UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS

Analysis of translation of 5' untranslated regions in cancer

"*Documento Definitivo*"

Doutoramento em Biologia
Especialidade de Biologia de Sistemas

Joana Filipa Pires Silva

Tese orientada por:
Doutora Luísa Romão
Doutor Augusto Luchessi

Documento especialmente elaborado para a obtenção do grau de doutor

2019

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS

Analysis of translation of 5' untranslated regions in cancer

Doutoramento em Biologia

Especialidade de Biologia de Sistemas

Joana Filipa Pires Silva

Tese orientada por:

Doutora Luísa Romão

Doutor Augusto Luchessi

Júri:

Presidente:

- Doutor Rui Manuel dos Santos Malhó, Professor Catedrático e Presidente do Departamento de Biologia Vegetal da Faculdade de Ciências da Universidade de Lisboa

Vogais:

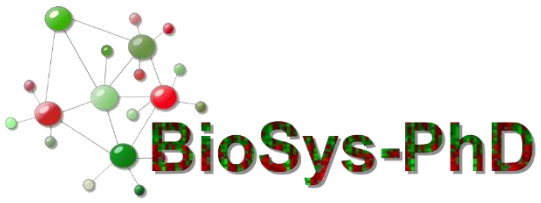
- Doutor Paulo Henrique Carrasquinho de Matos, Investigador Auxiliar do Instituto Nacional de Saúde Doutor Ricardo Jorge;
- Doutora Paula Duque, Investigadora Principal do Instituto Gulbenkian de Ciência da Fundação Calouste Gulbenkian;
- Doutora Ana Luísa Ribeiro da Silva, Técnica Superior Doutorada do Departamento de Endocrinologia Diabetes e Metabolismo do Hospital Santa Maria, na qualidade de individualidade de reconhecida competência na área científica;
- Doutora Luísa Maria Ferreira Romão Loison, Professora Associada Convidada da Faculdade de Ciências da Universidade de Lisboa (Orientadora).

Documento especialmente elaborado para a obtenção do grau de doutor

Bolsa de doutoramento financiada pela Fundação para a Ciência e a Tecnologia
(SFRH/BD/106081/2015)

2019

Joana Filipa Pires Silva foi bolseira de doutoramento no âmbito do Programa doutoral BioSys em Sistemas Biológicos, Genómica Funcional & Integrativa (FCT/PD/00065/2012) da Faculdade de Ciências da Universidade de Lisboa. Este projecto foi financiado pela Fundação para a Ciência e Tecnologia do Ministério da Ciência, Tecnologia e Ensino Superior com a bolsa de doutoramento SFRH/BD/106081/2015 e pelo BioISI – Biosystems & Integrative Sciences Institute da Faculdade de Ciências da Universidade de Lisboa (UID/MULTI/04046/2013).



De acordo com o disposto no artigo 24º do Regulamento de Estudos de Pós-Graduação da Universidade de Lisboa, Despacho nº 7024/2017, publicado no Diário da República – 2ª Série – nº 155 – 11 de Agosto de 2017, foram utilizados nesta dissertação secções incluídas no seguinte artigo:

Silva J, Fernandes R, Romão L (2019) Translational regulation by upstream open reading frames and human diseases. In: Romão L. (eds) The mRNA metabolism in human disease. Advances in Experimental Medicine and Biology, vol 1157, pages 99-116. Springer, Cham.

No cumprimento do disposto da referida deliberação, o autor esclarece serem da sua responsabilidade, exceto quando referido o contrário, a execução das experiências que permitiram a elaboração dos resultados apresentados, assim como da interpretação e discussão dos mesmos. Os resultados obtidos por outros autores foram incluídos com a autorização dos mesmos para facilitar a compreensão dos trabalhos e estão assinalados nas respetivas figuras e metodologias.

Index

Agradecimentos

Em primeiro lugar, dirijo os mesmos agredecimentos à direção da Faculdade de Ciências da Universidade de Lisboa, instituição que confere o grau.

Em segundo lugar, dirijo-me à direção do programa doutoral BioSys em Sistemas Biológicos, Genómica Funcional & Integrativa, integrado no BioISI – Biosystems & Integrative Sciences Institute da Faculdade de Ciências da Universidade de Lisboa, agradecendo pelo desenvolvimento deste programa doutoral e pelo financiamento durante os quatro anos de projeto.

Em terceiro, os meus agradecimentos são para a direção do Instituto Nacional de Saúde Dr. Ricardo Jorge (INSA), instituição de acolhimento que permitiu e co-financiou o desenvolvimento deste projeto de doutoramento.

À minha orientadora, a Doutora Luísa Romão do Departamento de Genética do INSA, agradeço por me considerar uma aluna com conhecimento científico para o desenvolvimento deste projeto de doutoramento, bem como pela ajuda científica para a resolução de problemas que permitiu o enriquecimento do meu pensamento científico.

Ao meu co-orientador, Doutor Augusto Luchessi do Depertamento de Biotecnologia da Faculdade de Ciências Aplicadas da Universidade Estadual de Campinas, agradeço pela coloboração e disponibilidade durante estes quatro anos, participando com questões científicas para a realização deste projeto.

À Doutora Margarida Gama-Carvalho e Doutora Ana Luísa Silva, agradeço por todas as 'discussões' científicas que contribuíram para à construção deste projeto de doutoramento. Ainda à Doutora Margarida Gama-Carvalho, agradeço a disponibilização dos equipamentos necessários à realização dos imensos ensaios de luminometria.

A todos que fazem parte do Depertamento de Genética do INSA, obrigada por de uma forma ou de outra terem ajudado à execução deste projeto. Um especial obrigada ao José Manuel Furtado, por toda a ajuda nos momentos de crise e por ter sempre uma palavra reconfortante e animadora.

Aos meus colegas de doutoramento da 2ª edição do programa doutoral BioSys, com especial ênfase à Andreia Henriques, que é mais que uma colega e virou uma excelente amiga, minha primeira colega de casa, que partilhou comigo momentos bons e menos bons desta jornada académica bem como da vida pessoal, um muito obigada.

Ao ex-colega de casa, colega da 3ª edição do programa doutoral e colega de laboratório (quase de bancada), o meu querido amigo Rafael Fernandes por ser, em muitas ocasiões, a voz

da minha consciência, partilhando comigo as alegrias e os dissabores deste projeto de doutoramento, e por ouvir os meus desabafos infinitas vezes. Por tudo isto e muito mais, um muito obrigada.

Aos meus colegas de laboratório, permanentes e passageiros, Rafaela, Juliane, Rafael, Cláudia, Paulo, Patrícia, Ana Rita, Andreia, Zé Ferrão, Juan, Bruna, Manuel, Cláudia Estima, Gonçalo, Pedro Nina, a todos vós um muito obrigada por me ouvirem e aturarem, mas também por partilharem comigo todos os meus momentos de loucura. Foi um começo atribulado, mas no final concluo esta etapa sabendo que ganhei amigos verdadeiros.

Aos homónimos de nome, Rafaela e Rafael, e à Juliane, um especial obrigada pelo tempo disponibilizado para a conclusão deste trabalho, tanto escrito como prático.

Às minhas lindas onconetes, agradeço por todo o apoio quer no trabalho como num contexto mais pessoal. Adorei conhecer-vos a todas e que continuemos a partilhar mais e bons momentos.

Aos meus amigos mais antigos e família, agradeço toda a paciência que tiveram comigo durante estes quatro anos por não poder estar sempre presente e nem sempre o mais bem-humorada possível. Ao meu padrinho, um especial agradecimento, pois deste sempre me fez ver que tudo o que fazemos deve ser aproveitado e não suportado. Um carinho especial ao meu afilhado e ao meu sobrinho que têm sido duas luzinhas de amor incondicional que me fazem ver a vida com outra cor.

Aos novos amigos que apareceram de súbito e que sem se aperceberem tornaram estes últimos meses mais fácies e mais bonitos, muito obrigada.

Aos meus pais e irmão, os mais sacrificados durante todo este processo de enrequecimento académico, de coração o meu muito obrigada! São o meu porto de abrigo, as pessoas em quem posso confiar sem qualquer dúvida, que me enxugam as lágrimas e me dizem “Vamos à luta que vamos conseguir!”, e que sei que me amam incondicionalmente. Amo-vos a todos!

Um especial Muito Obrigada, por tudo e desde sempre, à minha mãe. A esta pessoa linda, de coração gigante, que nunca desistiu de mim e sempre me fez ver que para a frente é que é o caminho. A ti que muitas lágrimas me viste derramar para chegar até aqui, e que tens sofrido comigo, digo-te que tens sido a minha inspiração, a minha força. Podemos agora chorar de alegria que chegámos ao fim desta etapa. Amo-te muito!

Abbreviations

5’UTR	5’-untranslated region
5-FU	5-Fluorouracil
5MP	eIF5-mimic protein
aa	amino acid
ABC	ATP-binding cassette
ABCE1	ATP-binding cassette subfamily E member 1
AdoMetDC	S-adenosylmethionine decarboxylase
AS	argininosuccinate synthase
A-site	acceptor-site
ATF4	activating transcription factor 4
ATP	adenosine triphosphate
BCL-2	B-cell lymphoma
BiP	binding immunoglobulin protein
bp	base pair
BSA	bovine serum albumin
C/EBPα	CCAAT-enhancer-binding protein-α
C/EBPβ	CCAAT-enhancer-binding protein-β
Cdk1	cyclin-dependent kinase 1
CDKN1B	cyclin dependent kinase inhibitor 1B
CDKN2A	cyclin dependent kinase inhibitor 2A
cDNA	complementary DNA
CHOP	CCAAT-enhancer-binding protein homologous protein
C-terminal	carboxyl-terminal
CTNNB1	catenin beta 1
Dbp5	DEAD-box RNA helicase
DMEM	Dulbecco’s Modified Eagle Medium
DMSO	dimethyl sulfoxide
DRB	adenosine analogue 5,6-dichloro-1-3-D-ribofuranosylbenzimidazole
DRD	DOPA responsive dystonia
eEF	eukaryotic elongation factor
eIF	eukaryotic initiation factor
eIF2α-P	phosphorylated eIF2α

EPHB1	ephrin receptor B1
EPRS	glutamyl-propyl-tRNA synthase
ER	endoplasmic reticulum
ERBB2	Erb-B2 receptor tyrosine kinase 2
ERCC5	excision repair cross-complementation group 5
eRF	eukaryotic release factor
E-site	exit-site
FBS	fetal bovine serum
FLuc	*Firefly* luciferase
FXII	factor XII
GADD34	growth arrest and DNA damage-inducible 34
GADD45A	growth arrest and DNA damage-inducible 45 alpha
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
GCH1	guanosine triphosphate cyclohydrolase 1
GCN2	general control non-derepressible-2
GCN4	general control protein
GDP	guanosine 5’-diphosphate
GTP	guanosine 5’-triphosphate
hCMV	human cytomegalovirus
HER-2	human epidermal growth factor receptor 2
HIF1α	hypoxia inducible factor 1 subunit alpha
HIV-1	human immunodeficiency virus type 1
HJV	hemojuvelin
HLA	human leukocyte antigen
HLH	helix-loop-helix
HRH2	histidine receptor H2
HRI	heme-regulated inhibitor
IBTKα	inhibitor of Bruton’s tyrosine kinase isoform α
IFN	interferon
IFRD1	interferon related developmental regulator 1
IRES	internal ribosome entry sites
IRF6	interferon regulatory factor 6
ISR	integrated stress response

JAK2	janus kinase 2
Leu	leucine
LUC	luciferase
m^7GTP	7-methylguanosine
MAP2K6	mitogen-activated protein kinase kinase 6
MEN	multiple endocrine neoplasia
Met	methionine
MKKS	McKusick-Kaufman syndrome
mORF	main open reading frame
mRNA	messenger RNA
MS	mass spectrometry
MSH5	MutS homolog 5
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
NBD	nucleotide-binding domain
NMD	nonsense-mediated mRNA decay
nt	nucleotide
NTC	non-template control
N-terminal	amino-terminal
o/n	overnight
ORF	open reading frame
PABPC1	poly(A) binding protein cytoplasmic 1
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
pDNA	plasmid DNA
PERK	PKR-like ER kinase
PIC	pre-initiation complex
PKR	protein kinase double-stranded RNA-dependent
PLB	passive lysis buffer
PP1c	protein phosphatase 1
PPP1R15b	protein phosphatase 1 regulatory subunit 15B
PPS	popliteal pterygium syndrome
P-site	peptidyl-site
PTC	premature termination codon

PTEN	phosphatase and tensin homolog
PVDF	polyvinyldene difluoride
Rac1	Rac family small GTPase 1
RiboSeq	ribosome profiling
RLI	RNaseL inhibitor
RLuc	*Renilla* luciferase
RNAse L	endoribonuclease L
RPMI	Roswell Park Memorial Institute 1640
rRNA	ribosomal RNA
RT-PCR	reverse transcription-PCR
RT-qPCR	reverse transcription-quantitative PCR
SCS	Saethre-Chotzen syndrome
SDS	sodium dodecyl sulphate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
SEM	standard error of the mean
SERCA	sarcoplasmic/ER Ca^{2+} ATPases
siRNA	small interfering RNA
SNP	single nucleotide polymorphism
SNV	single nucleotide variant
SOEing	synthesis by overlap extension
SPINK1	serine protease inhibitor kazal-type 1
SRY	sex determining region Y
TBS	tris-buffered saline
Tg	thapsigargin
Tip60	Tat interactive protein 60 kDa
TM	transmembrane
TP53	tumor protein p53
tRNA	transfer RNA
tRNAi	initiator tRNA
$tRNA_i^{Met}$	methionyl-initiator tRNA
TWIST1	twist-related protein 1
uAUG	upstream AUG
uORF	upstream open reading frame

UPF1	up-frameshift 1 regulator of nonsense transcripts yeast homolog
UPR	unfolded protein response
UV	ultraviolet
VWS	Van der Woude syndrome

Abstract

Short upstream open reading frames (uORFs) are *cis*-acting elements located within the 5'-leader sequence of transcripts. Recent genome-wide ribosome profiling (RiboSeq) studies have demonstrated the widespread presence of uORFs in the transcriptome and have shown that many uORFs can initiate with non-AUG codons. uORFs can impact gene expression of the downstream main open reading frame (mORF) by triggering messenger RNA (mRNA) decay or by regulating translation. Thus, disruption, elimination or creation of uORFs can elicit the development of several genetic diseases, such as cancer.

The ATP-binding cassette subfamily E member 1 (*ABCE1*) gene belongs to the ABC gene transporter superfamily. However, it does not behave as a drug transporter like the other members of this family. ABCE1 actively participates in the different stages of the translation process and is involved in cell proliferation and anti-apoptotic signaling processes, associating ABCE1 to a potential oncogenic function. RiboSeq occupancy profiles of the *ABCE1* mRNA 5'-leader sequence indicate an active translation associated with the presence of uORFs, which is suggestive of a high translational regulation.

Our aim was to study the translational regulation mediated by the five AUG and five non-AUG uORFs present in the human *ABCE1* 5'-leader sequence in colorectal cancer. With this purpose, we constructed a set of *Firefly* luciferase (FLuc) reporter vectors derived from the wild-type one containing the native configuration of the human *ABCE1* 5'-leader sequence upstream of the FLuc ORF, and transiently transfected colorectal cancer HCT116 cells. Here we show that *ABCE1* mORF expression is regulated by its uORFs. Our results are consistent with a model wherein uORF1 recruits ribosomes onto the mRNA, behaving like a ribosomal barrier. The ribosomes that efficiently bypass uORF1 and/or uORF2, must probably reinitiate at uORF3 and/or uORF5, while uORF4 is greatly bypassed. uORF3 and uORF5 function as repressive uORFs that may cooperate to reach a maximum repression of the mORF. Thus, both bypass and reinitiation events of the AUG uORFs within *ABCE1* 5'-leader sequence contribute for the translational control of the mORF. In constrast, the non-AUG uORFs seem to be devoid of a significant inhibitory activity. The AUG uORF-mediated translational control is maintained in normal and in endoplasmic reticulum (ER) stress conditions, which keeps the expression level of ABCE1 at a minimum, showing that *ABCE1* is a stress-resistant transcript whose functions are equally essential in normal and in coditions of global translation impairment. In addition, we show that *ABCE1* uORF-mediated translational regulation is preserved in non-tumorigenic and cancerous cells, which is consistent with a lack of an

oncogenic function by the uORFs, as well as ABCE1 itself, in the colorectal cancer cell line tested.

This study contributes with an additional example of how uORF-mediated translational regulation can occur. In addition, it reveals how important is to screen the 5'-leader sequence of the transcripts in search for potential disease-related variants. This information might be relevant for the implementation of new diagnostic and/or therapeutic tools for diseases associated with the deregulation of uORF-mediated translational control.

Key words

Translational regulation · uORFs · non-AUG uORFs · ABCE1 · colorectal cancer

Resumo

A expressão génica é extremamente regulada ao nível da tradução do RNA mensageiro (do inglês, *messenger RNA*, mRNA), sendo que por sua vez, este processo é dividido em diferentes fases: iniciação, alongamento, terminação e reciclagem dos ribossomas. A etapa da iniciação é vista como o passo limitante do processo de tradução, e é por isso altamente regulada pela existência de vários elementos regulatórios localizados na região 5'-não traduzida dos transcritos. De entre estes elementos que atuam numa configuração em *cis*, salientamos as pequenas grelhas de leitura a montante (do inglês, *upstream open reading frames*, uORFs). Estes elementos reguladores são potencialmente traduzidos e são definidos por um codão de iniciação na região 5'-não traduzida do mRNA e um codão de terminação localizado a montante ou sobreposto com a grelha de leitura principal (do inglês, *main open reading frame*, mORF). Estudos rescentes do perfil de ribossomas (do inglês, *ribosome profiling*, RiboSeq) monstraram a presença generalizada de uORFs no transcritoma, revelando ainda que uma grande maioria das uORFs inicia a sua sequência com codões não-canónicos diferindo em um nucléotido do codão de iniciação comummente usado, o AUG. É de facto estimado que aproximadamente metade dos transcritos têm pelo menos uma uORF. Estes elementos surgem em determinadas classes de transcritos, como é o caso dos fatores de transcrição, recetores celulares, oncogenes e genes envolvidos no crescimento e diferenciação celular.

