**, 184 (2011).

514. Tait, S. W. G. & Green, D. R. Mitochondria and cell death: outer membrane permeabilization and beyond. *Nature Reviews Molecular Cell Biology 2010 11:9* **11**, 621–632 (2010).
515. Woehlbier, U. & Hetz, C. Modulating stress responses by the UPRosome: A matter of life and death. *Trends Biochem Sci* **36**, 329–337 (2011).
516. Hetz, C. The unfolded protein response: controlling cell fate decisions under ER stress and beyond. *Nat Rev Mol Cell Biol* **13**, 89–102 (2012).
517. Riaz, T. A. *et al.* Role of Endoplasmic Reticulum Stress Sensor IRE1α in Cellular Physiology, Calcium, ROS Signaling, and Metaflammation. *Cells* **9**, (2020).
518. Adams, C. J., Kopp, M. C., Larburu, N., Nowak, P. R. & Ali, M. M. U. Structure and molecular mechanism of ER stress signaling by the unfolded protein response signal activator IRE1. *Front Mol Biosci* **6**, 11 (2019).
519. Credle, J. J., Finer-Moore, J. S., Papa, F. R., Stroud, R. M. & Walter, P. On the mechanism of sensing unfolded protein in the endoplasmic reticulum. *Proc Natl Acad Sci U S A* **102**, 18773–18784 (2005).
520. Karagöz, G. E. *et al.* An unfolded protein-induced conformational switch activates mammalian IRE1. *Elife* **6**, (2017).
521. Kimata, Y., Oikawa, D., Shimizu, Y., Ishiwata-Kimata, Y. & Kohno, K. A role for BiP as an adjustor for the endoplasmic reticulum stress-sensing protein Ire1. *J Cell Biol* **167**, 445 (2004).
522. Wang, P., Li, J., Tao, J. & Sha, B. The luminal domain of the ER stress sensor protein PERK binds misfolded proteins and thereby triggers PERK oligomerization. *J Biol Chem* **293**, 4110–4121 (2018).
523. Fun, X. H. & Thibault, G. Lipid bilayer stress and proteotoxic stress-induced unfolded protein response deploy divergent transcriptional and non-transcriptional programmes. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* **1865**, 158449 (2020).
524. Bertolotti, A., Zhang, Y., Hendershot, L. M., Harding, H. P. & Ron, D. Dynamic interaction of BiP and ER stress transducers in the unfolded-protein response. *Nature Cell Biology 2000 2:6* **2**, 326–332 (2000).
525. Amin-Wetzel, N. *et al.* A J-Protein Co-chaperone Recruits BiP to Monomerize IRE1 and Repress the Unfolded Protein Response. *Cell* **171**, 1625 (2017).
526. Carrara, M., Prischi, F., Nowak, P. R., Kopp, M. C. & Ali, M. M. U. Noncanonical binding of BiP ATPase domain to Ire1 and Perk is dissociated by unfolded protein CH1 to initiate ER stress signaling. *Elife* **4**, (2015).
527. Nikawa, J. -I & Yamashita, S. IRE1 encodes a putative protein kinase containing a membrane-spanning domain and is required for inositol phototrophy in Saccharomyces cerevisiae. *Mol Microbiol* **6**, 1441–1446 (1992).
528. Jonikas, M. C. *et al.* Comprehensive characterization of genes required for protein folding in the endoplasmic reticulum. *Science* **323**, 1693 (2009).

529. Feng, B. *et al.* The endoplasmic reticulum is the site of cholesterol-induced cytotoxicity in macrophages. *Nature Cell Biology 2003 5:9* **5**, 781–792 (2003).
530. Cunha, D. A. *et al.* Initiation and execution of lipotoxic ER stress in pancreatic beta-cells. *J Cell Sci* **121**, 2308–2318 (2008).
531. Spassieva, S. D., Mullen, T. D., Townsend, D. M. & Obeid, L. M. Disruption of ceramide synthesis by CerS2 down-regulation leads to autophagy and the unfolded protein response. *Biochem J* **424**, 273 (2009).
532. Fu, S. *et al.* Aberrant lipid metabolism disrupts calcium homeostasis causing liver endoplasmic reticulum stress in obesity. *Nature* **473**, 528–531 (2011).
533. Li, Y. *et al.* Enrichment of Endoplasmic Reticulum with Cholesterol Inhibits Sarcoplasmic-Endoplasmic Reticulum Calcium ATPase-2b Activity in Parallel with Increased Order of Membrane Lipids. *Journal of Biological Chemistry* **279**, 37030–37039 (2004).
534. Lajoie, P., Moir, R. D., Willis, I. M. & Snapp, E. L. Kar2p availability defines distinct forms of endoplasmic reticulum stress in living cells. *Mol Biol Cell* **23**, 955 (2012).
535. Hou, N. S. *et al.* Activation of the endoplasmic reticulum unfolded protein response by lipid disequilibrium without disturbed proteostasis in vivo. *Proc Natl Acad Sci U S A* **111**, (2014).
536. Promlek, T. *et al.* A Highlights from MBoC Selection: Membrane aberrancy and unfolded proteins activate the endoplasmic reticulum stress sensor Ire1 in different ways. *Mol Biol Cell* **22**, 3520 (2011).
537. Volmer, R., van der Ploeg, K. & Ron, D. Membrane lipid saturation activates endoplasmic reticulum unfolded protein response transducers through their transmembrane domains. *Proc Natl Acad Sci U S A* **110**, 4628–4633 (2013).
538. Tam, A. B. *et al.* The UPR Activator ATF6 Responds to Proteotoxic and Lipotoxic Stress by Distinct Mechanisms. *Dev Cell* **46**, 327-343.e7 (2018).
539. Chattopadhyay, A. *et al.* Cholesterol-Induced Phenotypic Modulation of Smooth Muscle Cells to Macrophage/Fibroblast-like Cells Is Driven by an Unfolded Protein Response. *Arterioscler Thromb Vasc Biol* **41**, 302–316 (2021).
540. Pineau, L. *et al.* Lipid-induced ER stress: synergistic effects of sterols and saturated fatty acids. *Traffic* **10**, 673–690 (2009).
541. Rennert, C. *et al.* Prolonged Lipid Accumulation in Cultured Primary Human Hepatocytes Rather Leads to ER Stress than Oxidative Stress. *Int J Mol Sci* **21**, 1–23 (2020).
542. Radanović, T. & Ernst, R. The Unfolded Protein Response as a Guardian of the Secretory Pathway. *Cells* **10**, (2021).
543. Halbleib, K. *et al.* Activation of the Unfolded Protein Response by Lipid Bilayer Stress. *Mol Cell* **67**, 673-684.e8 (2017).
544. Tam, A. B. *et al.* The UPR Activator ATF6 Responds to Proteotoxic and Lipotoxic Stress by Distinct Mechanisms. *Dev Cell* **46**, 327 (2018).
545. Goldstein, J. L., DeBose-Boyd, R. A. & Brown, M. S. Protein Sensors for Membrane Sterols. *Cell* **124**, 35–46 (2006).