As uORFs regulam a expressão génica da sua mORF ao desencadearem a degradação do mRNA ou por regularem a tradução. Em condições fisiológicas normais, estes pequenos elementos regulatórios inibem a tradução da respetiva mORF, obtendo-se uma repressão da expressão de 30 a 80%. A inibição da expressão da mORF depende em grande parte de um contexto de Kozak forte envolvendo o codão de iniciação da uORF, bem como de outras caraterísticas, como por exemplo, uma longa distância entre a extremidade 5' do transcrito e o início da sequência da uORF, a presença de múltiplas uORFs, uma longa uORF e/ou uma pequena distância entre o codão de terminação da uORF e o início da mORF. A repressão da tradução induzida por uma uORF pode ser devida a um bloqueio dos ribossomas aquando da tradução da uORF ou por dissociação e reciclagem dos mesmos após terminação da tradução da uORF. Em condições de stress, as uORFs podem ser vistas como elementos que permitem a tradução da mORF. Isto pode acontecer através de mecanismos de reiniciação da tradução após tradução da uORF ou por não reconhecimento do codão de iniciação da uORF (do inglês, *leaky scanning* ou *ribosomal bypass*), levando a que os ribosomas iniciem a tradução no codão de

iniciação da mORF. Desta forma, compreende-se que a disrupção, eliminação ou criação de uORFs pode levar ao desenvolvimento de várias doenças genéticas, como é o caso do cancro.

O gene *ABCE1* (do inglês, *ATP-binding cassette subfamily E member 1*) pertence à superfamília de genes transportadores ABC. Contudo não se comporta como um transportador de drogas como os outros membros desta família pelo facto de não apresentar domínios transmembranares. A proteína ABCE1 participa ativamente nas diferentes fases do processo de tradução, bem como, nos processos de proliferação celular e evasão da morte celular por apoptose, sugerindo que esta proteína pode ter funções oncogénicas. A ocupação ribosomal da sequência 5'-não traduzida do mRNA do gene *ABCE1* obtida por estudos de RiboSeq, é indicativa de uma tradução ativa nesta região que por sua vez parece estar associada com a presença de uORFs, as quais poderão estar envolvidas no controlo traducional da expressão da mORF. Para além disso, dados de RiboSeq numa linha celular de cancro coloretal, HCT116, mostraram a existência de pequenas ORFs que iniciam com codões de iniciação não-canónicos na região 5'-não traduzida do transcrito *ABCE1*. Paralelamente, foram também identificadas bioinfomaticamente uORFs iniciadas em AUG para este transcrito. Contudo, não foi realizado nenhum estudo experimental para investigar a função biológica das uORFs no mRNA do *ABCE1* humano.

Tendo em conta a falta de informação relativamente às uORFs presentes na região 5'-não traduzida do mRNA do *ABCE1*, o objetivo deste projeto foi o estudo da função biológica destas dez uORFs, tanto iniciadas em AUGs como em codões não-canónicos, em células do cancro coloretal. Para tal estabeleceram-se as seguintes tarefas: (i) determinar o impacto destas uORFs na expressão da respetiva mORF; (ii) identificar os mecanismos pelos quais estas uORFs regulam a tradução da mORF; (iii) determinar o impacto de condições de stress, nomeadamente condições de stress do retículo endoplasmático (do inglês, *endoplasmic reticulum*, ER), na regulação da tradução mediada por estas uORFs; e (iv) verificar se a regulação da tradução promovida por estas uORFs tem um papel no desenvolvimento tumoral do cancro coloretal. Para tal, a sequência de DNA complementar (do inglês, *complementary DNA*, cDNA) da região 5'-não traduzida do mRNA do *ABCE1* humano foi clonada a montante da ORF da luciferase do pirilampo (do inglês, *Firefly luciferase*, FLuc), num vetor plasmídico, obtendo-se assim a contrução ABCE1_5'UTR. Foi realizada mutagénese dirigida neste vetor de forma a obter todas os construções necessárias. Cada uma destas construções foi transfetada transitóriamente na linha celular HCT116, simultaneamente com um vetor que expressa a luciferase da *Renilla* como controlo. Após lise celular, os lisados proteicos foram analisados através de ensaios de luminometria. Os nossos resultados mostram que as uORFs presentes na região 5'-não

traduzida do mRNA do *ABCE1* humano têm impacto na tradução da respetiva mORF, como demonstrado pela redução de 70% da atividade relativa da luciferase após transfeção com a contrução ABCE1_5'UTR. Esta repressão da mORF do mRNA do *ABCE1* é promovida principalmente por duas uORFs iniciadas em AUG, a uORF3 e a uORF5, as quais parecem ter um efeito aditivo de forma a alcançar um máximo de repressão da mORF. As restantes uORFs iniciadas em AUG do transcrito *ABCE1* (uORF1, uORF2 e uORF4) não apresentam capacidade inibitória significativa. Embora todos os AUGs das cinco uORFs tenham o potencial de serem reconhecidas pela maquinaria de tradução, destas uORFs, a uORF1 e uORF5 são eficientemente reconhecidas devido aos seus codões de iniciação terem uma sequência de Kozak mais forte. Após a sua tradução, ocorre reiniciação da tradução no codão de iniciação seguinte que se encontre nas condições mais adequadas à tradução. Contrariamene, a uORF2, uORF3 e uORF4 são menos eficientemente reconhecidas. No conjunto, as cinco uORFs iniciadas em AUG regulam a tradução da sua mORF através de mecanismos de reiniciação da tradução e *leaky scanning*. Em contraste, as uORFs iniciadas em codões não-canónicos não apresentam uma capacidade inibitória significativa. A regulação da tradução do mRNA do *ABCE1* promovida pelas uORFs iniciadas em AUG foi mantida quer em condições normais ou de stress do ER, indicando que a expressão da proteína ABCE1 deve ser mantida num determinado nível para exercer as suas funções na célula mesmo em condições de inibição global da tradução. O padrão de regulação por estas uORFs é igualmente mantido em células do cancro coloretal bem como em células não tumorais do cólon (linha celular NCM460), mostrando que as mesmas não têm impacto no processo tumorigénico, o que por sua vez é concordante com um papel não-tumorigénico da proteína ABCE1 na linha de cancro coloretal testada.

Este estudo constitui um exemplo adicional sobre os mecanismos subjacentes à regulação da tradução do mRNA mediada por uORF(s). Além disso, revela o quão importante é a análise da região 5'-não traduzida dos transcritos de modo a identificar potenciais variantes associadas a doenças genéticas. Esta informação poderá ser relevante para a implementação de novas ferramentas de diagnóstico e/ou terapêuticas para doenças associadas à desregulação do controlo traducional mediado por uORF(s).

Palavras-chave

Regulação da tradução · uORFs · uORFs contendo codões de iniciação não-canónicos · ABCE1 · cancro coloretal

List of Figures

List of Tables

1. Introduction

1.1. General overview of the mRNA translation process

Gene expression is largely modulated at the level of mRNA translation, which is itself divided into initiation, elongation, termination and ribosome recycling steps [1–4]. The canonical process of translation initiation involves mRNA cap recognition and ribosomal scanning, and thus it is described as a scanning mechanism (Figure 1.1) [reviewed in 4]. The first step is the formation of an active ternary complex that consists of eukaryotic initiation factor 2 (eIF2) binding to the methionyl-initiator tRNA ($tRNA_i^{Met}$) and guanosine 5'-triphosphate (GTP), which is controlled by the guanine nucleotide exchange factor eIF2B that recycles and converts eIF2 bound to guanine 5'-diphosphate (eIF2-GDP) to eIF2-GTP [6, 7]. This is followed by the formation of the 43S pre-initiation complex (PIC) composed of eIF3, eIF1, eIF1A, eIF5, the eIF2·GTP·$tRNA_i^{Met}$ ternary complex, and the small 40S ribosomal subunit, that will attach to the 7-methylguanosine (m^7GTP) cap structure at the 5'-end of the mRNA (5'-cap), with the formation of the 48S initiation complex [4, 5, 8]. This process is mediated by the eIF4F complex composed by eIF4E, eIF4G and eIF4A, where eIF4E establishes a link with the mRNA 5'-cap, eIF4G recruits PIC by interaction with eIF3, and eIF4A unwinds the mRNA secondary structure to allow ribosomal scanning [4, 5, 9]. The 48S complex scans the 5'-leader sequence from 5' to 3' until it reaches an initiation codon (usually an AUG) at the peptidyl-site (P-site) of the ribosome. At this point, eIF5 mediates hydrolysis of eIF2-GTP, which promotes $tRNA_i^{Met}$ anticodon base-pairing with the start codon, the release of eIF2-GDP along with other eIFs, and the binding of the 60S large ribosomal subunit by eIF5B to form the 80S translating ribosome [4, 5, 8, 10, 11]. One of the factors that can influence AUG recognition is its surrounding context, defined as the Kozak consensus sequence [12, 13]. The optimal sequence surrounding the AUG initiation codon was postulated to be (GCC)A/GCC**AUG**G, where the positions -3 and +4 relative to the A of the AUG, are the most important ones [13]. The G purines at positions -3 and +4 establish a link with the α subunit of eIF2 (eIF2α) and the 18S ribosomal RNA (rRNA), respectively [14]. Additionally, the nucleotide (nt) G at the position -6, as well as C at positions -1 and -2, seem to improve translation initiation efficiency [13, 15].

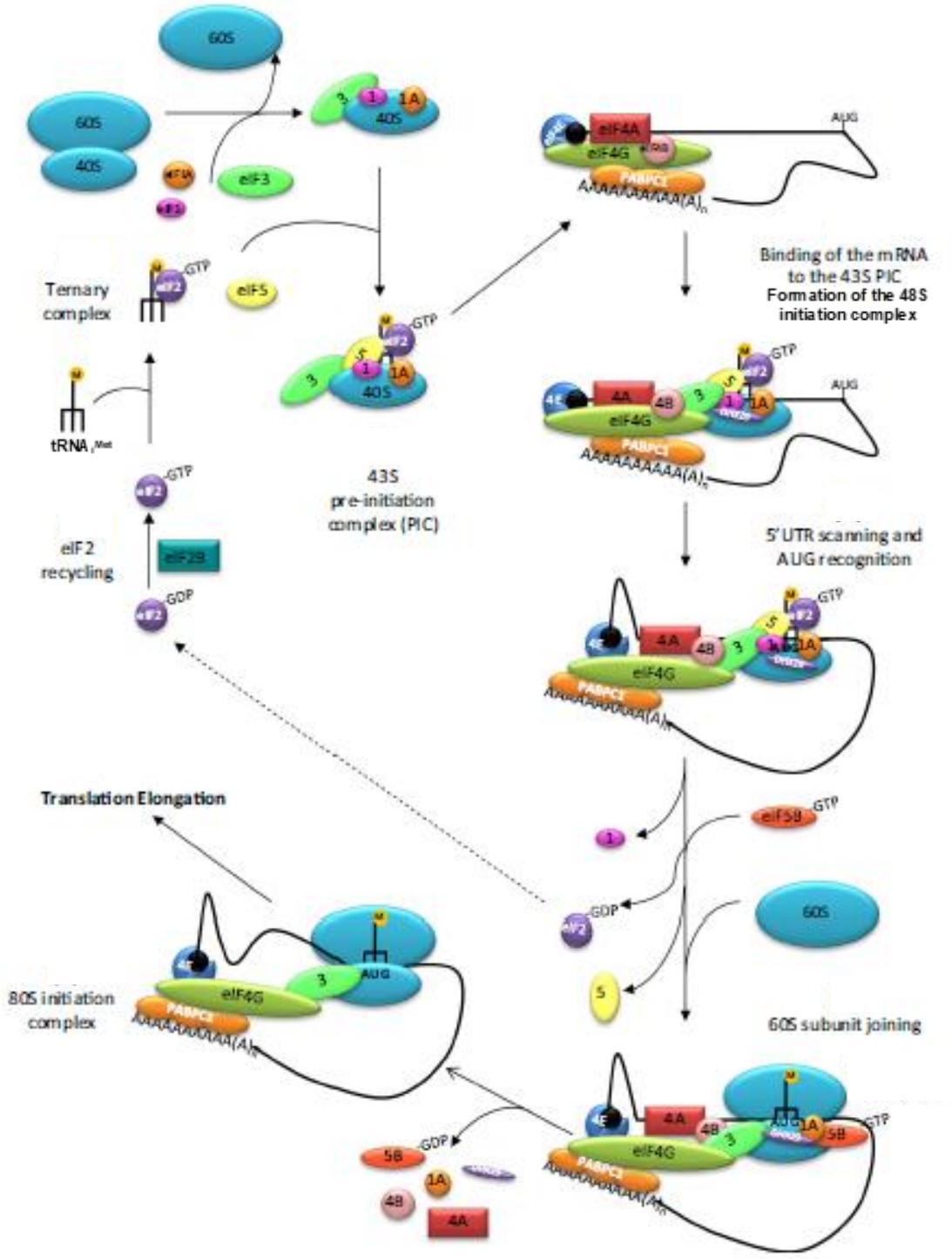


Figure 1.1 – Eukaryotic translation initiation process. The translation initiation process begins with the assembling of the ternary complex composed of eIF2, GTP and $tRNA_i^{Met}$, after eIF2B-mediated recycling of eIF2-GDP to eIF2-GTP. This is followed by the formation of 43S PIC (eIF3, eIF1, eIF1A, eIF5, the eIF2·GTP·$tRNA_i^{Met}$ ternary complex, and the small 40S ribosomal subunit) that binds to the 5′-cap structure by the eIF4F complex composed by eIF4E, eIF4G and eIF4A, forming the 48S initiation complex. The 48S complex scans the 5'-leader sequence from 5' to 3' until it reaches an initiation codon (usually an AUG) at the P-site of the ribosome. eIF5 mediates hydrolysis of eIF2-GTP, with consequent $tRNA_i^{Met}$ anticodon base-pairing with the start codon, the release of eIF2-GDP along with other eIFs, and the binding of the 60S large ribosomal subunit by eIF5B to form the 80S translating ribosome (Adapted from Lacerda and co-workers [31]).

Introduction

The elongation phase of translation begins after tRNAi anticodon base-pairing to the AUG in the ribosomal P-site, when the ribosome moves forward to the second codon of the transcript open reading frame (ORF) being in the ribosomal acceptor-site (A-site) [16]. The eukaryotic elongation factor eEF1A, binds to a cognate aminoacyl-tRNA in a GTP-dependent manner, delivering it to the A-site of the small ribosomal subunit of the elongating ribosome [16, 17]. tRNA recognition of the second codon promotes GTP hydrolysis by eEF1A, followed by its release and recycle by the exchange factor eEF1B in the form of eEF1A-GTP, allowing for anticodon base-pairing [16, 17]. Peptide bound formation between the incoming aminoacyl moiety of the A-site tRNA with the P-site peptidyl-tRNA occurs rapidly [16, 18]. This is followed by a ratchet-like motion of the ribosomal subunits that promotes tRNAs movement to a so-called hybrid P/E and A/P states where the tRNA acceptor ends are in the exit (E)- and P-sites and the anticodon loops are still in the P- and A-sites, respectively. Binding of eEF2-GTP to the small 40S subunit, promote translocation of the tRNAs into the P- and E-sites. The hydrolysis of GTP promotes conformational changes in eEF2 that unlock the ribosome, which will allow for tRNA and mRNA movement with consequent locking the subunits in a post-translocation state. In this post-translocation state, a deacylated tRNA occupies de E-site and a peptidyl-tRNA occupies the P-site. The A-site of the ribosome is now free for the binding of the next aminoacyl-tRNA by eEF1A [16].

The termination step of translation is not so well described as is the initiation phase, but some new evidences support a model where the ATP-binding cassette subfamily E member 1 (ABCE1) has a central role, being considered as the first termination factor that binds to the ribosome [19]. The stop codon (either UAA, UAG and UGA) context for an efficient recognition and efficient translation termination requires a purine residue in position +4, and in the case of a pyrimidine at position +4, it also needs a purine at position +5 [20]. As soon as the A-site is unoccupied, ABCE1 binds to the ribosome [19]. ABCE1 and its interactor eukaryotic release factor 3 (eRF3) bound to GDP, bind to the 40S ribosomal subunit at the ribosome termination complexes, and wait for the entry of eRF1 delivered by DEAD-box RNA helicase (Dbp5) to the stop codon at the A-site of the ribosome after its recognition by eRF1 [19, 21]. Only after Dbp5 dissociation by ATP-hydrolysis, both release factors interact with each other. The quick interaction between eRF1 and eRF3 initiates with GTP binding to eRF3 (with consequent release of eRF3 from ABCE1), followed by GTP hydrolysis, which triggers the dissociation of these two factors, resulting in the activation of eRF1 [19]. Dissociation of eRF3–GDP by the j subunit of eIF3 (eIF3j) allows for peptidyl-tRNA hydrolysis and subsequent eRF1-ABCE1 interaction that positions eRF1 at the peptidyl-transferase center of

the ribosome with the following release of the formed polypeptide [19, 22–24].

After translation termination, ABCE1 splits the 80S ribosome into its large 60S and small 40S subunits allowing for ribosome recycling [25, 26]. Post-termination ribosomal complexes are only efficiently dissociated by ABCE1 after peptide release by both eRF1 and eRF3 that are maintained associated to the ribosome [19, 26]. Binding of ABCE1 to the post-terminating complexes slightly stimulates its NTPase activity needed for rybosomal recycling [26]. In the pre-splitting complex, binding of adenosine triphosphate (ATP) to the ABCE1 nucleotide-binding domain (NBD) active sites promote their closure, positioning its Fe-S cluster towards the A-site where it interacts with eRF1 that will further increase the NTPase activity of ABCE1 [4, 5]. Thus, hydrolysis of ATP allows for the ABCE1 NDBs to open and the Fe-S clusters repositioning at the 40S subunit as in the pre-splitting complexes [25–27]. This ATP-dependent tweezer-like motion of the ABCE1 NBDs creates a power stroke that promotes ribosomal splitting [26]. The presence of eIF3/eIF1/eIF1A and eIF3j in the 40S ribosomal subunit in association with ABCE1 empowered the dissociation of the 60S and 40S ribosomal subunits (with bound mRNA) [26]. Here, eIF1 promotes P-site tRNA release and eIF3j allows for mRNA dissociation [26]. Ligatin, also known as eIF2D, enhances deacylated tRNA and mRNA release from the recycled 40S ribosomal subunit [28]. After ribosomal subunit dissociation, ABCE1 remains stably bound to the 40S subunit and in this 40S-ABCE1 post-splitting complex, other factors remain also bound, as the case of the initiation factors eIF1A, eIF2, and subunits of eIF3 [25, 27]. Additionally, Heuer and co-workers identified tRNAi at the ribosomal P-site [25]. The presence of initiation factors in the 40S-ABCE1 post-splitting complexes shows that ABCE1 is actually participating in the formation of new translation initiation complexes [25]. In fact, ABCE1 seems to function as an anti-association factor by inhibiting the formation of the 80S active ribosome until the recognition of a new AUG codon by the tRNAi still bound to the 40S subunit [29].