546. Espenshade, P. J. & Hughes, A. L. Regulation of Sterol Synthesis in Eukaryotes. *https://doi.org/10.1146/annurev.genet.41.110306.130315* **41**, 401–427 (2007).
547. Hetz, C. & Glimcher, L. H. Fine tuning of the Unfolded Protein Response: Assembling the IRE1α interactome. *Mol Cell* **35**, 551 (2009).
548. Hetz, C. *et al.* Proapoptotic BAX and BAK modulate the unfolded protein response by a direct interaction with IRE1α. *Science (1979)* **312**, 572–576 (2006).
549. Gupta, S. *et al.* HSP72 Protects Cells from ER Stress-induced Apoptosis via Enhancement of IRE1α-XBP1 Signaling through a Physical Interaction. *PLoS Biol* **8**, (2010).
550. Pinkaew, D. *et al.* Fortilin binds IRE1α and prevents ER stress from signaling apoptotic cell death. *Nat Commun* **8**, (2017).
551. Urra, H., Pihán, P. & Hetz, C. The UPRosome - decoding novel biological outputs of IRE1α function. *J Cell Sci* **133**, (2020).
552. Hetz, C. & Papa, F. R. The Unfolded Protein Response and Cell Fate Control. *Mol Cell* **69**, 169–181 (2018).
553. Zhu, X., Zelmer, A., Kapfhammer, J. P. & Wellmann, S. Cold-inducible RBM3 inhibits PERK phosphorylation through cooperation with NF90 to protect cells from endoplasmic reticulum stress. *The FASEB Journal* **30**, 624–634 (2016).
554. Fonseca, S. G. *et al.* Wolfram syndrome 1 gene negatively regulates ER stress signaling in rodent and human cells. *J Clin Invest* **120**, 744 (2010).
555. Karagöz, G. E., Aragón, T. & Acosta-Alvear, D. Recent advances in signal integration mechanisms in the unfolded protein response. *F1000Res* **8**, (2019).
556. Grootjans, J., Kaser, A., Kaufman, R. J. & Blumberg, R. S. The unfolded protein response in immunity and inflammation. *Nat Rev Immunol* **16**, 469 (2016).
557. Martinon, F. & Glimcher, L. H. Regulation of Innate Immunity by signaling pathways emerging from the endoplasmic reticulum. *Curr Opin Immunol* **23**, 35 (2011).
558. Martinon, F., Chen, X., Lee, A. H. & Glimcher, L. H. Toll-like receptor activation of XBP1 regulates innate immune responses in macrophages. *Nat Immunol* **11**, 411 (2010).
559. Woo, C. W. *et al.* Adaptive Suppression of the ATF4-CHOP Branch of the Unfolded Protein Response by Toll-Like Receptor Signaling. *Nat Cell Biol* **11**, 1473 (2009).
560. Hotamisligil, G. S. Endoplasmic Reticulum Stress and the Inflammatory Basis of Metabolic Disease. *Cell* **140**, 900 (2010).
561. Back, S. H. *et al.* Translation attenuation through eIF2α phosphorylation prevents oxidative stress and maintains the differentiated state in beta cells. *Cell Metab* **10**, 13 (2009).
562. Lipson, K. L. *et al.* Regulation of insulin biosynthesis in pancreatic beta cells by an endoplasmic reticulum-resident protein kinase IRE1. *Cell Metab* **4**, 245–254 (2006).
563. Qiu, Y. *et al.* A crucial role for RACK1 in the regulation of glucose-stimulated IRE1α activation in pancreatic β-cells. *Sci Signal* **3**, ra7 (2010).
564. Lee, A. H., Scapa, E. F., Cohen, D. E. & Glimcher, L. H. Regulation of hepatic lipogenesis by the transcription factor XBP1. *Science* **320**, 1492 (2008).

565. Hotamisligil, G. S. Endoplasmic Reticulum Stress and the Inflammatory Basis of Metabolic Disease. *Cell* **140**, 900 (2010).
566. Özcan, U. *et al.* Endoplasmic reticulum stress links obesity, insulin action, and type 2 diabetes. *Science (1979)* **306**, 457–461 (2004).
567. Yang, Y., Cheung, H. H., Tu, J. J., Miu, K. K. & Chan, W. Y. New insights into the unfolded protein response in stem cells. *Oncotarget* **7**, 54010 (2016).
568. Hu, C. C. A., Dougan, S. K., McGehee, A. M., Love, J. C. & Ploegh, H. L. XBP-1 regulates signal transduction, transcription factors and bone marrow colonization in B cells. *EMBO J* **28**, 1624 (2009).
569. Zhang, K. *et al.* The unfolded protein response sensor IRE1α is required at 2 distinct steps in B cell lymphopoiesis. *Journal of Clinical Investigation* **115**, 268 (2005).
570. Huh, W. J. *et al.* XBP1 controls maturation of gastric zymogenic cells by induction of MIST1 and expansion of the rough endoplasmic reticulum. *Gastroenterology* **139**, 2038 (2010).
571. Hayashi, A. *et al.* The role of Brain-derived Neurotrophic Factor (BDNF)-induced XBP1 splicing during brain development. *Journal of Biological Chemistry* **282**, 34525–34534 (2007).
572. Christianson, J. C. & Carvalho, P. Order through destruction: how ER-associated protein degradation contributes to organelle homeostasis. *EMBO J* **41**, e109845 (2022).
573. Bhattacharya, A. & Qi, L. ER-associated degradation in health and disease – from substrate to organism. *J Cell Sci* **132**, (2019).
574. Reggiori, F. & Molinari, M. ER-PHAGY: MECHANISMS, REGULATION, AND DISEASES CONNECTED TO THE LYSOSOMAL CLEARANCE OF THE ENDOPLASMIC RETICULUM. *Physiol Rev* **102**, 1398–1448 (2022).
575. Wang, S. & Kaufman, R. J. The impact of the unfolded protein response on human disease. *J Cell Biol* **197**, 857 (2012).
576. Kaser, A., Martínez-Naves, E. & Blumberg, R. S. Endoplasmic reticulum stress: implications for inflammatory bowel disease pathogenesis. *Curr Opin Gastroenterol* **26**, 318–326 (2010).
577. Lee, J. & Ozcan, U. Unfolded Protein Response Signaling and Metabolic Diseases. *J Biol Chem* **289**, 1203 (2014).
578. Mao, T. *et al.* PKA phosphorylation couples hepatic inositol-requiring enzyme 1α to glucagon signaling in glucose metabolism. *Proc Natl Acad Sci U S A* **108**, 15852–15857 (2011).
579. Zhou, Y. *et al.* REGULATION OF GLUCOSE HOMEOSTASIS THROUGH XBP1-FOXO1 INTERACTION. *Nat Med* **17**, 356 (2011).
580. Wang, S. *et al.* IRE1α -XBP1s Induces PDI Expression to Increase MTP Activity for Hepatic VLDL Assembly and Lipid Homeostasis. *Cell Metab* **16**, 473 (2012).
581. Madden, E., Logue, S. E., Healy, S. J., Manie, S. & Samali, A. The role of the unfolded protein response in cancer progression: From oncogenesis to chemoresistance. *Biol Cell* **111**, 1–17 (2019).

582. Grandjean, J. M. D. & Luke Wiseman, R. Small molecule strategies to harness the unfolded protein response: where do we go from here? *J Biol Chem* **295**, 15692 (2020).
583. Hetz, C., Axten, J. M. & Patterson, J. B. Pharmacological targeting of the unfolded protein response for disease intervention. *Nature Chemical Biology 2019 15:8* **15**, 764–775 (2019).
584. van Heesch, S. *et al.* Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes. *Genome Biol* **15**, 1–12 (2014).
585. al Bitar, S. & Gali-Muhtasib, H. The Role of the Cyclin Dependent Kinase Inhibitor p21cip1/waf1 in Targeting Cancer: Molecular Mechanisms and Novel Therapeutics. *Cancers (Basel)* **11**, (2019).
586. Duggimpudi, S., Larsson, E., Nabhani, S., Borkhardt, A. & Hoell, J. I. The cell cycle regulator CCDC6 is a key target of RNA-binding protein EWS. *PLoS One* **10**, 1–12 (2015).
587. Levine, M. S. & Holland, A. J. The impact of mitotic errors on cell proliferation and tumorigenesis. *Genes Dev* **32**, 620 (2018).
588. Li, K., Hope, C. M., Wang, X. A. & Wang, J. P. RiboDiPA: a novel tool for differential pattern analysis in Ribo-seq data. *Nucleic Acids Res* **48**, 12016–12029 (2020).
589. Schmidt, E. K., Clavarino, G., Ceppi, M. & Pierre, P. SUnSET, a nonradioactive method to monitor protein synthesis. *Nature Methods 2009 6:4* **6**, 275–277 (2009).
590. Goodman, C. A. & Hornberger, T. A. Measuring protein synthesis with SUnSET: a valid alternative to traditional techniques? *Exerc Sport Sci Rev* **41**, 107 (2013).
591. Chassé, H., Boulben, S., Costache, V., Cormier, P. & Morales, J. Analysis of translation using polysome profiling. *Nucleic Acids Res* **45**, e15–e15 (2017).
592. Takahama, K., Sugimoto, C., Arai, S., Kurokawa, R. & Oyoshi, T. Loop lengths of G-quadruplex structures affect the G-quadruplex DNA binding selectivity of the RGG motif in Ewing's sarcoma. *Biochemistry* **50**, 5369–5378 (2011).
593. Mayr, C. What Are 3′ UTRs Doing? *Cold Spring Harb Perspect Biol* **11**, (2019).
594. Leppek, K., Das, R. & Barna, M. Functional 5′ UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nat Rev Mol Cell Biol* **19**, 158 (2018).
595. Ray, D. *et al.* A compendium of RNA-binding motifs for decoding gene regulation. *Nature* **499**, 172 (2013).
596. Dassi, E. Handshakes and Fights: The Regulatory Interplay of RNA-Binding Proteins. *Front Mol Biosci* **4**, 67 (2017).
597. Loughlin, F. E. *et al.* The Solution Structure of FUS Bound to RNA Reveals a Bipartite Mode of RNA Recognition with Both Sequence and Shape Specificity. *Mol Cell* **73**, 490-504.e6 (2019).
598. van Nostrand, E. L. *et al.* A large-scale binding and functional map of human RNA-binding proteins. *Nature* **583**, 711 (2020).
599. Lin, J. C. & Tarn, W. Y. RNA-binding Motif Protein 4 Translocates to Cytoplasmic Granules and Suppresses Translation via Argonaute2 during Muscle Cell Differentiation. *J Biol Chem* **284**, 34658 (2009).