1.2. Upstream open reading frames act as translational regulators

Translation initiation is the rate-limiting step of translation and is tightly controlled by mechanisms that involve different regulatory elements located in the 5'-leader sequence of the transcripts [2]. Among them are: (i) internal ribosome entry sites (IRES), which are highly structured RNA regions that recruit ribosomes to or near to the translation initiation codon and thus induce translation in a cap-independent manner; (ii) 5'-untranslated region RNA structures (hairpins, G-quadruplexes and pseudoknots) that impair start codon scanning by the ribosome and repress translation initiation; (iii) protein binding sites where different molecular ligands

can interact or form stable ribonucleoprotein complexes, thus promoting or repressing translation; (iv) RNA modifications that unfold RNA and are usually associated with an efficient translation initiation process; and (v) upstream AUG or non-AUG open reading frames (uORFs) that usually inhibit the downstream translation initiation at the main open reading frame (mORF) [2, 4, 30, 31]. A uORF consists of a well-studied class of small ORFs potentially translated, with its initiation codon within the 5'-leader sequence of a mRNA and its in-frame termination codon upstream or overlapped with the mORF, and ranging typically from two to one hundred codons [2, 32–34]. It is bioinformatically estimated that approximately 49 to 58% of human transcripts carry at least one uORF [33, 35]. Indeed, uORFs are conspicuous in certain classes of genes, such as transcription factors, cellular receptors, oncogenes and genes involved in cell growth and differentiation control [35–39]. These regulatory elements are usually seen to be evolutionarily conserved, which may suggest an important biological function [33, 40].

With the advent of ribosome profiling (RiboSeq), a highly precise technique for monitoring *in vivo* translation based on RNA deep-sequencing of mRNA fragments covered with ribosomes, a significant ribosomal occupancy (signature of active translation) was shown in regions thought to be non-coding, as the case of the 5'-leader sequences, which is consistent with the widespread presence of translatable uORFs among the transcriptome [33, 34, 41]. It was recently shown that, among the uORFs identified by RiboSeq, uORFs containing near-cognate codons are more frequent compared to AUG-containing uORFs, and their recognition involves alternative translation initiation mechanisms. The most prevalent non-AUG codon is CUG [34, 42]. Non-AUG uORFs are important for the regulation of protein synthesis of specific transcripts, including several with a relevant role in stress responses, and also oncogenes [34, 42–45].

Under physiological conditions, uORFs are typically described as repressors of translation initiation at the downstream mORF [2, 33, 46]. This is explained by the process of start codon recognition during translation initiation: the 43S PIC binds to the 5′-cap structure of the mRNA and forms the 48S initiation complex that scans from 5′ to 3′ until it reaches an initiation codon [2, 4, 35]. Then, eIF2 ternary complex and other initiation factors dissociate, which reduces the levels of mORF expression in about 30 to 80% [33, 40, 47]. The inhibitory activity of a uORF can be associated with a strong upstream initiation codon context. Other features may also positively influence the repression activity of a uORF, such as a long distance from the 5'-end of the transcript to the beginning of the uORF, the existence of multiple uORFs (additive effect), the consumption of active pre-initiation complexes, a long uORF and/or a short intercistronic distance (distance from the uORF stop codon to the mORF start codon) [33, 35, 46, 48, 49].

Translational repression can be achieved by ribosome dissociation and recycling after uORF translation or by ribosome stalling of the elongating/terminating ribosomes during the process of uORF translation (Figure 1.2A) [1, 2]. While ribosome dissociation can be regulated either by the nucleotide sequence or the resulting uORF-encoded peptide [50, 51], ribosome stalling is associated with unique features in the 5'-leader sequence, such as the presence of secondary structures, the interaction with *trans*-acting factors, the nascent peptide sequence, or codon usage (rare versus common codons) [46, 48, 52–54]. Ribosomal stalling at the uORF termination codon can also trigger nonsense-mediated mRNA decay (NMD), since the uORF stop codon can be recognized as a premature termination codon (PTC) (Figure 1.2A) [55–57]. As an example, an inhibitory overlapping uORF in human and yeast *STN1* mRNA was recently described. In yeast, this overlapping uORF targets *STN1* to NMD, maintaining the low abundance of STN1 responsible for the normal activity of telomeres [58].

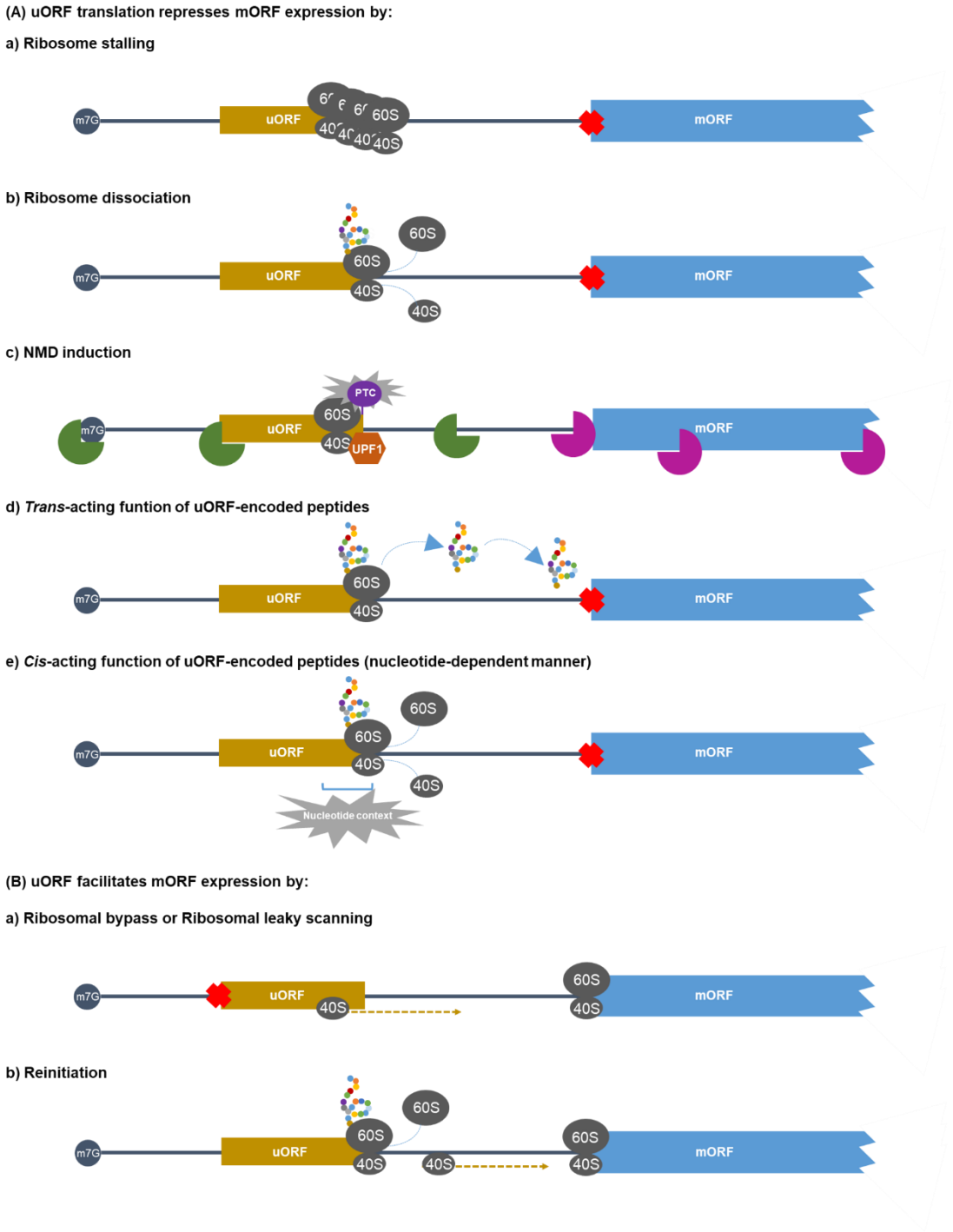


Figure 1.2 – Mechanisms of uORF-mediated translational regulation. (A) uORF translation mediates mORF repression by different mechanisms: a) ribosome stalling, in which the elongating/terminating ribosomes are blocked usually due to the presence of secondary structures, the nucleotide context, and the interaction with *trans*-acting factors; b) ribosome dissociation with consequent ribosome recycling that is regulated mainly in a nucleotide-dependent manner; c) NMD induction, due to stalling of the ribosomes at the uORF termination codon that is recognized as a PTC; d) *trans*- and e) *cis*-inhibition of mORF expression by the uORF-encoded peptide, being the last one dependent on the peptide sequence and its interaction with the translational machinery, leading to ribosome stalling and dissociation. **(B)** uORF-mediated mORF translation by: a) ribossomal bypass or ribosomal leaky scanning, where the scanning ribosomes pass through the uORF initiation codon without recognizing it, initiating translation further downstream, a mechanism mainly associated with a poor Kozak context; and b) reinitiation, where the uORF is translated and the 40S ribosomal subunit remains attached, resumes scanning and reinitiates translation at the mORF, a mechanism associated with small uORFs and a long intercistronic distance.

However, in response to cellular stress, the presence of uORFs can increase the expression of certain mRNAs, a mechanism used by stress-responsive transcripts to alleviate the cell from stress, and also by oncogenes during the tumor initiating process [1, 42]. During a stress stimulus (oxidative and endoplasmic reticulum (ER) stress, hypoxia, nutrient deprivation, ultraviolet (UV) radiation, among others) the cell tends to respond rapidly by reprogramming its gene expression pattern at the level of protein synthesis [39, 59]. Depending on the stress, four different serine-threonine kinases can be activated, phosphorylating eIF2: PKR-like ER kinase, PERK; protein kinase double-stranded RNA-dependent, PKR; general control non-derepressible-2, GCN2; and heme-regulated inhibitor, HRI [59]. Phosphorylation of eIF2 at serine 51 on its α subunit (eIF2α-P) impairs GDP to GTP exchange by guanine nucleotide exchange factor eIF2B, necessary to obtain a new active form of eIF2α. Therefore, the ternary complex is not formed, which impairs translation initiation [60, 61]. The abundance of eIF2α-P determines global and gene-specific translation rates [51]. This gene expression reprogramming in conditions of global translation repression, aims to restore cell homeostasis and cell survival by activation of the integrated stress response (ISR). If the stress is too long or severe, cells can be committed to programmed cell death (apoptosis) [59, 62, 63]. Therefore, stressed cells repress global translation to avoid unnecessary protein synthesis, but at the same time, allow translation of specific proteins aiming to solve the stress [1, 39]. Many of these proteins are encoded by uORFs-containing transcripts [4, 39, 51, 56]. However, the existence of a uORF is not a straight indication that the corresponding transcripts will be translated in stress conditions [51]. There are several factors that can influence the translational regulatory function of a uORF: (i) its length; (ii) the intercistronic distance; (iii) the distance from the 5'-end of the 5'-leader sequence to the uORF start codon; (iv) its start codon context; (v) its secondary structure; (vi) number of uORFs; (vii) the type of initiation codon, and (viii) its stop codon context [2, 33, 50]. These features can operate alone or in combination to define the uORF-mediated mechanism of translational control orchestrated for a specific mRNA under physiological or stress conditions [1, 51].

1.2.1. Translation reinitiation

As mentioned above, if a scanning ribosome recognizes a uORF initiation codon, it can translate the uORF and dissociate or stall during either the elongation or termination phase of translation. However, another option exists for the ribosome: it can translate the uORF and remain associated with the mRNA, continue scanning, and reinitiate translation further downstream (Figure 1.2B). The mechanism of translation reinitiation implies that after uORF

translation termination, the ribosomal 60S subunit is released from the mRNA while the 40S subunit remains attached, and resumes scanning until a new competent ribosome is assembled for translation initiation at a downstream initiation codon [1, 3, 5]. One major factor influencing reinitiation is the availability of the ternary complex, more precisely the eIF2-GTP in this complex, which is hydrolyzed and released after uORF start codon recognition and ribosomal assembly. The abundance of the ternary complex can be linked to the distance scanned by the 40S ribosomal subunit until it reaches a new initiation codon [64]. In this case, short uORFs and long intercistronic distances are positively related to an efficient translation reinitiation at a downstream ORF [1, 38, 65]. On the other hand, a decrease in the reinitiation rate, for instance, by longer uORFs and/or the presence of secondary structures that promote ribosome stalling, can be explained by the loss of several eIFs along with the ternary complex after uORF translation is completed [3, 48, 54]. In fact, some of the eIFs are transiently maintained associated with elongating and terminating ribosomes, which in turn are the ones that will resume scanning and reinitiate translation downstream [66]. It seems that if the 40S-eIF3-eIF4F complex that was responsible for the engagement of the 40S ribosome to the uORF initiation codon is kept together until termination of uORF translation, it will allow the 40S subunit to resume scanning and initiate translation at a downstream initiation codon. Disruption of this complex before uORF translation is completed, destabilizes the interaction between the ribosome and the mRNA, leading to its dissociation after uORF translation termination, which in the end prevents reinitiation at the downstream mORF [54].

The mammalian activating transcription factor 4 (*ATF4*) and its yeast homolog, the general control protein (*GCN4*), are well described examples of transcripts translated via a mechanism of 'delayed' reinitiation (Figure 1.3A). Both are transcription factors that, when induced, activate several pathways of the ISR aiming to release the cell from stress [64, 67].

ATF4 5'-leader sequence bears two uORFs: a 3 amino acid (aa)-long first uORF (uORF1) that is 5'-proximal, and a lengthened, out-of-frame and coding sequence-overlapping uORF2 of 59 aa, both with an AUG in a strong Kozak consensus context. During basal conditions, the short uORF1 is translated, allowing the 40S ribosomal subunit to be maintained attached to the mRNA, which resumes scanning until it acquires a new ternary complex (eIF2·GTP·tRNAi^{Met}) in time to reinitiate translation at the uORF2 start codon. This last event leads to a translational repression of the mORF, since the ribosome terminates uORF2 translation at 3' of the mORF initiation codon [67].

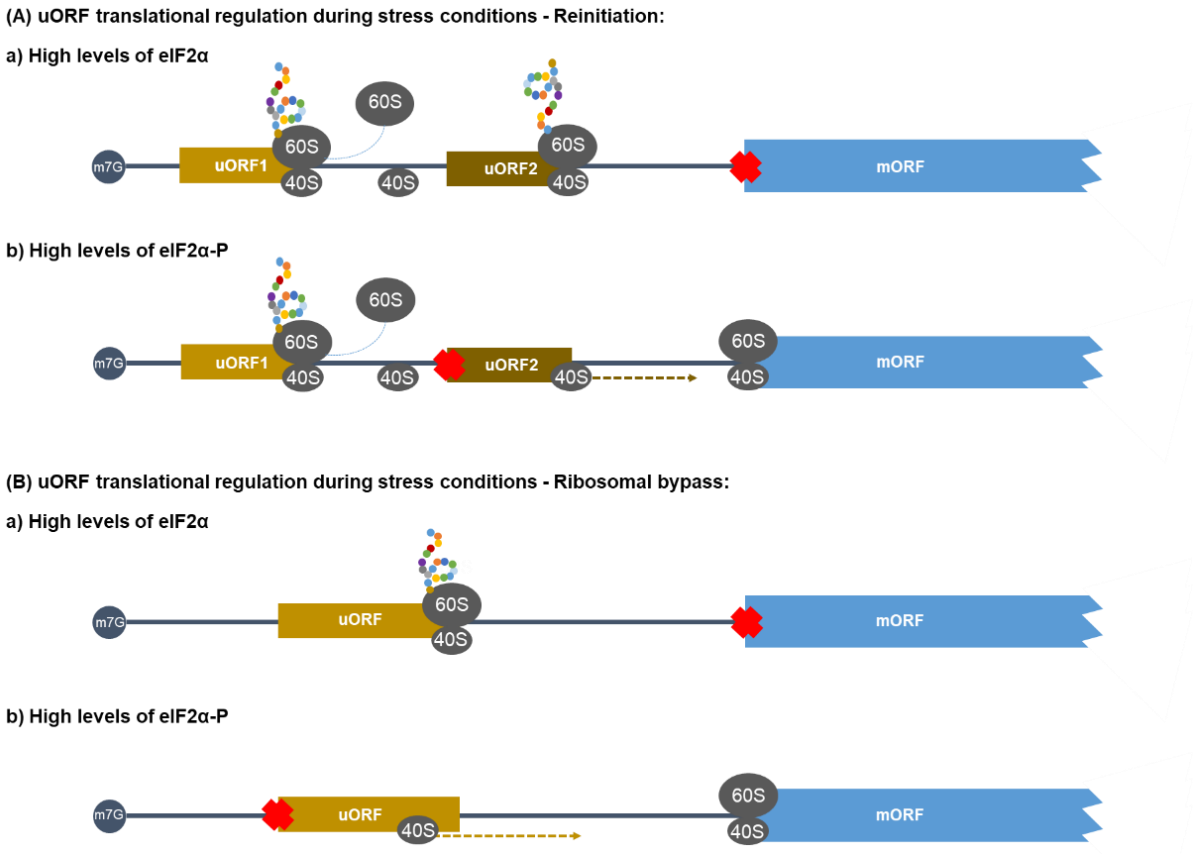


Figure 1.3 – Mechanisms of uORF-mediated translational regulation in stress conditions. During a stress stimulus (for example, ER stress), eIF2α is phosphorylated (eIF2α-P), reducing the formation of new ternary complexes (eIF2·GTP·$tRNA_i^{Met}$) necessary for translation initiation, thus impairing global translation, but at the same time, allowing the translation of stress-responsive transcripts. **(A)** Reinitiation as a mechanism to facilitate mORF expression of specific transcripts during stress: a) in basal conditions, where eIF2α is abundant, the first uORF start codon is recognized and the uORF translated. After uORF1 translation termination, the 40S ribossomal subunit remains bound to the mRNA and continues scanning until it acquires a new functional ternary complex in time to reinitiate translation at the second uORF start codon, rendering mORF downregulation; b) during stress, uORF1 is translated but reinitiation at uORF2 start codon is delayed due to the high levels of eIF2α-P that impairs the formation of new ternary complexes, thus uORF2 is bypassed and the ribosome reinitiates at the mORF initation codon. **(B)** Ribosomal bypass as a mechanism to facilitate mORF expression of specific transcripts during stress: a) in physiological conditions (high levels of eIF2α), translation of the inhibitory uORF impairs mORF expression; b) during stress, the weak uORF Kozak context together with the low levels of eIF2α (high levels of eIF2α-P) promote uORF bypass by the scanning ribosome that will be further loaded with active ternary complexes in time to initiate translation at the mORF initiation codon.

During stress, phosphorylation of eIF2α reduces the cellular levels of eIF2-GTP, which is a limiting step for the formation of more ternary complexes [1, 64]. In these conditions, after *ATF4* uORF1 translation, there is a delay in the reload of a new ternary complex to the 40S ribosomal subunit, important to initiate translation at the uORF2 initiation codon. Therefore, the ribosome bypasses the inhibitory uORF2 and reinitiates translation at *ATF4* main start codon [67]. As exemplified by the inhibitory role of *ATF4* uORF2, overlapped uORFs are obvious translational repressors since their termination codons are located 3' of the mORF initiation codon [1, 3]. However, it is speculated that the post-terminating ribosomes can scan

backwards in a 3' to 5' fashion at least for a critical number of nucleotides (less than ten) and reinitiate translation at an upstream initiation codon, although less efficiently [48].

In the case of *GCN4*, its 5'-leader sequence contains four short uORFs and from them only uORF1 is translated in both unstressed (good nutrition) and stressed (starvation) conditions. In fact, abolishing translation of the uORF1 by mutating its initiation codon, significantly impairs the expression of *GCN4* mORF [64]. In good nutritional conditions, reinitiation occurs in the following uORF-initiation codons after uORF1 translation, since the terminating ribosomes are rapidly reloaded with the necessary factors to initiate translation, which therefore repress GCN4 expression [36, 64]. From the three downstream uORFs, uORF3 and uORF4 are the most repressive ones [64]. During stress, as for *ATF4* mRNA, the terminating ribosomes can no longer reload the required initiation factors in time to translate the inhibitory uORFs, resulting in the translation of the mORF [36, 64]. Interestingly, it is postulated that reinitiation in yeast diverges from mammalian reinitiation by the existence of specific *cis*-acting elements located upstream or downstream of the uORF [66]. The last codon and the 10 nt AU-rich sequence located at the 3'-side of the uORF1 termination codon promote the continued binding of the ribosome to the mRNA to allow permissive reinitiation. This is explained by the lack of a strong base-pairing between the rRNA and the AU-rich sequence that, in contrast to the GC-rich uORF4 termination codon context, will not allow the terminating ribosome to dissociate, and therefore the ribosome will resume scanning [50]. Additionally, permissive reinitiation is also associated with an upstream sequence that interacts with the initiation factor eIF3, mainly the eIF3a subunit via its amino-terminal (N-terminal) domain, which stabilizes the post-terminating 40S ribosome subunit in the mRNA, promoting reinitiation at the *GCN4* main initiation codon [68–70].

Another relevant aspect of the uORF-mediated reinitiation process is the ability to select translation initiation codons responsible for the synthesis of different protein isoforms encoded from the same mRNA molecule [1]. This is illustrated by the transcription factors CCAAT-enhancer-binding protein-α (*C/EBPα*) and -β (*C/EBPβ*), which regulate the proliferation and differentiation of multiple cell types, that have four alternative initiation sites: one for an out-of-frame uORF and three others that encode functionally different isoforms (extended, p42 and p30 for *C/EBPα*, and LAP*, LAP and LIP for *C/EBPβ*) [36, 71]. The presence of a translatable uORF between the first and third initiation codons of both mRNAs regulates the expression of each isoform, being responsible for the translation reinitiation of the truncated isoforms of *C/EBPα* and *C/EBPβ*, respectively, p30 and LIP [36, 71, 72].

1.2.2. Ribosomal bypass or leaky scanning

Ribosomal bypass or ribosomal leaky scanning is a mechanism in which the scanning ribosomes bypass the uORF start codon and initiate translation at the subsequent start codon (Figure 1.2B) [1–3]. Ribosomal bypass is usually associated with the proximity of the upstream AUG (uAUG) to the 5'-end and, to a greater extent, to the surrounding context of the uORF initiation codon [1, 3, 73, 74]. In fact, during ER stress, a weak uORF initiation codon context, together with a stronger initiation codon at the mORF, are seen in mRNAs that are preferentially translated (Figure 1.3B) [51, 75, 76]. However, it is unclear if this mechanism is triggered by a reduced availability of the ternary complex or an alteration in other factors critical for the translational process [3]. For instance, in the case of *C/EBPα* and *C/EBPβ* transcripts, the uORF initiation codon is bypassed in conditions of reduced levels of eIF4E and eIF2α (high abundance of eIF2α-P) resulting in the expression of the full-length isoform in detriment of the truncated ones [71]. Another example of a transcript that is regulated by a bypass mechanism is the growth arrest and DNA damage-inducible 34 (*GADD34)*. GADD34 is an important protein that regulates the ISR by a negative feedback mechanism. GADD34 controls the magnitude and duration of gene expression reprogramming by controlling the levels of eIF2α-P via interaction with the catalytic subunit of protein phosphatase 1 (PP1c), thus leading to dephosphorylation of eIF2α and ISR attenuation [51, 60]. Both *GADD34* uORFs are bypassed during stress conditions due to their weak Kozak context, thus increasing GADD34 expression [51, 62]. The α isoform of inhibitor of Bruton's tyrosine kinase (*IBTKα*) mRNA is also translationally regulated by a mechanism of ribosomal bypass during conditions of eIF2α phosphorylation. In the case of the transcript *IBTKα* that encodes a protein important for the adaption and resolution of stress by promoting cell survival, it has a highly conserved 588 nt-long 5'-leader sequence bearing four uORFs. The *IBTKα* 5'-leader sequence cloned upstream of the *Firefly* luciferase reporter construct, when transfected into mouse embryonic fibroblast cells treated with thapsigargin, showed an increase in the luciferase activity, which is associated with a translational deregulation of the inhibitory uORFs. By mutational analysis of the uORFs' initiation codons, it was shown that uORF1 is the main inhibitory uORF of IBTKα expression. The weak *IBTKα* uORF1 initiation codon context, together with low levels of eIF2α during stress, allow its bypass and thus, mORF expression [75].

Interestingly, some of the transcripts containing NMD-inducing uORFs in physiological settings, develop NMD-resistance under stress conditions [56, 77]. This is achieved by ribosomal bypass, where the ribosome scans through the uORF initiation codon, not recognizing the uORF termination codon as a PTC, and therefore saving the transcript from

NMD [2]. This seems to be the case of the transcript 1 of interferon related developmental regulator 1 (*IFRD1*) that has a 52 aa-long uORF that, when translated in physiological conditions, represses *IFRD1* expression by triggering mRNA decay [77]. In conditions of ER stress induced by tunicamycin, *IFRD1* mRNA stabilizes, as there is ribosomal bypass of its uORF, leading to increased protein levels [77]. Other stress-related mRNAs, like *ATF4* [55, 56], as well as CCAAT-enhancer-binding protein homologous protein (*CHOP*) [56, 65], are also well described examples of uORF-containing transcripts that commit them to NMD. These mRNAs are upregulated in stress conditions, where eIF2α-P causes uORF ribosomal bypass and consequent NMD impairment, thus enhancing the ISR [56]. For instance, *CHOP* 5'-leader sequence contains a uORF with two in-frame uAUGs with inhibitory activity when in physiological conditions [65, 76]. In thapsigargin-induced ER stress, a translational increase of CHOP is reported due to leaky scanning of the uORF initiation codons that are seen to be in a weak Kozak context [65, 76, 78]. Persistent elevated levels of CHOP during a prolonged ISR, where the cells can no longer resolve the stress and survive, will induce the apoptotic signaling cascade by transcriptional activation of pro-apoptotic genes [79].

1.2.3. Recognition of non-canonical initiation codons

A large group of the uORFs identified by RiboSeq is seen to initiate at near-cognate start codons [34, 42]. However, the real number of non-canonical codons acting as initiator sites of translation is possibly underestimated, since initiation at these codons shows some resistance to the recognition by translation initiation-inhibitor drugs and also because there are no reliable bioinformatic tools for their discrimination [12, 34, 80]. The CUG codon, encoding for a leucine (Leu), is the most prevalent non-AUG codon used in uORFs, followed by GUG (valine) that is present in uORFs at roughly the same extent as AUG, and then by UUG (Leu) [12, 34, 42, 43]. Consistent with these data are the results obtained by a peptidomics study, where small peptides synthesized by non-AUG uORFs were detected [81].

Although translation initiation is poorly efficient when started at non-canonical initiation codons compared to the AUG, uORFs containing near-cognate start codons seem to be equally translated and important for the regulation of protein synthesis of specific transcripts during stress, where global translation is repressed [34, 42–45]. uORFs bearing non-AUG codons appear to regulate the expression of their downstream ORFs by a mechanism of ribosomal bypass [3]. For instance, the glutamyl-propyl-tRNA synthase (*EPRS*) mRNA is highly expressed in conditions of eIF2α-P abundance induced by the ER stressor thapsigargin, due to the regulation of two inhibitory uORFs with non-canonical start codons. The first uORF

initiates with a CUG and is overlapped and out-of-frame with the mORF and the second one is a UUG-containing uORF. Mutation of the CUG codon at uORF1 to an AUG in an optimal Kozak context represses the downstream ORF in physiological conditions and during thapsigargin-induced stress. Thus, bypass of the uORF1, due to its non-canonical start codon, is necessary for the translation of the *EPRS* mORF. The same holds for uORF2 that when its initiation codon is mutated to AUG, it obtains low downstream expression during normal and stress conditions. Thus, as occurs for uORF1, the presence of the UUG initiation codon in uORF2 facilitates its bypass by some scanning ribosomes that will initiate translation at the mORF, a mechanism exacerbated under stress conditions due to high levels of eIF2α-P [82]. The growth arrest and DNA-damage inducible gamma (*GADD45G*) transcript, with a relevant function in cell growth and apoptosis regulation, also bears a CUG-containing uORF that, due to its overlapping and out-of-frame context, represses GADD45G expression in unstressed conditions. During starvation, the inhibitory uORF is bypassed essentially due to the non-canonical nature of its initiation codon, which increases GADD45G protein levels [83]. One additional example of this mechanism is observed for the WNT signaling pathway regulator *FRAT2* transcript. A construct carrying the bicistronic *FRAT2* mRNA sequence, with the uORF and the mORF tagged differentially, shows co-expression of the uORF-encoded peptide and the main protein in HEK293T cells. The uORF translation is initiated at an ACG codon that when mutated to AUG abolishes the expression of its mORF with concomitant expression of the uORF. This agrees with a bypass of the *FRAT2* uORF to allow downstream translation [81].

The conspicuous translation initiation at non-AUG codons raises the question of how these non-canonical codons are recognized as initiation sites by the translation machinery. It was seen that a CUG initiation codon is translated via a cap-dependent mechanism, although translation initiation seems to be independent of the $tRNA_i^{Met}$ [80, 84, 85]. For instance, using methionine sulfamide, an inhibitor of the methionyl-tRNA synthase (responsible for the aminoacylation of the tRNA with methionine, Met), in toeprinting assays, does not produce a significant effect on the CUG codon recognition. This result suggests the 'replacement' of the $tRNA_i^{Met}$ by another aminoacyl-tRNA in the ribosomal P-site of the pre-initiation complexes for the translation of CUG codons [80]. In fact, previous studies revealed that CUG is translated as a Leu and not as a Met residue [84]. Furthermore, it is suggested that some of the 40S ribosomal subunits are not pre-loaded with any initiator tRNA, contrasting with the current model of translation, and that they are able to load the $tRNA^{Leu}$ when CUG is encountered. Delivery of $tRNA^{Leu}$ to the P site of the ribosome is also independent of eIF2·GTP·$tRNA_i^{Met}$ and is mediated by the eukaryotic initiation factor 2A (eIF2A) [85]. In conditions of low abundance of

eIF2·GTP·$tRNA_i^{Met}$, the levels of eIF2A are found to be elevated [86]. Conversely to eIF2α, depletion of eIF2A does not promote measurable global translational repression, but has a negative effect on uORFs containing CUG or UUG as initiation codons, which supports the function of eIF2A in translation initiation of non-canonical start codons [85, 86]. This is shown for the binding immunoglobulin protein (*BiP*) transcript, where depletion of eIF2A impairs translation of its UUG-containing uORF. During ER stress induced by thapsigargin, this depletion leads to a significant decrease in BiP expression, due to the downregulation of the UUG-uORF [86]. In parallel to eIF2A, it is shown that other factors are involved in non-canonical translational events. Controlling the rate of non-AUG initiation is the equilibrium between the translation initiation factor eIF5 and the eIF5-mimic protein (5MP), acting as an enhancer and a repressor, respectively, of translation initiation at near-cognate codons [43, 87]. The repression exerted by 5MP, is suggested to be achieved by preventing the interaction of eIF5 with eIF2 in PIC in favor of its own interaction [43]. Moreover, start codon selection is also the result of a cross-regulation between eIF5 and eIF1, where overexpression of eIF1 represses non-AUG initiation [87]. Yet, the interplay between eIF5 and eIF2A in the promotion of translation initiation at non-canonical start codons was not determined [43].

Similar to what is postulated for the AUG, recognition of the non-AUG codons is associated with a strong Kozak consensus sequence. The optimal sequence context for translation initiation at a CUG is almost the same already described for AUG, being TCCACC**CUG**G, and as expected, modifications of this sequence will weaken the recognition of the CUG [84]. In the case of *FRAT2* mRNA, when the ACG start codon of the uORF is placed in a less optimal Kozak sequence, the uORF-encoded peptide is expressed at lower levels [81]. As for CUG, efficient translation at the GUG codon requires the presence of a CCACC sequence immediately upstream [88]. Additionally, it was also demonstrated that non-AUG start codons enhance their own translation by blocking the scanning ribosomes in their vicinity. This is explained by a highly GC-rich content immediately after the non-AUG codon allowing the formation of hairpins that will impair scanning [44].

As pointed out before, it seems that there is a pool of ribosomes that scan specifically for the non-AUG codons, and that those ribosomes may be different from the ones initiating at the AUG [84]. In fact, ribosomal heterogeneity can be used as an explanation of the differential recognition of AUG and non-AUG codons [80]. This seems to be dependent on the rRNA composition, on ribosomal proteins that can be differentially expressed and post-translationally modified, and on the modification of the translation factors interacting with the ribosome, as well as the tRNAs [80, 89, 90]. These so-called specialized ribosomes can modulate translation

of mRNAs depending on the presence of specific features such as uORFs, thus leading to the preferential translation of classes of mRNAs, adding another layer of translational regulation centered on the ribosome [89, 91]. Additionally, even the constitutive components of the ribosome that have little or no variation, can have specialized activities when interacting with regulatory elements in the 5'-leader sequence [89]. Moreover, it is postulated that changes in the ribosomal components is crucial for gene expression reprogramming depending on the cell environment, differentiation and development [90].

1.2.4. Functional uORF-encoded peptides

The uORF-encoded peptides are usually poorly detected by the conventional proteomic methods based on mass spectrometry (MS) [92]. Moreover, despite the wide detection of uORF translation by RiboSeq approaches, this technique does not give information on the peptides encoded by these elements [34, 81, 93]. Thus, complementation between mRNA deep-sequencing and RiboSeq with proteomics approaches, such as MS, expand overall protein identification, including the detection of uORF-encoded peptides [81, 92–94]. These small peptides were encountered in cells at concentrations equivalent to the cellular proteins and also with specific subcellular localizations, providing evidence that they have biological functions, others than mORF repression [81].

The uORF-encoded peptides can act as *cis*- or *trans*-regulators of the mORF expression (Figure 1.2A) [3]. The ability of the peptides to stall the translation machinery in a *cis*-fashion can be explained by their specific sequence and/or their direct or indirect interaction with small molecules ('peptoswitch'), as occurs for S-adenosylmethionine decarboxylase (*AdoMetDC*) mRNA [32, 95]. The AdoMetDC uORF-encoded peptide has 6 aa with the sequence MAGDIS that seems to interact directly, via the 4^{th} and 5^{th} residues, with the last tRNA ($tRNA^{Ser}$), stabilizing the ribosome near the uORF termination codon in the presence of high levels of polyamines [53, 95]. This process stalls the ribosomes and allows their dissociation, which results in the translational repression of the mORF [95]. In the case of the already mentioned *GADD34* mRNA, its 5'-leader sequence contains two inhibitory uORFs, in which uORF2 translation significantly represses mORF expression during basal conditions in a nucleotide-dependent manner [51, 62]. The uORF2-encoded small peptide contains a conserved Pro-Pro-Gly sequence at the carboxyl-terminal (C-terminal) region that allows the release of the translating ribosomes, thus inhibiting mORF translation [51]. An example of the *trans*-acting regulation activity of the uORF-encoded peptides is also represented for the argininosuccinate synthase (*AS*) transcript [96]. Pendleton and co-workers showed that the peptide encoded by

the overlapped out-of-frame uORF of the *AS* transcript inhibits the expression of endogenous AS protein through a process dependent on the uORF length and sequence [96].

The uORF-encoded peptides can also interact with different cellular proteins, and not only with the products of their own coding sequences, functioning as *trans*-acting factors, and thus having several biological functions in the cell [32, 86]. For instance, the small peptides encoded by both non-AUG uORFs of the *BiP* transcript function as human leukocyte antigen (HLA)-presented epitopes recognized by human T cells [86]. The uORF-encoded peptides can also be developmentally and/or spatially regulated, and for that are endowed of distinct functions compared to the protein encoded by the mORF [81, 97, 98]. An example of this developmental regulation was already described for the spliced transcript variant of MYCN (*MYCNΔ1b*) where the presence of a uORF does not influence the translational level of MYCNΔ1b, but instead produces a small peptide (MYCNOT) that is expressed in fetal but not in adult brains [98]. Regarding spatial regulation, there is the McKusick-Kaufman syndrome (*MKKS*) transcript, where two of the three uORFs in its 5'-leader sequence seem to be translated into highly conserved peptides with mitochondrial localization, distinct from the cytoplasmic localization of the MKKS protein. This suggests that the uORF-encoded peptides and the main protein have distinct functions [97].