600. Audano, M. *et al.* Zc3h10 regulates adipogenesis by controlling translation and F-actin/mitochondria interaction. *J Cell Biol* **220**, (2021).
601. Cho, J. *et al.* LIN28A is a suppressor of ER-associated translation in embryonic stem cells. *Cell* **151**, 765–777 (2012).
602. Hu, W., Yuan, B. & Lodish, H. F. Cpeb4-mediated translational regulatory circuitry controls terminal erythroid differentiation. *Dev Cell* **30**, 660 (2014).
603. Kumar, R., Poria, D. K. & Ray, P. S. RNA-binding proteins La and HuR cooperatively modulate translation repression of PDCD4 mRNA. *Journal of Biological Chemistry* **296**, 100154 (2021).
604. Piecyk, M. *et al.* TIA-1 is a translational silencer that selectively regulates the expression of TNF-α. *EMBO J* **19**, 4154 (2000).
605. Ostareck, D. H. *et al.* mRNA Silencing in Erythroid Differentiation: hnRNP K and hnRNP E1 Regulate 15-Lipoxygenase Translation from the 3′ End. *Cell* **89**, 597–606 (1997).
606. Barreau, C., Watrin, T., Beverley Osborne, H. & Paillard, L. Protein expression is increased by a class III AU-rich element and tethered CUG-BP1. *Biochem Biophys Res Commun* **347**, 723–730 (2006).
607. Zhang, T. *et al.* FUS Regulates Activity of MicroRNA-mediated Gene Silencing. *Mol Cell* **69**, 787 (2018).
608. Kuzuo Glu-Öztürk, D. *et al.* miRISC and the CCR4–NOT complex silence mRNA targets independently of 43S ribosomal scanning. *EMBO J* **35**, 1186–1203 (2016).
609. Francisco-Velilla, R., Fernandez-Chamorro, J., Ramajo, J. & Martinez-Salas, E. The RNA-binding protein Gemin5 binds directly to the ribosome and regulates global translation. *Nucleic Acids Res* **44**, 8335 (2016).
610. Gan, P. *et al.* RBPMS is an RNA-binding protein that mediates cardiomyocyte binucleation and cardiovascular development. *Dev Cell* **57**, 959-973.e7 (2022).
611. Angenstein, F. *et al.* Proteomic characterization of messenger ribonucleoprotein complexes bound to nontranslated or translated poly(A) mRNAs in the rat cerebral cortex. *J Biol Chem* **280**, 6496–6503 (2005).
612. Cenik, C. *et al.* Integrative analysis of RNA, translation, and protein levels reveals distinct regulatory variation across humans. *Genome Res* **25**, 1610 (2015).
613. Hao, P. *et al.* Eukaryotic translation initiation factors as promising targets in cancer therapy. *Cell Communication and Signaling* **18**, 1–20 (2020).
614. Keiper, B. D. Cap-Independent mRNA Translation in Germ Cells. *Int J Mol Sci* **20**, (2019).
615. Sweeney, T. R., Abaeva, I. S., Pestova, T. v. & Hellen, C. U. T. The mechanism of translation initiation on Type 1 picornavirus IRESs. *EMBO J* **33**, 76 (2014).
616. Deniz, N., Lenarcic, E. M., Landry, D. M. & Thompson, S. R. Translation initiation factors are not required for Dicistroviridae IRES function in vivo. *RNA* **15**, 932 (2009).
617. Kuzuoğlu-Öztürk, D. *et al.* miRISC and the CCR4–NOT complex silence mRNA targets independently of 43S ribosomal scanning. *EMBO J* **35**, 1186 (2016).

618. Sun, L. *et al.* The oncomicropeptide APPLE promotes hematopoietic malignancy by enhancing translation initiation. *Mol Cell* **81**, 4493-4508.e9 (2021).
619. Onal, G., Kutlu, O., Gozuacik, D. & Dokmeci Emre, S. Lipid Droplets in Health and Disease. *Lipids in Health and Disease 2017 16:1* **16**, 1–15 (2017).
620. Han, J. & Kaufman, R. J. The role of ER stress in lipid metabolism and lipotoxicity. *J Lipid Res* **57**, 1329–1338 (2016).
621. Frakes, A. E. & Dillin, A. The UPRER: Sensor and Coordinator of Organismal Homeostasis. *Mol Cell* **66**, 761–771 (2017).
622. Volmer, R. & Ron, D. Lipid-dependent regulation of the unfolded protein response. *Curr Opin Cell Biol* **33**, 67 (2015).
623. Sicari, D., Delaunay-Moisan, A., Combettes, L., Chevet, E. & Igbaria, A. A guide to assessing endoplasmic reticulum homeostasis and stress in mammalian systems. *FEBS J* **287**, 27–42 (2020).
624. Schultz, A. M. & Oroszlan, S. Tunicamycin inhibits glycosylation of precursor polyprotein encoded by env gene of Rauscher murine leukemia virus. *Biochem Biophys Res Commun* **86**, 1206–1213 (1979).
625. Thastrup, O., Cullen, P. J., Drobak, B. K., Hanley, M. R. & Dawson, A. P. Thapsigargin, a tumor promoter, discharges intracellular Ca2+ stores by specific inhibition of the endoplasmic reticulum Ca2(+)-ATPase. *Proc Natl Acad Sci U S A* **87**, 2466 (1990).
626. van Herpen, N. A. & Schrauwen-Hinderling, V. B. Lipid accumulation in non-adipose tissue and lipotoxicity. *Physiol Behav* **94**, 231–241 (2008).
627. Haze, K., Yoshida, H., Yanagi, H., Yura, T. & Mori, K. Mammalian Transcription Factor ATF6 Is Synthesized as a Transmembrane Protein and Activated by Proteolysis in Response to Endoplasmic Reticulum Stress. *Mol Biol Cell* **10**, 3787 (1999).
628. Yang, Y. *et al.* Transcription Factor C/EBP Homologous Protein in Health and Diseases. *Front Immunol* **8**, (2017).
629. Mihailidou, C., Papavassiliou, A. G. & Kiaris, H. A crosstalk between p21 and UPR-induced transcription factor C/EBP homologous protein (CHOP) linked to type 2 diabetes. *Biochimie* **99**, 19–27 (2014).
630. Ariyasu, D., Yoshida, H. & Hasegawa, Y. Endoplasmic Reticulum (ER) Stress and Endocrine Disorders. *Int J Mol Sci* **18**, (2017).
631. Cubillos-Ruiz, J. R., Bettigole, S. E. & Glimcher, L. H. Tumorigenic and Immunosuppressive Effects of Endoplasmic Reticulum Stress in Cancer. *Cell* **168**, 692–706 (2017).
632. Sehgal, P. *et al.* Inhibition of the sarco/endoplasmic reticulum (ER) Ca2+-ATPase by thapsigargin analogs induces cell death via ER Ca2+ depletion and the unfolded protein response. *J Biol Chem* **292**, 19656 (2017).
633. Guha, P., Kaptan, E., Gade, P., Kalvakolanu, D. v. & Ahmed, H. Tunicamycin induced endoplasmic reticulum stress promotes apoptosis of prostate cancer cells by activating mTORC1. *Oncotarget* **8**, 68191 (2017).