1.3. uORFs and human genetic disorders

The importance of uORFs in the regulation of different patterns of gene expression under normal and stress conditions highlights their relevance, if altered, for the development of several human diseases, such as metabolic, hematologic and neurologic disorders, inherited syndromes, cancer and its susceptibility [2, 36–39, 45]. In different disease-associated variant databases, more than 3700 variants were identified in the 5'-leader sequence of human transcripts that can alter a uORF [35]. Among these variations, the most harmful are the ones that originate or eliminate an initiation or termination codon, and thus regulate the presence or absence of a uORF [35, 36]. Additionally, other alterations could deregulate uORF features like the Kozak consensus context, the uORF length, the uORF number, and its distance to the 5' end of the mRNA or to the main coding sequence [36]. Several bioinformatic analyses were performed to map possible variations (polymorphisms and mutations) within the 5'-leader sequence of transcripts that can interfere with the regulatory function of uORFs [33, 35]. Calvo and co-workers identified 509 transcripts bearing common polymorphisms associated with the creation or elimination of a uORF [33]. More recently, Wethmar and co-workers identified in 2610 genes, 1375 single nucleotide polymorphisms (SNPs) disrupting uAUGs and 2724 SNPs

affecting the Kozak context of a uORF. Additionally, in a small percentage of genes, 697 SNPs are present at uORF stop codons. Eight uORF-disruptive SNPs already have clinical association [99]. A well-known example of a disruptive polymorphism is the one identified in the factor XII (*FXII*) gene that plays a role in the coagulation process [100]. The alteration of a C for a T at the position -4 (c.-4C>T) of the *FXII* 5'-leader sequence forms a novel uORF that negatively regulates the FXII plasma levels, associating this polymorphism to the occurrence of stroke episodes [33, 100, 101]. Calvo and co-workers have also identified eleven novel patient mutations in disease-associated genes that create or eliminate a uORF [33]. One example is the interferon regulatory factor 6 (*IRF6*) gene associated with Van der Woude syndrome (VWS) and the popliteal pterygium syndrome (PPS), two disorders characterized by facial alterations, such as cleft lip and palate. From the several mutations identified in the *IRF6* gene, a frameshift mutation in the position -48 of the 5'-leader sequence that alters a T to an A (c.-48T>A), creates a new initiation codon [102]. This mutation originates a uORF that significantly represses IRF6 protein levels (70-100% of inhibition), which is typically associated with the disease phenotype [33]. Additionally, there are mutations that create another uORF in a uORF-containing mRNA, such as the case of the sex determining region Y (*SRY*) gene in gonadal dysgenesis (c.-75G>A) and the serine protease inhibitor kazal-type 1 (*SPINK1*) gene in hereditary pancreatitis (c.-53C>T). This leads to the almost complete inhibition of mORF expression, which is explained by a cumulative effect of multiple inhibitory uORFs [33, 103, 104].

In the reviews of Barbosa and co-workers and Silva and co-workers are listed several examples of uORF creation, elimination or modification in the development of several types of diseases, including rare disorders [2, 105], such as hereditary thrombocythemia (uORF elimination) [106, 107], melanoma predisposition (uORF creation) [108, 109] and Marie Unna hereditary hair loss (uORF modification) [110]. Several new examples of genetic diseases have been described to have an association with uORFs' deregulation. For instance, haploinsufficiency of twist-related protein 1 (TWIST1) is correlated to the development of Saethre-Chotzen syndrome (SCS), characterized by a malformation of the skull (craniosynostosis). From a screening of 14 genetically undiagnosed SCS patients, two novel single nucleotide variants (SNVs) were detected in the *TWIST1* 5'-leader sequence (c.-263C>A and c.-255G>A) that contribute to the formation of novel start sites (AUG) in a good Kozak context. The c.-263C>A SNV generates an out-of-frame uORF of 68 codons and the c.-255G>A SNV generates an in-frame uAUG with the mORF that possibly forms an N-extended isoform of TWIST1 protein. Both alterations repress *TWIST1* mORF expression, which is associated with the typical disease phenotype, when no common mutations in the main coding

sequence are present [111]. Kitano and co-workers identified the presence of three SNPs - rs542483929, rs188349884 and rs759579732 - in the 5'-leader sequence of the histidine receptor H2 (*HRH2*) gene that create transposon-derived upstream ATGs, which originate uORFs. These new formed uORFs downregulate mORF expression and can be potentially associated with gastric cancer susceptibility, a disease already related to mutations in the enhancer region of *HRH2* [112].

Alterations in the pattern and function of uORFs in proto-oncogenes and tumor suppressors genes can explain how cancer is triggered: mutations that lead to a loss-of-function of the uORF in proto-oncogenes resulting in their overexpression or mutations that promote a gain-of-function of the uORF in tumor suppressor genes resulting in a decrease of these protective proteins [36]. RiboSeq analysis studies provide evidence of the widespread presence of exomic cancer mutations that alter uORF start codons impairing their regulatory role in several oncogenes and tumor-suppressor genes, such as *MYC*, B-cell lymphoma (*BCL-2*), phosphatase and tensin homolog (*PTEN*), tumor protein p53 (*TP53*), MutS homolog 5 (*MSH5*), among others [113].

As mentioned above, familial melanoma predisposition is a well-established uORF-related condition, associated with mutations in the cyclin dependent kinase inhibitor 2A (*CDKN2A*) gene [109]. *CDKN2A* impairs cell cycle progression by encoding two tumor suppressor proteins, $p16^{INK4A}$ and $p14^{ARF}$ [108, 114]. A transversion of the nucleotide G to T at position -34 (c.-34G>T) in the *CDKN2A* 5'-leader sequence creates an out-of-frame uAUG in the mRNA. This uAUG forms a translatable uORF responsible for the low expression levels of the mORF, thus providing predisposition to melanoma [109].

In addition to its role in disease development, uORF creation has also been related to poor prognosis in the response to drug treatment. The creation of a uORF in the mRNA of the excision repair cross-complementation group 5 (*ERCC5*) gene generated by a polymorphic variation (rs751402) confers resistance to platinum-based chemotherapeutics in childhood ependymoma (malignant brain tumor). This is observed by the ERCC5 up-regulation mediated by its uORF following cisplatin-induced bulky adduct DNA damage. The elevated levels of ERCC5 lead to a higher degree of cisplatin-induced DNA damage repair, thus lowering the efficiency of this chemotherapeutic [115].

An example of a genetic alteration that modifies a uORF was reported in cyclin dependent kinase inhibitor 1B (*CDKN1B*) gene, associated with the induction of inherited multiple endocrine neoplasia (MEN) syndrome, which is characterized by several distinct tumors affecting at least two endocrine organs [116]. *CDKN1B* gene encodes for $p27^{KIP1}$, a tumor

suppressor that regulates cell cycle and cell proliferation by promoting cell cycle arrest at G1-phase [114]. A 4-base pair (4bp) deletion (c.-456-453delCCTT) was identified in the sequence of the *CDKN1B* uORF, which disrupts and shifts its termination codon, resulting in a lengthened uORF sequence and a reduced intercistronic space in relation to the downstream initiation codon. This germline mutation allows the translation of a lengthened uORF-encoded peptide that reduces mORF expression by preventing reinitiation events to occur that in the end can be associated with cancer formation. The patient carrying this germline mutation has pituitary adenomas and tumors in the endocrine pancreas, consistent with the MEN4 phenotype. No other biological functions were associated with this uORF-encoded peptide that can be related to the disease phenotype [116]. A uORF-encoded peptide with an association with the disease phenotype was recently identified in the familial DOPA responsive dystonia (DRD). The c.-22C>T SNP in the 5'-leader sequence of the guanosine triphosphate cyclohydrolase 1 (*GCH1*) gene creates an uAUG that promotes the formation of an overlapped and out-of-frame uORF. This uORF encodes a 73 aa-long peptide responsible for the low levels of GCH1 protein, which impairs the dopamine biosynthesis pathway, resulting in reduced levels of dopamine and dopaminergic dysfunction in the brain, characteristic of DRD. This 73 aa peptide also accumulates at considerable levels within the nucleus where it is predicted to be involved in transcription factor activities, promoting cytotoxic effects with reduction of cell viability [117].

A systematic search for cancer-related uORF mutations has been performed, screening for loss-of-function uORF mutations in 404 uORF initiation sites of 132 potential proto-oncogenes in 308 human malignancies. Interestingly, mutations were identified in both the uAUG and the uORF Kozak consensus sequence. Four novel uORF-associated mutations caused the loss of an uAUG in the Src family tyrosine kinase BLK proto-oncogene (*BLK*) in a colon adenocarcinoma; the ephrin receptor B1 (*EPHB1*) in a mammary carcinoma; the janus kinase 2 (*JAK2*) in chronic lymphocytic leukemia; and the mitogen-activated protein kinase kinase 6 (*MAP2K6*) in a colon adenocarcinoma. There was no detectable function of the mutations found in *BLK* and *JAK2*. Although, in the cases of *EPHB1* and *MAP2K6*, the uAUGs were mutated to an uGUG and an uACG, respectively, which were responsible for the up-regulation of the downstream ORF. It is worth noting that the *MAP2K6* uORF mutation was also present in normal control tissue of the affected patient, which suggests a relationship between this variant and the predisposition to tumor development. In an additional study, a whole exome sequencing computational analysis of datasets of 464 colon adenocarcinomas revealed 53 non-recurrent somatic mutations that delete either the uORF initiation or termination codon [118]. This

highlights the importance of uORF mutations in the tumorigenic process, although further functional studies are needed.

In addition to the role of genetic alterations that deregulate uORF-mediated translational regulation, there are other mechanisms that can overcome or take advantage of the repression exerted by naturally occurring uORFs to promote a disease phenotype. For instance, MDM2 is an oncoprotein that antagonizes, in a feedback loop, the function of the tumor suppressor p53 and has been seen overexpressed in many tumors, such as osteosarcomas, gliomas and soft tissue sarcomas. MDM2 can be produced from two spliced isoforms with different 5'-leader sequences as a result of the function of two cryptic promoters: (i) the first promoter (P1) transcribes a long mRNA (L-mdm2) without exon 2, and (ii) the second promoter (P2) transcribes a short mRNA (S-mdm2) with exon 2, but lacking exon 1. L-mdm2 mRNAs contain two inhibitory uORFs in exon 1 that will repress MDM2 expression. However, those uORFs do not exist in the S-mdm2 mRNA. In choriocarcinoma cells, MDM2 expression is translated from the S-mdm2 transcript. Apparently, the transcriptional switch from promoter P1 to P2 is related to the binding of p53 to elements in intron 1 of MDM2 gene which activates P2 and consequently induces MDM2 expression via transcription of S-mdm2 [119]. Additionally, BRCA1, DNA repair associated gene, is partially regulated by a similar mechanism. *BRCA1* is well known to be the major susceptibility gene for breast and ovarian cancers by showing a reduced expression level. It was shown that two distinct promoters are used to produce two *BRCA1* transcript variants distinct in their 5'-leader sequences (named 5'UTRa and 5'UTRb). The mRNA containing the 5'UTRb is expressed in breast cancer but not in normal tissues, and it contains three uORFs that negatively modulate mORF expression [120]. Another alternative mechanism of uORF regulation with pathophysiological relevance was described for Erb-B2 receptor tyrosine kinase 2 (*ERBB2*) transcript, also known as the human epidermal growth factor receptor 2 (*HER-2*). HER-2 is an oncoprotein that appears overexpressed in breast cancer cells (33). *HER-2* mRNAs have an inhibitory small uORF in which the stop codon is located 5 nt upstream of the mORF initiation codon, impairing translation reinitiation and keeping HER-2 at basal levels under physiological conditions. However, another post-transcriptional mechanism seems to occur to promote overexpression of HER-2 in cancer cells, without having alterations in the mRNA sequence or size. A U-rich translational derepression element was identified in the 3'UTR of the *HER-2* mRNA, that associates with several *trans*-acting factors, among them RNA-binding proteins, to repress the inhibitory activity of the translatable uORF, allowing an efficient translation of the mORF in breast cancer cells [121].

Recently, RiboSeq data has shown that, during initiation of the tumorigenic process of epidermal cells, translation of cancer-related mRNAs is dependent on the translation of the uORFs present in their 5'-leader sequences. Moreover, those uORFs initiate to a great extent at non-canonical start codons, with the CUG being the most prevalent. High levels of eIF2A were also detected, and this, as mentioned before, has a role in the translational regulation of those non-AUG uORFs, and thus this alternative translation factor is also associated with tumor progression. Given this, the authors speculate that for tumorigenic initiation the translational apparatus needs to be redirected to the translation of uORFs in a cohort of cancer-related mRNAs such as catenin beta 1 (*CTNNB1*), hypoxia inducible factor 1 subunit alpha (*HIF1α*), Rac family small GTPase 1 (*Rac1*), cyclin-dependent kinase 1 (*Cdk1*), among others [42]. The above-specified examples well illustrate how deregulation of uORF-mediated translational regulation can be associated with human diseases.

1.4. The ABCE1 protein

The ATP-binding cassette (ABC) protein superfamily is the larger family of transporters [123]. A typical configuration of an ABC transporter contains two NBDs and two transmembrane (TM) domains: the NBDs provide the necessary energy for the transport and the TMs provide substrate specificity [25, 123]. ABC gene superfamily of transporters is divided into seven subfamilies based on similarity in gene structure, order of domains, and sequence homology in the cytoplasmic NBD and TM domains [123]. ABCE1 protein belongs to the ABC superfamily of proteins, being the only member of the subfamily E (ABCE) [123, 124]. The *ABCE1* gene is localized in the chromosome 4, more precisely at position 4q31, and its main coding sequence encodes a 599 aa protein (~67 kDa), found in the nucleus and in the cytoplasm of mammalian cells [125–127]. *ABCE1* gene is highly conserved and transversally found in Eukarya and Archea, but not in Bacteria [126, 128]. ABCE1 protein is composed of two NBDs (NBD1 and NBD2) and a hinge domain at the C-terminal of the protein that arranges the two NBDs in a head-to-tail orientation (Figure 1.4) [123, 129]. However, contrary to what is seen for this family of proteins, ABCE1 does not have the TM domains that confer to these proteins the ability to behave like drug transporters [123, 128]. Additionally, at the N-terminal of the ABCE1 protein there is an iron-sulfur (Fe-S) domain and in its core there are two $(4Fe\text{-}4S)^{2+}$ clusters (Figure 1.4), which confer to ABCE1 a wide range of biological roles in cellular processes; for instance, the link of ABCE1 protein to nucleic acids or to the endoribonuclease L (RNase L) [126, 129].

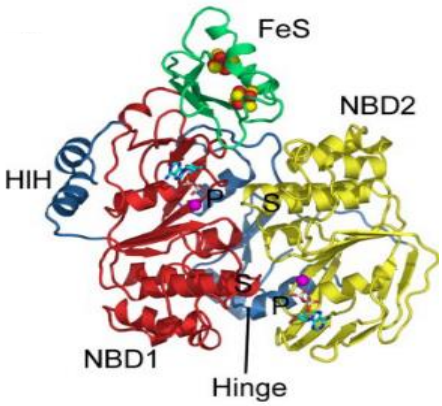


Figure 1.4 – The ABCE1 protein structure. The NBDs 1 and 2 are highlighted in red and yellow, respectively; the NDB1 contains a helix-loop-helix (HLH) motif showed in dark blue. Both NBDs, with the ATP binding site, are orientated in a head-to-tail conformation by the specific hinge domain (blue). Fe-S domain (green), with both $(4Fe\text{-}4S)^{2+}$ clusters (iron in red, sulfur in yellow), binds to NBD1 (Adapted from Karcher and co-workers [129]).

ABCE1 protein was first described as a 2-5A linked oligonucleotides (2-5A)-dependent RNaseL inhibitor (RLI; another common name assigned to ABCE1), by the formation of an heterodimer with RNase L, dependent on the ratio between the two proteins. Increasing level of RNase L by interferons (IFNs), changes the ratio between ABCE1 and RNase L, thus allowing the activation of the 2-5A/RNase L pathway [126]. The 2-5A/RNase L system is one of the effective pathways of the IFN-antiviral response controlling protein synthesis by promoting RNA degradation at early stages of viral infection [130, 131]. ABCE1 will inhibit the interaction of 2-5A oligonucleotides to RNase L and thus inhibits its 2-5A-dependent activation, therefore inhibiting the 2-5A/RNase L pathway. This associates ABCE1 with a potential role in this IFN system (although not regulated by it as occurs for RNase L) and in RNA metabolism of mammalian cells, by antagonizing the functions of RNase L [126]. During human immunodeficiency virus type 1 (HIV-1) viral infection, ABCE1 is upregulated, which leads to the inhibition of RNase L function, enhancing the pool of HIV RNA and proteins, with consequent viral production, in an human T lymphocyte cell line. Lower ABCE1 expression has the opposite effect: RNase L is no longer inhibited and the viral load decreases [132]. Additionally, ABCE1 is reported to be involved in a more direct way to the viral process of HIV-1, by participating in the assembly of viral capsids in mammalian cells. In HIV-1 infection, ABCE1 is up-regulated and associates with Gag polypeptides after their translation, acting as a molecular chaperone during the assembly of those HIV-Gag polypeptides at the plasma membrane, creating a complete immature capsid [133]. Moreover, ABCE1 is only recruited for the locations at the plasma membrane where there are assembling Gag polypeptides, and ABCE1-Gag association is maintained only until virus maturation (capsid formation), followed

its dissociation and consequent virus release [134]. Moreover, an association of ABCE1 with the RNA metabolism was established by Kärblane and co-workers by demonstrating that ABCE1 functions as an RNA silencing suppressor in *N. benthamiana* plants and in the human kidney cell line HEK293 [135].

Since the 2-5A/RNase L pathway only functions in vertebrates cells, it is seen that the inhibitory function of ABCE1 towards RNase L is only a secondary role for this protein [126, 135]. ABCE1 was also postulated to have relevant functions in the translation process. Its overall participation in the translation process seems to explain its evolutionary conservation [125]. It was reported by Chen and co-workers a crucial role of ABCE1 in the initial stage of translation, by its involvement in the assembly of the pre-initiation complex. This is postulated by the interaction of ABCE1 with the eukaryotic initiation factor eIF5 and eIF2α, demonstrated in the human cervival cancer HeLa cells by co-immunoprecipitation studies. Chen and co-workers also showed that mRNA translation is inhibited by small interfering RNA (siRNA)-depletion of *ABCE1*, as shown by a shift of occupancy by large polysomes to free monosomes [128]. Dong and co-workers showed that ABCE1 is associated with 43S and 48S pre-initiation complexes, by interacting with the small 40S subunit. Depletion of ABCE1 in yeast cells impairs the association of eIF2, eIF1 and eIF5 to the 40S subunit, thus reducing the 43S pre-initiation complexes, demonstrating the involvement of ABCE1 in binding of these factors to the small 40S ribosomal subunit and thus with the formation of 43S, and consequently, 48S complexes for translation initiation [136].

In addition to its function in translation initiation, ABCE1 has also been reported to play a role in translation termination and recycling, mainly by its close interaction with the release factors eRF1 and eRF3, as above described in section 1.1 [19, 26]. Additionally, ABCE1 by forming a complex with Pelota and Hbs1, two paralogues of eRF1 and eRF3, was shown to promote the dissociation of vacant 80S ribosomes and stalled elongation complexes [137]. Another function that makes ABCE1 intrinsically associated with translation is its involvement in ribosomal biogenesis. Kispal and co-workers demonstrated in yeast that depletion of ABCE1 abrogates the export of both 40S and 60S subunits from the nucleus to the cytoplasm. This function of the ABCE1 is associated with its Fe-S clusters at the N-terminal domain [138].