634. Yang, L. *et al.* Oxidative and endoplasmic reticulum stresses are involved in palmitic acid-induced H9c2 cell apoptosis. *Biosci Rep* **39**, 20190225 (2019).
635. Chaitanya, G. V., Alexander, J. S. & Babu, P. P. PARP-1 cleavage fragments: Signatures of cell-death proteases in neurodegeneration. *Cell Communication and Signaling* **8**, 1–11 (2010).
636. Bentley, D. L. The union of transcription and mRNA processing: 20 years of coupling. *RNA* vol. 21 569–570 Preprint at https://doi.org/10.1261/rna.050740.115 (2015).
637. Espinosa-Cantú, A., Cruz-Bonilla, E., Noda-Garcia, L. & DeLuna, A. Multiple Forms of Multifunctional Proteins in Health and Disease. *Frontiers in Cell and Developmental Biology* vol. 8 451 Preprint at https://doi.org/10.3389/fcell.2020.00451 (2020).
638. Gerstberger, S., Hafner, M. & Tuschl, T. A census of human RNA-binding proteins. *Nat Rev Genet* **15**, 829–845 (2014).
639. Bagheri, A., Astafev, A., Al-Hashimy, T. & Jiang, P. Tracing Translational Footprint by Ribo-Seq: Principle, Workflow, and Applications to Understand the Mechanism of Human Diseases. *Cells* **11**, 2966 (2022).
640. Re, A., Waldron, L. & Quattrone, A. Control of Gene Expression by RNA Binding Protein Action on Alternative Translation Initiation Sites. *PLoS Comput Biol* **12**, 1005198 (2016).
641. Eastman, G., Smircich, P. & Sotelo-Silveira, J. R. Following Ribosome Footprints to Understand Translation at a Genome Wide Level. *Comput Struct Biotechnol J* **16**, 167 (2018).
642. Li, G. W., Burkhardt, D., Gross, C. & Weissman, J. S. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* **157**, 624 (2014).
643. Liu, Y., Beyer, A. & Aebersold, R. On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* **165**, 535–550 (2016).
644. Liu, T. Y. *et al.* Time-Resolved Proteomics Extends Ribosome Profiling-Based Measurements of Protein Synthesis Dynamics. *Cell Syst* **4**, 636 (2017).
645. Schwartz, J. C. *et al.* FUS binds the CTD of RNA polymerase II and regulates its phosphorylation at Ser2. *Genes Dev* **26**, 2690 (2012).
646. Ishigaki, S. *et al.* Position-dependent FUS-RNA interactions regulate alternative splicing events and transcriptions. *Sci Rep* **2**, (2012).
647. Mastrocola, A. S., Kim, S. H., Trinh, A. T., Rodenkirch, L. A. & Tibbetts, R. S. The RNA-binding Protein Fused in Sarcoma (FUS) Functions Downstream of Poly(ADP-ribose) Polymerase (PARP) in Response to DNA Damage. *J Biol Chem* **288**, 24731 (2013).
648. Fujii, R. & Takumi, T. TLS facilitates transport of mRNA encoding an actin-stabilizing protein to dendritic spines. *J Cell Sci* **118**, 5755–5765 (2005).
649. Morlando, M. *et al.* FUS stimulates microRNA biogenesis by facilitating co-transcriptional Drosha recruitment. *EMBO J* **31**, 4502–4510 (2012).
650. Yasuda, K. *et al.* The RNA-binding protein Fus directs translation of localized mRNAs in APC-RNP granules. *J Cell Biol* **203**, 737–746 (2013).