Chen and co-workers associated ABCE1 with cell proliferation control by demonstrating that when *ABCE1* is siRNA-depleted from HEK293 cells, there is a shift from large polysomes to monosomes which indicates a substantial impairment on protein synthesis with a consequent decrease in cell viability [128]. Toompuu and co-workers demonstrated that siRNA suppression of *ABCE1* in HEK293 cells and also in the HeLa cells, abrogates cell proliferation with the

accumulation of cells in the S-phase of the cell cycle [127]. Additionally, the delay in S-phase progression results only from *ABCE1* suppression and is completely independent of RNase L activity on the modulation of the expression of mRNAs involved in cell cycle control [127, 139]. Toompuu and colleagues also demonstrated that in *ABCE1* siRNA-depleted cells there is an insufficient DNA synthesis and histone expression, which can impair cell cycle progression and thus explains cell accumulation in S-phase [127]. In addition, in both studies of Chen and co-workers [128] and Toompuu and co-workers [127], there were morphological signs of cell death after *ABCE1* depletion, indicating that ABCE1 has an anti-apoptotic activity. Altogether, these results are in agreement with a potential role of ABCE1 in oncogenesis. In fact, the chromosomic localization (4q31) of *ABCE1* gene is a fragile site that was already associated with structural rearrangements in cancers [140].

ABCE1 overexpression has been associated to the tumorigenic process of several human cancers, for instance lung cancer [141–143], breast cancer [144], esophageal cancer [145], ovarian cancer [146], glioma [147], among others. *ABCE1* mRNA and protein expression were quantified in lung adenocarcinoma tissues, metastatic lymph nodes and also in normal lung tissues. By comparing the three groups, the *ABCE1* mRNA and protein levels were seen to increase in this order: normal lung tissues<lung adenocarcinoma tissues<metastatic lymph nodes. A decrease in cell proliferation was measured after transfection of the lung adenocarcinoma cell line A549 with a lentivirus expressing an *ABCE1*-specific short hairpin, showing the relevance of ABCE1 in the tumorigenesis of lung cancer. Ren and co-workers also demonstrated that after depleting *ABCE1* from A549 cells, there are 476 differentially expressed genes belonging to different classes, such as cell proliferation, development, cell adhesion and apoptosis, among others [141]. Despite these indications, the mechanism by which ABCE1 can lead to development of lung cancer is still unclear. Liang and co-workers have shown that the HIV-1 Tat interactive protein 60 kDa (Tip60), a lysine acetyltransferase, regulates the rate of acetylation of ABCE1 between the human normal bronchial epithelial HBE cell line and the human lung cancer cell line A549, with a significant acetylation rate in the A549 cells compared to that in HBE cells. Tip60 is upregulated in this A549 cell line, and when depleted, reduces the levels of acetylation in ABCE1. In fact, a decrease in ABCE1 acetylation by Tip60 suppression results in an impairment of cell proliferation as well as invasion and migration, with consequent induction of cell death by apoptosis [142]. The Fe-S clusters in the N-terminus of ABCE1 protein seems to be important in the proliferation and metastization state in lung adenocarcinoma. Yu and co-workers demonstrated that a deficiency of the Fe-S domain of ABCE1 reduces significantly the cell proliferation rate and migration ability of A549 cells.

In A549 cells expressing this defective variant of ABCE1 there is only minimal cytoskeletal rearrangements, and this is due to a lack of interaction between ABCE1 and β-actin, explaining the importance of the functional Fe-S clusters of ABCE1 in the context of A549 progression and migration [148]. ABCE1 protein was also shown to be overexpressed in breast cancer tissues compared to adjacent normal tissues, suggesting an involvement of this ABC protein in breast cancer progression. Huang and co-workers demonstrated that by siRNA-depletion of *ABCE1* in the breast cancer cell line MCF-7, the number of proliferating cells was reduced due to the induction of the apoptotic signaling cascade. The invasion ability of MCF-7 cells in *ABCE1* siRNA-transient transfected cells was also reduced. Suppression of ABCE1 expression was accompanied by an upregulation of RNase L, once more demonstrating that the oncogenic role of ABCE1 can be associated with a downregulation of RNase L in order to potentiate cancer cell proliferation [144].

On the other hand, Shichijo and co-workers showed that *ABCE1* mRNA is ubiquitously expressed in normal and in colorectal cancer cells. They also demonstrated that two ABCE1-derived antigens with HLA-A2 binding motifs were recognized by the reactive cytotoxic T lymphocytes from colon cancer, and therefore ABCE1 protein and its peptides could be used in targeted immunotherapy for HLA-$A2^+$ colon cancer patients [149]. Conversely, Hlavata and co-workers reported an up-regulation of *ABCE1* mRNA in a pool of colorectal cancers, putting ABCE1 on the spot for combined target therapy in colorectal cancer treatment [150]. Adding to this, in the lung cancer cell line A549, ABCE1 seems to be involved in the 5-Fluorouracil (5-FU) chemoresistance, and it was shown that a combined therapy with *ABCE1* siRNA and 5-FU chemotherapy, increases A549 susceptibility to the chemotherapeutic agent [151]. Seborova and co-workers using bioinformatics and experimental data, correlate *ABCE1* mRNA downregulation and an increase in chemosensitivity in ovarian primary tumors. They also showed that *ABCE1* mRNA upregulation in peritoneal metastasis of ovarian cancer, among other ABC transcripts, is associated with a worst disease progression and to a higher resistance [146]. In glioma cells, activation of the PI3K/Akt/NF-κB pathway is involved in ABCE1-mediated chemoresistance [147]. These and other examples well establish ABCE1 as a player in tumorigenesis.

1.5. Aims of the project

The ubiquitous presence of uORFs in the human transcriptome stresses their importance for the regulation of gene expression. Indeed, uORFs are usually seen as translational repressors of the correspondent mORF in normal physiological conditions. However, depending on the

cell environmental conditions, such as in stress, uORFs can potentiate the translation of specific transcripts in order to restore cell homeostasis. The presence of uORFs in certain classes of transcripts, as well as the genetic alterations in their sequences that can disrupt their regulatory functions, were already associated with the development of several human diseases, including cancer.

ABCE1 is a protein involved in different stages of the translation process. However, it is also involved in cell proliferation and anti-apoptotic signaling processes, which are determinant aspects that make ABCE1 an oncogenic protein. Interestingly, a bioinformatic study performed by Vanderperre and co-workers, predicted the presence of two AUG uORFs within the human *ABCE1* 5'-leader sequence [152]. In agreement, RiboSeq studies performed in different cell lines, have revealed a wide ribosome occupancy in the 5'-leader sequence of the human *ABCE1* transcript [93, 153, 154]. In addition, in the colorectal carcinoma cell line HCT116, Crappé and co-workers identified several small ORFs within the *ABCE1* 5'-untranslated region starting with non-canonical start codons [93]. These data suggest that ABCE1 may itself be subjected to a high translation regulation control, for instance, mediated by the uORFs identified in the *ABCE1* 5'-leader sequence. However, to date, no experimental validation of the *ABCE1* uORFs' functions was described.

Taking into account the unknown role of *ABCE1* uORFs, our aim was to study the biological function of the uORFs present in the human *ABCE1* 5'-unstranslated region in colorectal cancer. For that, we stablished several research objectives:

(i) Determine the impact of the *ABCE1* uORFs in mORF expression;

(ii) Dissect the mechanisms that drive uORF-mediated regulation of *ABCE1* mORF translation;

(iii) Assess if stress conditions, namely ER stress, play a role in uORF-mediated translational regulation;

(iv) Determine if the *ABCE1* uORFs and their translational regulatory function play any role in the tumorigenic process of colorectal cancer cells.

2. Material and Methods

2.1. *In silico* analyses

The Translate tool from ExPASy (web.expasy.org/translate) was used to empirically identify potential uORFs within the 5'-leader sequence of *ABCE1* mRNA, determining only uORFs starting with an AUG. The sORF.org database (sorfs.org/database) was used to get experimental information on the existence of uORFs in *ABCE1* 5'-leader sequence identified by RiboSeq. Bioedit software was used for the multiple alignment of *ABCE1* 5'-leader sequences from different species, which was then uploaded into the SMS software (bioinformatics.org/sms2/color_align_cons.html) to predict the degree of alignment conservation between them using a color-code system.

2.2. Reporter vectors

A *Firefly* luciferase (FLuc) reporter plasmid carrying the 515 nt-long 5'-leader sequence of the human *ABCE1* transcript (5'-leader sequence information obtained in Ensembl database (ensembl.org/index.html), assembly GRCh37, with the Transcript ID: ENST00000296577) was constructed using the synthesis by overlap extension (SOEing) polymerase chain reaction (PCR) approach. With the first PCR, we amplified the 5'-untranslated sequence of interest from a cDNA sample of NCM460 cells (non-tumorigenic colorectal cell line) with the primers #1 (with a linker for the HindIII restriction site; Table 2.1) and #2 (with a linker for the 5'-end of the FLuc ORF; Table 2.1). With the second PCR, we amplified the 5'-part of the FLuc ORF (region comprised between HindIII restriction site and a few nucleotides downstream of the BsrGI restriction site) using as template a modified pGL2-enhancer vector (Promega), and primers #3 (with a linker for the 3'-end of the *ABCE1* 5'-leader sequence; Table 2.1) and #4 (a primer designed for the FLuc ORF downstream of the BsrGI restriction site; Table 2.1). The pGL2-enhancer vector was previously modified to contain the human cytomegalovirus (hCMV) promotor in between BglII and HindIII restriction sites from the BglII/HindIII digested pcDNA 3.1/hygro+ (Invitrogen) and is called pGL2-FLuc or empty vector [155]. With a third PCR, the so called SOEing PCR, and the primers #1 and #4, both PCR fragments were joined in order to fuse the FLuc ORF downstream of the *ABCE1* 5'-leader sequence fragment, in a way that the FLuc AUG replaces the AUG of the transcript mORF.

Table 2.1 – Sequence of the primers used to generate the constructs needed to study the function of the AUG uORFs of the human *ABCE1* transcript.

Primer	Sequence (5' → 3')
#1	CCCAAGCTTAAGCCGTGTCGGCACCAGAC
#2	CTTTATGTTTTTGGCGTCTTCCATAACTGGGCGAAAGAATATCC
#3	GGATATTCTTTCGCCCAGTTATGGAAGACGCCAAAAACATAAAG
#4	GGACTCTGGTACAAAATCGT
#5	CTGACACCTCCAGCGTTAGATTGACTAAGGCTCCACTCCTG
#6	CAGGAGTGGAGCCTTAGTCAATCTAACGCTGGAGGTGTCAG
#7	GATTGAAACCGGAGAGGCGATTGCATCTGTTTACGCTAGGA
#8	TCCTAGCGTAAACAGATGCAATCGCCTCTCCGGTTTCAATC
#9	TGAACAAGGGTCGCAGCTCATTGACGTCATATCTCCCTACC
#10	GGTAGGGAGATATGACGTCAATGAGCTGCGACCCTTGTTCA
#11	AGGGTTCCGCCTCACGCTCTTTGTCGCGCGCGCGCACTACG
#12	CGTAGTGCGCGCGCGCGACAAAGAGCGTGAGGCGGAACCCT
#13	GCGCGCGCGCACTACGTCCTTTGGCTTGCGCGTGCGGCGGC
#14	GCCGCCGCACGCGCAAGCCAAAGGACGTAGTGCGCGCGCGC
#15	TGGATATTCTTTCGCCCAGTATGGAAGACGCCAAAAACAT
#16	ATGTTTTTGGCGTCTTCCATACTGGGCGAAAGAATATCCA
#17	CACTCCTGACCCACCGGCCTGGAGGAAAGCGCAAACGTAAA
#18	TTTACGTTTGCGCTTTCCTCCAGGCCGGTGGGTCAGGAGTG
#19	GCTGGCTTCGCCAACGGCGTGGAACAAGGGTCGCAGCTCA
#20	TGAGCTGCGACCCTTGTTCCACGCCGTTGGCGAAGCCAGC
#21	CACGCTGTGTGGCTGAAAAGGGAAGGCAAGAGCTGATTTGG
#22	CCAAATCAGCTCTTGCCTTCCCTTTTCAGCCACACAGCGTG
#23	GAACAAGGGTCGCAGCTCAAGGACGTCATATCTCCCTACCT
#24	AGGTAGGGAGATATGACGTCCTTGAGCTGCGACCCTTGTTC
#25	GTGAGAACACGCTGTGTGGCGGAAAAGTGAAGGCAAGAGCT
#26	AGCTCTTGCCTTCACTTTTCCGCCACACAGCGTGTTCTCAC
#27	TGGATATTCTTTCGCCCAGTGCGGAAGACGCCAAAAACAT
#28	ATGTTTTTGGCGTCTTCCGCACTGGGCGAAAGAATATCCA
#29	GGATATTCTTTCGCCCAGTTGCGGAAGACGCCAAAAACATA
#30	TATGTTTTTGGCGTCTTCCGCAACTGGGCGAAAGAATATCC
#31	TGACACCTCCAGCGTTACCATGGCTAAGGCTCCACTCCTG
#32	CAGGAGTGGAGCCTTAGCCATGGTAACGCTGGAGGTGTCA
#33	GATTGAAACCGGAGAGGACCATGGATCTGTTTACGCTAGG
#34	CCTAGCGTAAACAGATCCATGGTCCTCTCCGGTTTCAATC
#35	TGAACAAGGGTCGCAGCACCATGGCGTCATATCTCCCTAC
#36	GTAGGGAGATATGACGCCATGGTGCTGCGACCCTTGTTCA
#37	AGGGTTCCGCCTCACGCACCATGGCGCGCGCGCGCACTAC
#48	GTAGTGCGCGCGCGCGCCATGGTGCGTGAGGCGGAACCCT
#49	GCGCGCGCGCACTACGTACCATGGCTTGCGCGTGCGG
#40	CCGCACGCGCAAGCCATGGTACGTAGTGCGCGCGCGC
#41	AAGAGCTGATTTGGCCTCTGGCTCCCCTCCGCAAGGGGAT
#42	ATCCCCTTGCGGAGGGGAGCCAGAGGCCAAATCAGCTCTT

After digestion of the SOEing PCR product, and the pGL2-FLuc plasmid with the enzymes HindIII (Fermentas) and BsrGI (New England Biolabs), the two products were ligated with the T4 DNA ligase (NZYTech), forming the ABCE1_5'UTR (5'UTR as a simple way to refer the 5'-leader sequence). NZY5α competent cells (NZYTech) were transformed with ABCE1_5'UTR and grew in LB agar medium supplemented with ampicillin (1000x; Sigma) to select positive colonies. NZYMiniprep kit (NZYTech) was used for plasmid DNA (pDNA) extraction, following the manufacturer's instructions. The sequence of the constructed reporter plasmid was confirmed by sequencing, and the sequencing results analyzed using the Bioedit software.

All the variant constructs described below were obtained by site-directed mutagenesis following a standard protocol. Briefly, 30 ng of template (pDNA) were amplified in a PCR reaction using lengthened primers containing the genetic alteration of interest and a proof-reading enzyme (NZYSpeedy Proof DNA polymerase; NZYTech). The PCR product was digested with DpnI (Thermo Fisher Scientific) as instructed by the manufacturer, and NZY5α competent cells transformed and pDNA extracted as indicated before. All the mutagenesis constructs were confirmed by sequencing.

To study the translational regulatory function of each one of the five AUG-containing uORFs, the ABCE1_5'UTR plasmid was subjected to site-directed mutagenesis to alter the AUG start codons from an ATG to TTG (primers #5 to #14 in Table 2.1). The obtained constructs have: (i) only one functional AUG uORF (called: uORF1, uORF2, uORF3, uORF4 and uORF5); (ii) only one mutated AUG uORF (called: "no uORF1", "no uORF2", "no uORF3", "no uORF4" and "no uORF5"); (iii) different combinations of functional uORFs where uORF1 and uORF2 are always deleted (called: "no uORF1+2", "no uORF1+2+3", "no uORF1+2+4" and "no uORF1+2+5"); or (iv) none of the AUG uORFs functional (called: "no AUG uORFs").

To analyze if the AUG uORFs can in fact be translated, a site-directed mutagenesis was also applied to the constructs with only one functional AUG uORF to obtain reporter plasmids in which the uORF sequence is fused in-frame with the FLuc ORF. For that, the following mutagenesis reactions were performed: (i) for uORF3 and uORF4 constructs, 1 nt (T) was deleted in the intercistronic space between the shared uORF's stop codon and the beginning of the FLuc ORF (primers #15 and #16 in Table 2.1); and (ii) for the five uORF-individual constructs, each in-frame stop codon was mutated from TGA to GGA (primers #17 to #26 in Table 2.1). The resulting constructs are called: Fused_uORF1_AUG, Fused_uORF2_AUG, Fused_uORF3_AUG, Fused_uORF4_AUG and Fused_uORF5_AUG. Additionally, for the

five uORF-individual constructs, the FLuc start codon was mutated from ATG to GCG (primers #27 to #30 in Table 2.1), abolishing any translation initiation at the mORF. The obtained constructs are called: Fused_uORF1_GCG, Fused_uORF2_GCG, Fused_uORF3_GCG, Fused_uORF4_GCG and Fused_uORF5_GCG.

To study if reinitiation is a possible mechanism of uORF-translational regulation by these AUG uORFs, the Kozak consensus sequence of each uAUG in the constructs with only one functional uORF (uORF1-5) was altered by site-directed mutagenesis to an optimal Kozak consensus with the sequence A**ATG**G (primers #31 to #40 in Table 2.1). Those constructs are called: Optimal_uAUG1, Optimal_uAUG2, Optimal_uAUG3, Optimal_uAUG4 and Optimal_uAUG5.

To investigate if the AUG uORFs regulate translation by a mechanism of ribosomal bypass, series of sequential mutagenesis were made: (i) for uORF1, uORF2 and uORF5 constructs, 1 nt (T) was deleted in the intercistronic space between the uORFs' sequences and the beginning of the FLuc ORF (primers #41 and #42 in Table 2.1) to get out-of-frame uORFs relative to the mORF; and, (ii) for all the five uORF-individual constructs, each in-frame stop codon was mutated from TGA to GGA (primers #17 to #26 in Table 2.1). Those constructs were named: "mut stops_uORF1", "mut stops_uORF2", "mut stop_uORF3", "mut stop_uORF4" and "mut stops_uORF5".

To assess the translational regulatory function of each one of the five *ABCE1* non-canonical uORFs towards mORF expression, the construct "no AUG uORFs" was subjected to series of sequential mutagenesis to alter the near-cognate initiation codons from CTG/GTG to GCG (primers #1 to #16 in Table 2.2). The resulting constructs have only one functional non-AUG uORF (called: uORF6, uORF7, uORF8, uORF9 and uORF10) or none of the non-AUG uORFs functional (so called: "no uORFs").