651. Murakami, T. *et al.* ALS/FTD Mutation-Induced Phase Transition of FUS Liquid Droplets and Reversible Hydrogels into Irreversible Hydrogels Impairs RNP Granule Function. *Neuron* **88**, 678–690 (2015).
652. Tischbein, M. *et al.* The RNA-binding protein FUS/TLS undergoes calcium-mediated nuclear egress during excitotoxic stress and is required for GRIA2 mRNA processing. *J Biol Chem* **294**, 10194–10210 (2019).
653. Patel, A. *et al.* A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* **162**, 1066–1077 (2015).
654. Shelkovnikova, T. A., Robinson, H. K., Southcombe, J. A., Ninkina, N. & Buchman, V. L. Multistep process of FUS aggregation in the cell cytoplasm involves RNA-dependent and RNA-independent mechanisms. *Hum Mol Genet* **23**, 5211–5226 (2014).
655. Marko, M., Vlassis, A., Guialis, A. & Leichter, M. Domains involved in TAF15 subcellular localisation: Dependence on cell type and ongoing transcription. *Gene* **506**, 331–338 (2012).
656. Blechingberg, J., Luo, Y., Bolund, L., Damgaard, C. K. & Nielsen, A. L. Gene Expression Responses to FUS, EWS, and TAF15 Reduction and Stress Granule Sequestration Analyses Identifies FET-Protein Non-Redundant Functions. *PLoS One* **7**, e46251 (2012).
657. Qi, T., Xu, Y., Zhou, T. & Gu, W. The Evolution of G-quadruplex Structure in mRNA Untranslated Region. *Evol Bioinform Online* **17**, (2021).
658. Song, J., Perreault, J.-P., Topisirovic, I. & Richard, S. RNA G-quadruplexes and their potential regulatory roles in translation. *Translation* **4**, e1244031 (2016).
659. Haugen, R. J., Arvola, R. M., Connacher, R. P., Roden, R. T. & Goldstrohm, A. C. A conserved domain of Drosophila RNA-binding protein Pumilio interacts with multiple CCR4–NOT deadenylase complex subunits to repress target mRNAs. *Journal of Biological Chemistry* **298**, (2022).
660. Loedige, I. *et al.* The NHL domain of BRAT is an RNA-binding domain that directly contacts the hunchback mRNA for regulation. *Genes Dev* **28**, 749–764 (2014).
661. Kharel, P., Becker, G., Tsvetkov, V. & Ivanov, P. Properties and biological impact of RNA G-quadruplexes: from order to turmoil and back. *Nucleic Acids Res* **48**, 12534–12555 (2020).
662. Blackwell, E., Zhang, X. & Ceman, S. Arginines of the RGG box regulate FMRP association with polyribosomes and mRNA. *Hum Mol Genet* **19**, 1314 (2010).
663. Bouvet, P., Diaz, J. J., Kindbeiter, K., Madjar, J. J. & Amalric, F. Nucleolin interacts with several ribosomal proteins through its RGG domain. *Journal of Biological Chemistry* **273**, 19025–19029 (1998).
664. Lindström, M. S., Jin, A., Deisenroth, C., Wolf, G. W. & Zhang, Y. Cancer-Associated Mutations in the MDM2 Zinc Finger Domain Disrupt Ribosomal Protein Interaction and Attenuate MDM2-Induced p53 Degradation. *Mol Cell Biol* **27**, 1056 (2007).
665. Choi, S. G. *et al.* Maximizing binary interactome mapping with a minimal number of assays. *Nature Communications 2019 10:1* **10**, 1–13 (2019).

666. Paronetto, M. P. Ewing sarcoma protein: a key player in human cancer. *Int J Cell Biol* **2013**, 642853 (2013).
667. Song, P., Yang, F., Jin, H. & Wang, X. The regulation of protein translation and its implications for cancer. *Signal Transduction and Targeted Therapy 2021 6:1* **6**, 1–9 (2021).
668. Jishi, A., Qi, X. & Miranda, H. C. Implications of mRNA translation dysregulation for neurological disorders. *Semin Cell Dev Biol* **114**, 11 (2021).
669. Ricci, E. P. *et al.* miRNA repression of translation in vitro takes place during 43S ribosomal scanning. *Nucleic Acids Res* **41**, 586 (2013).
670. Colón-Ramos, D. A. *et al.* Direct ribosomal binding by a cellular inhibitor of translation. *Nat Struct Mol Biol* **13**, 103 (2006).
671. Wang, B., Yanez, A. & Novina, C. D. MicroRNA-repressed mRNAs contain 40S but not 60S components. *Proc Natl Acad Sci U S A* **105**, 5343–5348 (2008).
672. Al-Jubran, K. *et al.* Visualization of the joining of ribosomal subunits reveals the presence of 80S ribosomes in the nucleus. *RNA* **19**, 1669 (2013).
673. Hui, D. J., Terenzi, F., Merrick, W. C. & Sen, G. C. Mouse p56 blocks a distinct function of eukaryotic initiation factor 3 in translation initiation. *Journal of Biological Chemistry* **280**, 3433–3440 (2005).
674. Eshraghi, M. *et al.* Mutant Huntingtin stalls ribosomes and represses protein synthesis in a cellular model of Huntington disease. *Nat Commun* **12**, (2021).
675. Zhang, S. *et al.* Analysis of Ribosome Stalling and Translation Elongation Dynamics by Deep Learning. *Cell Syst* **5**, 212-220.e6 (2017).
676. Chyzyńska, K., Labun, K., Jones, C., Grellscheid, S. N. & Valen, E. Deep conservation of ribosome stall sites across RNA processing genes. *NAR Genom Bioinform* **3**, 1–13 (2021).
677. do Couto Bordignon, P. & Pechmann, S. Inferring translational heterogeneity from Saccharomyces cerevisiae ribosome profiling. *FEBS J* **288**, 4541–4559 (2021).
678. Ramazi, S. & Zahiri, J. Post-translational modifications in proteins: resources, tools and prediction methods. *Database (Oxford)* **2021**, (2021).
679. Kim, J., Lee, J. M., Branton, P. E. & Pelletier, J. Modification of EWS/WT1 functional properties by phosphorylation. *Proc Natl Acad Sci U S A* **96**, 14300–14305 (1999).
680. Kim, J., Lee, J. M., Branton, P. E. & Pelletier, J. Modulation of EWS/WT1 activity by the v-Src protein tyrosine kinase. *FEBS Lett* **474**, 121–128 (2000).
681. Guinamard, R., Fougereau, M. & SECKINGER, P. The SH3 domain of Bruton's tyrosine kinase interacts with Vav, Sam68 and EWS. *Scand J Immunol* **45**, 587–595 (1997).
682. Bachmaier, R. *et al.* O-GlcNAcylation is involved in the transcriptional activity of EWS-FLI1 in Ewing's sarcoma. *Oncogene* **28**, 1280 (2009).
683. Zhang, T. *et al.* Acetylation dependent translocation of EWSR1 regulates CHK2 alternative splicing in response to DNA damage. *Oncogene 2022 41:29* **41**, 3694–3704 (2022).
684. Araya, N. *et al.* Transcriptional down-regulation through nuclear exclusion of EWS methylated by PRMT1. *Biochem Biophys Res Commun* **329**, 653–660 (2005).

685. Bedford, M. T. Arginine methylation at a glance. *J Cell Sci* **120**, 4243–4246 (2007).
686. Tibshirani, M. *et al.* Cytoplasmic sequestration of FUS/TLS associated with ALS alters histone marks through loss of nuclear protein arginine methyltransferase 1. *Hum Mol Genet* **24**, 773–786 (2015).
687. Jobert, L., Argentini, M. & Tora, L. PRMT1 mediated methylation of TAF15 is required for its positive gene regulatory function. *Exp Cell Res* **315**, 1273–1286 (2009).
688. Park, J. H. & Lee, S. B. An essential role for Ewing sarcoma gene (EWS) in early white adipogenesis. *Obesity* **23**, 138–144 (2015).
689. Lee, S., Mardinoglu, A., Zhang, C., Lee, D. & Nielsen, J. Dysregulated signaling hubs of liver lipid metabolism reveal hepatocellular carcinoma pathogenesis. *Nucleic Acids Res* **44**, 5529–5539 (2016).
690. Dong, X. H. *et al.* The Long Noncoding RNA RP11-728F11.4 Promotes Atherosclerosis. *Arterioscler Thromb Vasc Biol* **41**, 1191–1204 (2021).
691. Reimold, A. M. *et al.* Plasma cell differentiation requires the transcription factor XBP-1. *Nature 2001 412:6844* **412**, 300–307 (2001).
692. Nakanishi, K., Sudo, T. & Morishima, N. Endoplasmic reticulum stress signaling transmitted by ATF6 mediates apoptosis during muscle development. *J Cell Biol* **169**, 555 (2005).
693. Benali-Furet, N. L. *et al.* Hepatitis C virus core triggers apoptosis in liver cells by inducing ER stress and ER calcium depletion. *Oncogene 2005 24:31* **24**, 4921–4933 (2005).
694. Shyu, P. *et al.* Membrane phospholipid alteration causes chronic ER stress through early degradation of homeostatic ER-resident proteins. *Scientific Reports 2019 9:1* **9**, 1–15 (2019).
695. Siwecka, N. *et al.* The Structure, Activation and Signaling of IRE1 and Its Role in Determining Cell Fate. *Biomedicines* **9**, 1–29 (2021).
696. Lipke, K., Kubis-Kubiak, A. & Piwowar, A. Molecular Mechanism of Lipotoxicity as an Interesting Aspect in the Development of Pathological States—Current View of Knowledge. *Cells* **11**, (2022).
697. Bhattarai, K. R., Riaz, T. A., Kim, H. R. & Chae, H. J. The aftermath of the interplay between the endoplasmic reticulum stress response and redox signaling. *Experimental & Molecular Medicine 2021 53:2* **53**, 151–167 (2021).
698. Gold, L. T. & Masson, G. R. GCN2: roles in tumour development and progression. *Biochem Soc Trans* **50**, 737 (2022).
699. Jin, H. O., Hong, S. E., Kim, J. Y., Jang, S. K. & Park, I. C. Amino acid deprivation induces AKT activation by inducing GCN2/ATF4/REDD1 axis. *Cell Death Dis* **12**, (2021).
700. Cabodevilla, A. G. *et al.* Cell Survival during Complete Nutrient Deprivation Depends on Lipid Droplet-fueled β-Oxidation of Fatty Acids. *J Biol Chem* **288**, 27777 (2013).
701. Mateus, D., Marini, E. S., Progida, C. & Bakke, O. Rab7a modulates ER stress and ER morphology. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1865**, 781–793 (2018).

702. Kerselidou, D. *et al.* Alternative glycosylation controls endoplasmic reticulum dynamics and tubular extension in mammalian cells. *Sci Adv* **7**, (2021).
703. Gani, A. R., Uppala, J. K. & Ramaiah, K. V. A. Tauroursodeoxycholic acid prevents stress induced aggregation of proteins in vitro and promotes PERK activation in HepG2 cells. *Arch Biochem Biophys* **568**, 8–15 (2015).
704. Blomme, A. *et al.* THEM6-mediated reprogramming of lipid metabolism supports treatment resistance in prostate cancer. *EMBO Mol Med* **14**, (2022).
705. Lee, J., Shin, M. K., Ryu, D. K., Kim, S. & Ryu, W. S. Insertion and deletion mutagenesis by overlap extension PCR. *Methods in Molecular Biology* **634**, 137–146 (2010).
706. Lykke-Andersen, J., Shu, M. di & Steitz, J. A. Human Upf proteins target an mRNA for nonsense-mediated decay when downstream of a termination codon. *Cell* **103**, 1121–1131 (2000).
707. Duarte, M. *et al.* Rotavirus Infection Alters Splicing of the Stress-Related Transcription Factor XBP1. *J Virol* **93**, (2019).
708. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 1–21 (2014).
709. Rambout, X. *et al.* The transcription factor ERG recruits CCR4–NOT to control mRNA decay and mitotic progression. *Nature Structural & Molecular Biology 2016 23:7* **23**, 663–672 (2016).
710. Miras, M., Allen Miller, W., Truniger, V. & Aranda, M. A. Non-canonical Translation in Plant RNA Viruses. *Front Plant Sci* **8**, 494 (2017).
711. Bravo, R. *et al.* Endoplasmic Reticulum and the Unfolded Protein Response: Dynamics and Metabolic Integration. *Int Rev Cell Mol Biol* **301**, 215 (2013).
712. Re-envisioning the endoplasmic reticulum | National Institutes of Health (NIH). https://www.nih.gov/news-events/nih-research-matters/re-envisioning-endoplasmic-reticulum.
713. Hetz, C. & Papa, F. R. The Unfolded Protein Response and Cell Fate Control. *Mol Cell* **69**, 169–181 (2018).