Additionally, to analyze the bypass of the non-canonical uORFs by the translating ribosomes, we mutated the CTG/GTG initiation codons of each one of the non-canonical uORFs to an ATG using as template the "no AUG uORFs" construct (primers #17 to #26 in Table 2.2). The resulting constructs are termed: uORF6_AUG, uORF7_AUG, uORF8_AUG, uORF9_AUG and uORF10_AUG.

Table 2.2 – Sequence of the primers used to generate the constructs needed to study the function of the non-AUG uORFs of the human *ABCE1* transcript.

Primer	Sequence (5' → 3')
#1	GCGCGCGCGCACTACGTCCTTTGGCTTGCGCGCGCGGCGGCT
#2	AGCCGCCGCGCGCGCAAGCCAAAGGACGTAGTGCGCGCGCGC
#3	TACGTCCTATGGCTTGCGCGCGCGGCGGCTGGGCACCGCCA
#4	TGGCGGTGCCCAGCCGCCGCGCGCGCAAGCCATAGGACGTA
#5	GGCACCGCCATTTTGGCCGGCGGCCGTGAGAACACGCTGTG
#6	CACAGCGTGTTCTCACGGCCGCCGGCCAAAATGGCGGTGCC
#7	ACACGCTGTGTGGCTGAAAAGCGAAGGCAAGAGCTGATTTG
#8	CAAATCAGCTCTTGCCTTCGCTTTTCAGCCACACAGCGTGT
#9	CTGAAAAGTGAAGGCAAGAGGCGATTTGGCCTCTGTGCTCCC
#10	GGGAGCACAGAGGCCAAATCGCCTCTTGCCTTCACTTTTCAG
#11	CTGAAAAGCGAAGGCAAGAGGCGATTTGGCCTCTGTGCTCCC
#12	GGGAGCACAGAGGCCAAATCGCCTCTTGCCTTCGCTTTTCAG
#13	AAGAGGCGATTTGGCCTCTGCGCTCCCCTCCGCAAGGGGAT
#14	ATCCCCTTGCGGAGGGGAGCGCAGAGGCCAAATCGCCTCTT
#15	GATCGTTTTCTCCAGAAGAGGCGGATATTCTTTCGCCCAGTT
#16	AACTGGGCGAAAGAATATCCGCCTCTTCTGGAGAAAACGATC
#17	ACACGCTGTGTGGCTGAAAAATGAAGGCAAGAGCTGATTTG
#18	CAAATCAGCTCTTGCCTTCATTTTTCAGCCACACAGCGTGT
#19	CAAGAGCTGATTTGGCCTCTATGCTCCCCTCCGCAAGGGGA
#20	TCCCCTTGCGGAGGGGAGCATAGAGGCCAAATCAGCTCTTG
#21	GATCGTTTTCTCCAGAAGAGATGGATATTCTTTCGCCCAGT
#22	ACTGGGCGAAAGAATATCCATCTCTTCTGGAGAAAACGATC
#23	CTACGTCCTTTGGCTTGCGCATGCGGCGGCTGGGCACCGCC
#24	GGCGGTGCCCAGCCGCCGCATGCGCAAGCCAAAGGACGTAG
#25	GGGCACCGCCATTTTGGCCGATGGCCGTGAGAACACGCTGT
#26	ACAGCGTGTTCTCACGGCCATCGGCCAAAATGGCGGTGCCC

2.3. Cell culture

The human colorectal carcinoma cell line HCT116, two human colorectal adenocarcinoma cell lines DLD-1 and SW480, and the human cervical adenocarcinoma cell line HeLa, were grown in Dulbecco's Modified Eagle Medium (DMEM; Gibco, Thermo Fisher Scientific) supplemented with 10% (v/v) fetal bovine serum (FBS; Gibco, Thermo Fisher Scientific). The non-tumorigenic cell line NCM460 and two colorectal adenocarcinoma cell lines CaCo-2 and HT-29, were grown in Roswell Park Memorial Institute 1640 (RPMI; Gibco, Thermo Fisher Scientific) supplemented with 10% (v/v) FBS. Cells were maintained in an incubator at 37 ºC, in a humidifier atmosphere with 5% (v/v) CO_2.

2.4. Cell transfection with pDNA and thapsigargin treatment

HCT116 and NCM460 cells were seeded at approximately 70-80% confluence in 35 mm plates in the respective culture medium. Twenty-four hours later, cells were transiently co-transfected with 1.5 µg ABCE1_5'UTR plasmid or each one of the other above mentioned constructs, and with 0.5 µg of pRL-TK (Promega), a reporter vector expressing *Renilla* luciferase (RLuc) used as a transfection efficiency control, using Lipofectamine 2000 Transfection Reagent (Invitrogen) and following the manufacturer's instructions. Briefly, the corresponding amount of each reporter plasmid was diluted in 250 µL of Opti-MEM medium (Thermo Fisher Scientific), and 4 µL of Lipofectamine 2000 were also diluted into an equal volume of Opti-MEM. Both solutions were mixed to obtain a ratio of 1:1, and rest for at least 20 min at room temperature. The culture medium was changed to fresh growth media where the mixture above was added dropwise and cells were incubated for 24h at 37 ºC, in a humidifier atmosphere with 5% (v/v) CO_2. When appropriated, HCT116 cells were transiently transfected, and four hours after transfection, treated with 4 µM of thapsigargin (Tg; Enzo Life Sciences) or dimethyl sulfoxide (DMSO; Sigma), the vehicle control of thapsigargin, for 24h. Then, cells were harvested for further analyses.

2.5. Cell transfection with siRNAs

HCT116 and NCM460 cells were seeded at approximately 30-40% confluence in 96-well plates (Nunc) in the respective culture medium. Twenty-four hours after, 40 pmol of the siRNA for *ABCE1* (5'-UCAUCAAACCUCAAUAUGU-3') or the siRNA for Luciferase (*LUC*; 5'-CGUACGCGGAAUACUUCGA-3'), purchased as annealed and ready-to-use duplexes (Thermo Fisher Scientific), were transiently transfected using Lipofectamine 2000 according to the manufacturer instructions (same protocol as for pDNA described above in section 2.4). The *LUC* siRNA was used as control condition. Twenty-four hours after transfection, the knockdown was reinforced with more 40 pmol of the corresponding siRNA. Cells were harvested 48h after the first siRNA transfection and cleared cell lysates used for Western blot analysis. For cell viability assays, cells were kept adherent to the plate.

HeLa cells were seeded at approximately 30-40% confluence in 35 mm plates. Twenty-four hours after, cells were transiently transfected with 200 pmol of the siRNA for up-frameshift 1 regulator of nonsense transcripts yeast homolog (*UPF1*; 5'-AAGAUGCAGUUCCGCUCCAUU-3'), purchased as annealed and ready-to use duplexes (Thermo Fisher Scientific) or *LUC* siRNA, using Lipofectamine 2000 according to the manufacturer. Twenty-four hours after transfection, the knockdown was reinforced with more

100 pmol of the corresponding siRNA. Cells were harvested 30h after the second siRNA transfection for further analyses.

2.6. Luminometry assays

After pDNA transfection, cells were washed with 1 mL of pre-chilled 1× (v/v) phosphate-buffered saline (PBS 1×), lysed in 100 µL of 1× (v/v) passive lysis buffer (PLB 1×; Promega) and centrifuged at maximum speed to obtain cleared cell lysates. Relative FLuc and RLuc activity was assessed using the Dual-Luciferase® Reporter Assay System (Promega), according to the manufacturer's instructions, on a GloMax® 96 Microplate Luminometer (Promega). Briefly, 10 µL of the cleared cell lysates were plated in a white 96-well plate. First, we measured the luminescence signal of FLuc reporter by adding 40 µL of the Luciferase Assay Reagent (containing the *Firefly* luciferase substrate; Promega) to each sample. Then, RLuc reporter luminescence was sequentially quantified from the same sample by adding 40 µL of the Stop & Glo® Reagent (Promega) that first quenches the FLuc luminescence reaction and contains the substrate for the RLuc reaction. The collected data were expressed in arbitrary light units. Luciferase activity is obtained by normalizing FLuc to RLuc luminescence of each sample.

2.7. SDS-PAGE and Western blot

Cells were washed with 1 mL of pre-chilled PBS 1×, lysed in 100 µL of PLB 1× and centrifuged at maximum speed to obtain cleared cell lysates. For Western blot analysis, we used the correspondent volume of cleared cell lysate to 30 µg of total protein, previously quantified with NZYBradford reagent (NZYTech), as instructed by the manufacturer, in a spectrophotometer (NanoDrop 1000; Thermo Fisher Scientific). To the total protein sample, we added 5× (v/v) sodium dodecyl sulphate (SDS) sample buffer (NZYTech), for denaturation at 95 ºC for 15min. Samples were resolved by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) in a 10% or 12% acrylamide/bisacrylamide gel, and transfered to methanol pre-activated polyvinyldene difluoride (PVDF) membranes (Bio-Rad) for 1h. Thus, depending on the primary antibody, membranes were blocked in 5% (w/v) non-fat dry milk or bovine serum albumin (BSA; Sigma), diluted in 1× (v/v) Tris-Buffered Saline (TBS 1×) supplemented with 0.05% (v/v) Triton x-100 (Sigma) or Tween 20 (Sigma). Membranes were then blotted overnight (o/n), at 4 ºC, with shacking, with the following primary antibody solutions: mouse anti-α-tubulin (Sigma) diluted 1:50000 in 5% (w/v) non-fat milk in 0.05% (v/v) Triton-TBS 1×; rabbit anti-Pospho-eIF2α (Thermo Fisher Scientific) diluted 1:250 in 5% (w/v) BSA in 0.05% (v/v) Tween-TBS 1×; goat anti-Firefly Luciferase (Abcam) diluted 1:250 in the solution

1 for primary antibodies from the SignalBoost™ Immunoreaction Enhancer Kit (Merck Milipore); rabbit anti-UPF1 (Sigma) diluted 1:250 in 5% (w/v) BSA in 0.05% (v/v) Tween-TBS 1×; and, rabbit anti-ABCE1 (Abcam) diluted 1:1000 in 5% (w/v) non-fat milk in 0.05% (v/v) Tween-TBS 1×. Detection was performed at room temperature, with shaking, using secondary peroxidase-conjugated anti-mouse IgG (Bio-Rad), anti-rabbit IgG (Bio-Rad) or anti-goat IgG (Bio-Rad), diluted 1:4000, 1:3000 or 1:5000, respectively, in 5% (w/v) non-fat milk in the corresponding buffers, followed by enhanced chemiluminescence.

As the case of eIF2α-P and eIF2α proteins that have a similar molecular weight, membranes were stripped off from the previously used antibody and probed again, according to standard protocols. Briefly, membranes were used immediately after chemiluminescence or as dried membranes that require an additional methanol-activation. Then, membranes were washed with bi-destilled water, incubated in a 250 mM solution of NaOH, washed again in bi-destilled water and then in the respective buffer of the antibody to be used. After a blockage of 30min in 5% (w/v) non-fat milk in 0.05% (v/v) Triton-TBS 1×, membranes are probed o/n, at 4 ºC, with shacking, with rabbit anti-eIF2α (Cell Signaling) diluted 1:1000 in 5% (w/v) non-fat milk in 0.05% (v/v) Triton-TBS 1×. Detection was performed as described above, followed by enhanced chemiluminescence.

2.8. RNA extraction and cDNA synthesis

Total RNA isolation from the cleared cell lysates was performed using TRIzol reagent (Invitrogen) following the manufacturer's protocol. Total RNA was then treated with RNase-free DNase I (Promega) for approximately 1h at 37 ºC, and the RNA purified using phenol-chloroform, as described in standard protocols. cDNA was synthesized from 500 ng to 2 µg of total RNA using the NZY Reverse Transcriptase (NZYTech) and the oligo(dT) primers (NZYTech), as instructed by the manufacturer.

2.9. Quantitative RT-PCR (RT-qPCR)

The reverse transcription-quantitative PCR (RT-qPCR) was performed using the SYBR® Green PCR Master Mix (Applied Biosystems), as instructed by the manufacturer, in an Applied Biosystems® 7500 Real-time PCR System (Applied Biosystems). The cDNA synthesized above was used as template for the quantification of the mRNA expression levels of *ABCE1* (primers #5 and #6 in Table 2.3), growth arrest and DNA damage-inducible 45 alpha (*GADD45A*; primers #9 and #10 in Table 2.3) and glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*; primers #7 and #8 in Table 2.3). The amplification program was the following: an initial

denaturation at 95 ºC for 10min, followed by 40 cycles of 95 ºC for 15sec and 60 ºC for 30sec. The mRNA levels of the transcript of interest, *ABCE1* or *GADD45A*, were normalized to the mRNA levels of the internal control, *GAPDH*, and the relative mRNA expression levels were calculated by applying the comparative Ct method (ΔΔCt). For that, amplification efficiencies were calculated for each set of primers using serial dilutions of a cDNA sample. Technical replicates were performed for each qPCR reaction.

Table 2.3 – Primers used in the RT-qPCR and in the semi-quantitative RT-PCR.

Primer	Sequence (5' → 3')
#1	CAACTGCATAAGGCTATGAAGAGA
#2	ATTTGTATTCAGCCCATATCGTTT
#3	AACGCGGCCTCTTCTTATTT
#4	ACCAGATTTGCCTGATTTGC
#5	ATTAAGAAATGCCCCTTTGGCG
#6	GAAGTTTGAAGGCATTGGCACA
#7	CCATGAGAAGTATGACAACAGCC
#8	GGGTGCTAAGCAGTTGGTG
#9	GGAGGAATTCTCGGCTGGAG
#10	CGTTATCGGGGTCGACGTT

2.10. mRNA half-life analysis

After *UPF1* or *LUC* siRNA transfection, cultured cells were treated with 60 µM of adenosine analogue 5,6-dichloro-1-3-D-ribofuranosylbenzimidazole (DRB; Sigma) 28h-after the second siRNA transfection in order to inhibit transcription. Cells were then harvested at several time points (0, 0.5, 1 and 2h) and used for further analyses.

2.11. Semi-quantitative RT-PCR

The semi-quantitative reverse transcription-PCR (RT-PCR) was performed using the GoTaq® DNA polymerase (Promega) following the manufacturer's protocol. The amplification reaction was set for 1 µL of cDNA and contains 1× (v/v) Green GoTaq® Flexi buffer (5×; Promega), 2 mM of $MgCl_2$ (25 mM; Promega), 0.2 mM of dNTPs mix (10 mM; Bioline), 0.4 µM of primer forward (10 µM) and primer reverse (10 µM) for FLuc and RLuc ORF amplification (primers #1 and #2 and #3 and #4, respectively, in Table 2.3), 1.25u of GoTaq® DNA polymerase (5 u/µL), and nuclease-free water until 50 µL. The thermal cycling conditions were the following: initial denaturation at 95 ºC for 4min; 30 cycles of 95 ºC for 30sec (denaturation), 55 ºC for 30sec (annealing), and 72 ºC for 1min (extension); and then a final

extension at 72 ºC for 10 min. After amplification, 15 µL of the PCR products were resolved by electrophoresis in a 3% (w/v) agarose gel stained with ethidium bromide. The density of the bands was quantified using ImageJ software (version 1.52n), and the absolute mRNA amount determined using a standard calibration curve performed with serial dilutions of a cDNA sample. The results were expressed as mRNA levels by the normalization of FLuc to RLuc mRNA levels in each sample.

2.12. Cell proliferation assay

Cell proliferation was measured using 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT; Sigma). A 5 mg/mL stock solution prepared in sterile PBS 1× was diluted in serum-free DMEM to a final concentration of 0.5 mg/mL. After *ABCE1* or *LUC* siRNA transfection, culture medium was removed, cells were washed twice with sterile PBS 1×, and 100 µL of the above mentioning 0.5 mg/mL MTT solution was added to each well of a 96-well plate for at least 2h in an incubator at 37 ºC, in a humidifier atmosphere with 5% (v/v) CO_2. Then, MTT solution was removed and 100 µL of DMSO was added to each well for formazan crystals dissolution, by incubating for 30min with shacking and protected from light. Then absorbance at 570 nm was read in the microplate reader Multiscan Ascent spectrophotometer (Labsystems). Series of dilutions of the initial number of cells (without any treatment) were made to stablish a linear relation between the absorbance signal and the number of viable cells and thus determine the rate of cell proliferation, and thus viability, from the test conditions.

2.13. Statistical analysis

All the above experimental analyses were performed in a minimal of three independent assays. The results are expressed as mean ± standard error of the mean (SEM). Student's *t*-test, two-tailed and unpaired, was applied for statistical significance. Significance for statistical analysis was defined as $p<0.05$ (*), $p<0.01$ (**) and $p<0.001$ (***).

3. Results

3.1. Characterization of the human *ABCE1* 5'-leader sequence

The human *ABCE1* transcript has a 515 nt-long 5'-leader sequence, as annotated in Ensembl database (assembly GRCh37) with the transcript ID ENST00000296577. Using the ExPASy Translate tool, we identified five potential uAUG start codons in-frame with stop codons (UGA) located within the 5'-leader sequence of the *ABCE1* mRNA, constituting five AUG uORFs (uORF1-5; Table 3.1 and Figure 3.1A). In a bioinformatic study performed by Vanderperre and co-workers, these authors predicted the presence of two AUG uORFs within the *ABCE1* 5'-leader sequence [152], those corresponding to uORF2 and uORF3 identified by ExPASy. The first AUG uORF (uORF1) is the smallest of the five AUG uORFs, with 36 nts (11 aa) and is located close to the 5'-end of the transcript with an intercistronic distance of 426 nts to the mAUG. Downstream of uORF1 is uORF2, a 132 nt-long (43 aa) uORF leaving an intercistronic distance of 252 nts relative to the mORF. uORF3 and uORF4 have, respectively, 144 nts (47 aa) and 96 nts (31 aa) of length, both in-frame and sharing the same stop codon that is located 94 nts upstream of the mAUG. The last AUG uORF, uORF5, is 78 nt-long (25 aa) and is the most proximal uORF relative to the mORF AUG with an intercistronic distance of 84 nts. Of note, uORF5 overlaps with uORF3 and uORF4 the first 71 nts (Figure 3.1A). None of the five AUG uORFs has the initiation codon in an optimal Kozak consensus context, determined as (GCC)A/GCC**AUG**G [13] (Table 3.1). Only uORF1 and uORF5 have a good Kozak context, presenting at least in one position the adequate nucleotide: uORF1 has an A at position -3 and uORF5 has a G at position +4, relative to the A of the AUG that occupies the position +1. Additionally, uORF1 has also a G at postion -6 that is described to improve translation initiation efficiency [13, 15]. Conversely, the AUG of the *ABCE1* mORF is in a strong Kozak context with -3 and +4 positions occupied by a G nucleotide.