8 PUBLICATIONS

Alternative glycosylation controls endoplasmic reticulum dynamics and tubular extension in mammalian cells.

(*Science advances. May. 2021).* PMID: 33962942.

Kerselidou D, Dohai BS, Nelson DR, Daakour S, De Cock N, **Hassoun ZAO**, Kim DK, Olivet J, El Assal DC, Jaiswal A, Alzahmi A, Saha D, Pain C, Matthijssens F, Lemaitre P, Herfs M, Chapuis J, Ghesquiere B, Vertommen D, Kriechbaumer V, Knoops K, Lopez-Iglesias C, van Zandvoort M, Lambert JC, Hanson J, Desmet C, Thiry M, Lauersen KJ, Vidal M, Van Vlierberghe P, Dequiedt F, Salehi-Ashtiani K, Twizere JC.

Abstract

The endoplasmic reticulum (ER) is a central eukaryotic organelle with a tubular network made of hairpin proteins linked by hydrolysis of guanosine triphosphate nucleotides. Among posttranslational modifications initiated at the ER level, glycosylation is the most common reaction. However, our understanding of the impact of glycosylation on the ER structure remains unclear. Here, we show that exostosin-1 (EXT1) glycosyltransferase, an enzyme involved in *N*-glycosylation, is a key regulator of ER morphology and dynamics. We have integrated multiomics and superresolution imaging to characterize the broad effect of *EXT1* inactivation, including the ER shape-dynamics-function relationships in mammalian cells. We have observed that inactivating *EXT1* induces cell enlargement and enhances metabolic switches such as protein secretion. In particular, suppressing *EXT1* in mouse thymocytes causes developmental dysfunctions associated with the ER network extension. Last, our data illuminate the physical and functional aspects of the ER proteome-glycome-lipidome structure axis, with implications in biotechnology and medicine.

ERG transcription factors have a splicing regulatory function involving RBFOX2 that is altered in the EWS-FLI1 oncogenic fusion.

(*Nucleic Acids Res. May 2021)*. PMID: 34009296.

Saulnier O, Guedri-Idjouadiene K, Aynaud MM, Chakraborty A, Bruyr J, Pineau J, O'Grady T, Mirabeau O, Grossetête S, Galvan B, Claes M, **Al Oula Hassoun Z**, Sadacca B, Laud K, Zaïdi S, Surdez D, Baulande S, Rambout X, Tirode F, Dutertre M, Delattre O, Dequiedt F.

Abstract

ERG family proteins (ERG, FLI1 and FEV) are a subfamily of ETS transcription factors with key roles in physiology and development. In Ewing sarcoma, the oncogenic fusion protein EWS-FLI1 regulates both transcription and alternative splicing of pre-messenger RNAs. However, whether wild-type ERG family proteins might regulate splicing is unknown. Here, we show that wild-type ERG proteins associate with spliceosomal components, are found on nascent RNAs, and induce alternative splicing when recruited onto a reporter minigene. Transcriptomic analysis revealed that ERG and FLI1 regulate large numbers of alternative spliced exons (ASEs) enriched with RBFOX2 motifs and co-regulated by this splicing factor. ERG and FLI1 are associated with RBFOX2 via their conserved carboxy-terminal domain, which is present in EWS-FLI1. Accordingly, EWS-FLI1 is also associated with RBFOX2 and regulates ASEs enriched in RBFOX2 motifs. However, in contrast to wild-type ERG and FLI1, EWS-FLI1 often antagonizes RBFOX2 effects on exon inclusion. In particular, EWS-FLI1 reduces RBFOX2 binding to the ADD3 pre-mRNA, thus increasing its long isoform, which represses the mesenchymal phenotype of Ewing sarcoma cells. Our findings reveal a RBFOX2-mediated splicing regulatory function of wild-type ERG family proteins, that is altered in EWS-FLI1 and contributes to the Ewing sarcoma cell phenotype.

Expression and purification of recombinant extracellular sulfatase HSulf-2 allows deciphering of enzyme sub-domains coordinated role for the binding and 6-O-desulfation of heparan sulfate.

(Cell Mol Life Sci. May 2019). PMID: 30788513

Amal Seffouh, Rana El Masri, Olga Makshakova, Evelyne Gout, **Zahra El Oula Hassoun**, Jean-Pierre Andrieu, Hugues Lortat-Jacob, Romain R Vivès

Abstract

Through their ability to edit 6-O-sulfation pattern of Heparan sulfate (HS) polysaccharides, Sulf extracellular endosulfatases have emerged as critical regulators of many biological processes, including tumor progression. However, study of Sulfs remains extremely intricate and progress in characterizing their functional and structural features has been hampered by limited access to recombinant enzyme. In this study, we unlock this critical bottleneck, by reporting an efficient expression and purification system of recombinant HSulf-2 in mammalian HEK293 cells. This novel source of enzyme enabled us to investigate the way the enzyme domain organization dictates its functional properties. By generating mutants, we confirmed previous studies that HSulf-2 catalytic (CAT) domain was sufficient to elicit arylsulfatase activity and that its hydrophilic (HD) domain was necessary for the enzyme 6-O-endosulfatase activity. However, we demonstrated for the first time that high-affinity binding of HS substrates occurred through the coordinated action of both domains, and we identified and characterized 2 novel HS binding sites within the CAT domain. Altogether, our findings contribute to better understand the molecular mechanism governing HSulf-2 substrate recognition and processing. Furthermore, access to purified recombinant protein opens new perspectives for the resolution of HSulf structure and molecular features, as well as for the development of Sulf-specific inhibitors.