Table 3.1 – **Characterization of the AUG and non-AUG uORFs identified in the human *ABCE1* 5'-leader sequence.** The nucleotide and protein sequence and corresponding lengths, intercistronic distance, and Kozak sequence context are sown.

uORF ID	uORF nucleotide sequence	uORF length (nts)	Intercistronic distance (nts)	uORF-peptide sequence	uORF-peptide length (aa)	Kozak sequence context
uORF1	**ATG**ACTAAGGCTCCACTCCTGACCCACCGGCCTTGA	36	426	MTKAPLLTHRP*	11	GTTAGA**ATG**A
uORF2	**ATG**CATCTGTTTACGCTAGGACCACGCTCGACGTCGGAGAAAA GCCCACACACTCACAGTTTCCAGACCTGGCTGCTTGCCGAAACT CAGATTCTCGGCACCTCCAGCAGCTGGCTTCGCCAACGGCGTT GA	132	252	MHLFTLGPRSTSEK SPHTHSFQTWLLAE TQILGTSSSWLRQR R*	43	AGGCGA**ATG**C
uORF3	**ATG**ACGTCATATCTCCCTACCTACCTCCAGGGTTCCGCCTCACG CTCTATGTCGCGCGCGCGCACTACGTCCTATGGCTTGCGCGTG CGGCGGCTGGGCACCGCCATTTTGGCCGGTGGCCGTGAGAAC ACGCTGTGTGGCTGA	144	91	MTSYLPTYLQGSAS RSMSRARTTSYGL RVRRLGTAILAGGR ENTLCG*	47	AGCTCA**ATG**A
uORF4	**ATG**TCGCGCGCGCGCACTACGTCCTATGGCTTGCGCGTGCGGC GGCTGGGCACCGCCATTTTGGCCGGTGGCCGTGAGAACACGCT GTGTGGCTGA	96	91	MSRARTTSYGLRV RRLGTAILAGGREN TLCG*	31	CGCTCT**ATG**T
uORF5	**ATG**GCTTGCGCGTGCGGCGGCTGGGCACCGCCATTTTGGCCG GTGGCCGTGAGAACACGCTGTGTGGCTGAAAAGTGA	78	84	MACACGGWAPPF WPVAVRTRCVAEK*	25	CGTCCT**ATG**G
uORF6	**GTG**CGGCGGCTGGGCACCGCCATTTTGGCCGGTGGCCGTGAG AACACGCTGTGTGGCTGA	60	91	MRRLGTAILAGGRE NTLCG*	29	TTGCGC**GTG**C
uORF7	**GTG**GCCGTGAGAACACGCTGTGTGGCTGAAAAGTGA	36	84	MAVRTRCVAEK*	11	TGGCCG**GTG**G
uORF8	**GTG**AAGGCAAGAGCTGATTTGGCCTCTGTGCTCCCCTCCGCAA GGGGATCGTTTTCTCCAGAAGAGCTGGATATTCTTTCGCCCAGT TATGGCAGACAAGTTAACGAGAATTGCTATTGTCAACCATGA	129	- (CDS overlap)	MKARADLASVLPSA RGSFSPEELDILSPS YGRQVNENCYCQP *	42	TGAAAA**GTG**A
uORF9	**GTG**CTCCCCTCCGCAAGGGGATCGTTTTCTCCAGAAGAGCTGG ATATTCTTTCGCCCAGTTATGGCAGACAAGTTAACGAGAATTGCT ATTGTCAACCATGA	102	- (CDS overlap)	MLPSARGSFSPEEL DILSPSYGRQVNEN CYCQP*	33	GCCTCT**GTG**C
uORF10	**CTG**GATATTCTTTCGCCCAGTTATGGCAGACAAGTTAACGAGAA TTGCTATTGTCAACCATGA	63	- (CDS overlap)	MDILSPSYGRQVNE NCYCQP*	20	GAAGAG**CTG**G

Along with the five AUG uORFs, the human *ABCE1* mRNA 5'-leader sequence presents five more uORFs with non-canonical start codons (uORF6-10; Table 3.1 and Figure 3.1A), identified by RiboSeq analysis performed in HCT116 cells [93]. These uORFs are described in the sORF.org database, a repository of small ORFs identified by RiboSeq and they bear the most common non-AUG initiation codons, the CUG and the GUG. From the five non-AUG uORFs, only two have their stop codons within the 5'-leader sequence. The uORF6 is a 60 nt-long (29 aa) uORF, starts with a GUG and is in-frame with uORF3 and uORF4, sharing the same stop codon (UGA) (Figure 3.1A). The uORF7 is the smallest uORF from all the non-canonical ones, encompassing 36 nts (11 aa), it starts also with a GUG and shares the stop codon with uORF5 (Figure 3.1A). The uORF6 and uORF7 non-AUG uORFs and the last three AUG uORFs (uORF3, uORF4 and uORF5), are thus co-localized within the *ABCE1* 5'-leader sequence (Figure 3.1A). From the remaining non-AUG uORFs (uORF8-10), they share the same stop codon (UGA) that is located 41 nts downstream of the mAUG, creating overlapped and out-of-frame uORFs with the mORF (Figure 3.1A).

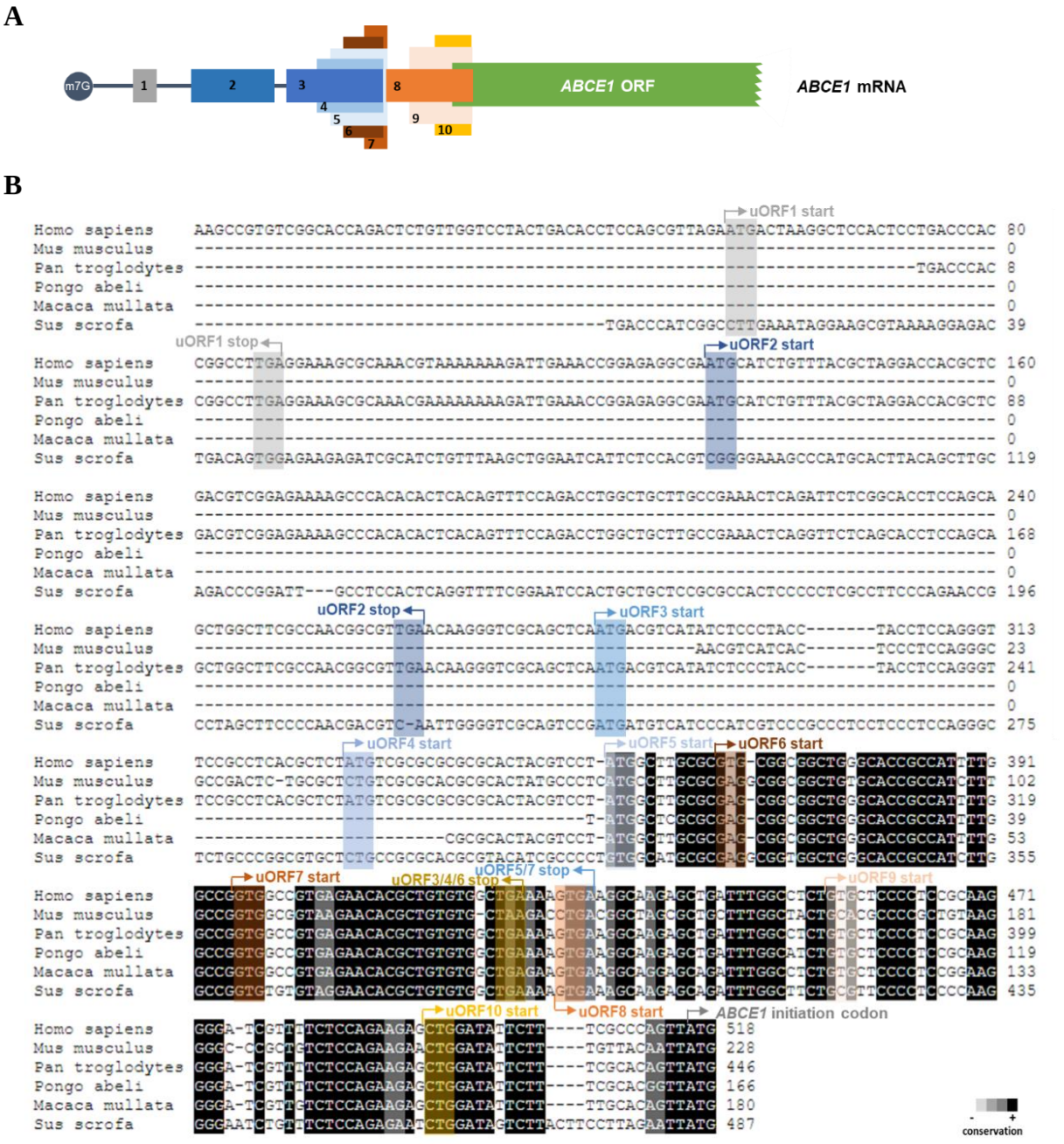


Figure 3.1 – Multiple uORFs in the human *ABCE1* mRNA 5'-leader sequence. (A) Schematic representation of the native configuration of the five AUG and five non-AUG uORFs in the 5'-leader sequence of the human *ABCE1* mRNA; **(B)** Conservation analysis of the *ABCE1* 5'-leader sequence between human and other mammalian species. The 5'-leader nucleotide sequences (plus the initiation codon of *ABCE1* mORF) of human (*Homo sapiens*), mouse (*Mus musculus*), chimpanzee (*Pan troglodytes*), orangutan (*Pongo abeli*), macaque (*Macaca mullata*) and pig (*Sus scrofa*) were aligned using Bioedit and the degree of conservation was determined using the SMS software (bioinformatics.org/sms2/color_align_cons.html). The degree of conservation is indicated by a grey scale system where white (-) represents the less conserved and black (+) the most conserved regions. Start and stop codons of each AUG and non-AUG uORF within the analyzed *ABCE1* 5'-leader sequences are shown in the color corresponding to the respective box illustrated in (A).

uORF8 is 129 nt-long (42 aa) and its GUG start codon is located 88 nts upstream of the mORF AUG. uORF9 and uORF10 have 102 nts (33 aa) and 63 nts (20 aa) in length, respectively. Additionally, uORF9 starts with a GUG that is located 61 nts upstream of the

mAUG and uORF10 has a CUG at 22 nts upstream of the main start codon. None of the five non-AUG initiation codons is in an optimal sequence context (TCCACC**NUG**G, with N being C or G [84, 88]) (Table 3.1). Only three of them have at least one nucleotide in a suitable position: uORF7 and uORF10 have a G in position +4, and uORF8 presents an A at position -3, in relation to the first nucleotide of the start codon. Additionally, uORF6, uORF7 and uORF8 have a T at position -6.

We also analyzed the conservation of *ABCE1* uORFs by aligning the sequences corresponding to the 5'-untranslated regions from human and other mammalian species using Bioedit software and then the SMS software to predict the degree of conservation by a color-code system. In Figure 3.1B, it is shown an almost perfect alignment between the 5'-leader sequences of human and chimpanzee (*Pan troglodytes*), with some pontual mismatches and a missing fragment of 72 nts at the 5'-end of *ABCE1* 5'-leader sequence of chimpanzee. When considering several species, a high degree of similarity was obtained exclusively at the 3'-end part of the 5'-leader sequence. This region fully comprises the sequences of uORF5, and consequently, uORF6 and uORF7. uORF5 start codon is highly conserved and uORF7 initiation codon is completely conserved among all the species tested. The stop codon shared between these two uORFs is totally conserved. An almost complete conservation was also observed for the uORF3/4 stop codon, shared with uORF6. The start codon of uORF8 is also very conserved while the start codon of uORF10 is totally conserved. The localization of the start codons relative to the mAUG is generally maintained. The high degree of conservation of these *ABCE1* uORFs can be associated with an important regulatory function. The complexity of the 5'-leader sequence of the human *ABCE1* mRNA and its high sequence conservation among species raises the question about the function of these uORFs in the translational regulation of the mORF.

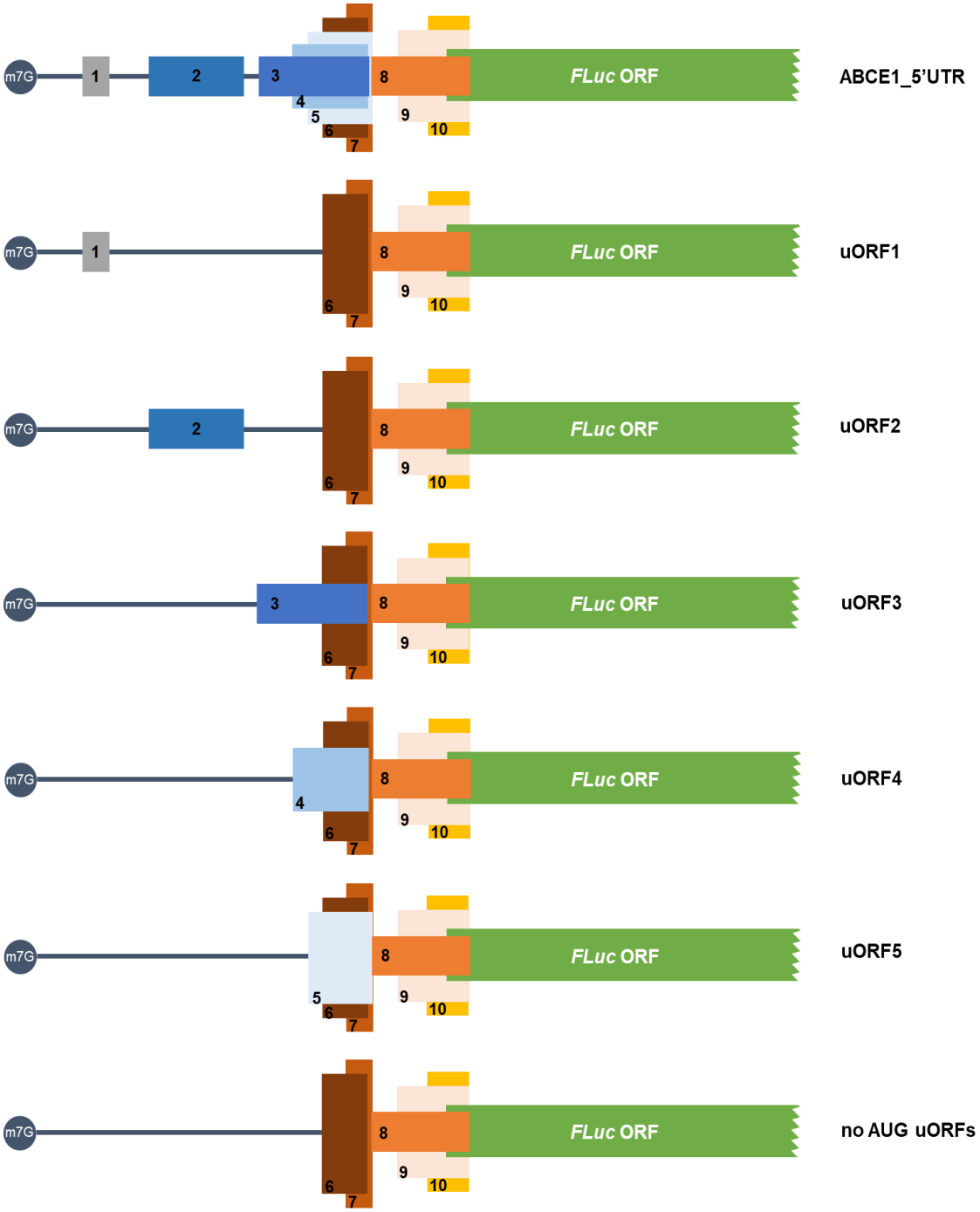
3.2. The *ABCE1* AUG uORFs repress downstream mORF translation

To study if the *ABCE1* uORFs listed above have a translational regulatory potential towards the main coding sequence, we constructed a plasmid carrying the cDNA sequence of the 515 nt-long 5'-leader sequence of the human *ABCE1* mRNA fused upstream of the FLuc ORF in the pGL2-enhancer vector (already modified to contain the hCMV promoter), called ABCE1_5'UTR (Figure 3.2A). This reporter vector was then subjected to site-directed mutagenesis, in order to inhibit the function of each AUG start codon (in cDNA: ATG → TTG) to obtain constructs with no functional AUG uORFs (called "no AUGs uORFs") or with only one functional AUG uORF (called uORF1, uORF2, uORF3, uORF4 and uORF5) (Figure 3.2A). In order to determine if the AUG uORFs in the 5'-leader sequence of *ABCE1* mRNA

have the potential to repress its mORF and which one of the five AUG-uORFs can be involved in this translational regulation, HCT116 cells were transiently co-transfected with each one of the above mentioned constructs and the reporter vector expressing RLuc, pRL-TK, as a transfection efficiency control. HCT116 cells were chosen since at least some of the *ABCE1* uORFs were identified by RiboSeq in this cell line, making it a good model to study their function in translational control. Twenty-four hours' post-transfection, cells were harvested and the lysates analyzed by luminometry assays. Relative luciferase activity was quantified by normalyzing FLuc activity to that of RLuc and then by normalizing the FLuc/RLuc ratio from each transfected vector to the one of "no AUG uORFs" construct, arbitrarily defined as 1. In addition, we also quantified the luciferase mRNA levels by semi-quantitative RT-PCR. Relative mRNA levels were determined by normalyzing FLuc mRNA levels to that of RLuc for each transfected construct followed by a normalization to the results obtained from the "no AUG uORFs" construct, arbitrarily defined as 1.

As shown in Figure 3.2B, ABCE1_5'UTR construct induced a decrease in relative luciferase activity in HCT116 cells, being oberved a significant ~3.2-fold repression when compared to that of the "no AUG uORFs". This indicates that the AUG uORFs at *ABCE1* mRNA 5'-leader sequence are able to repress translation of the mORF. Regarding the analysis of each one of the AUG uORFs, we observed that after transfection of uORF1 or uORF2 constructs, the relative luciferase activity levels were unchanged (~1.1-fold) when comparing to the results obtained from the "no AUG uORFs" construct (Figure 3.2B). This result indicates that these two AUG uORFs are not repressive. Conversely, the analysis of the uORF3, uORF4 and uORF5 constructs show that their relative luciferase activity decreases relative to the effect promoted by the "no AUG uORFs" reporter construct (Figure 3.2B). uORF3 and uORF4, contributed to a repression of ~1.8-fold and ~1.6-fold, respectively, whereas uORF5 inhibited around 2.2-fold the relative luciferase activity, making uORF5 the most repressive uORF.

A


m7G
1
2
3
4
5
6
7
8
9
10
FLuc ORF
ABCE1_5'UTR
uORF1
uORF2
uORF3
uORF4
uORF5
no AUG uORFs

B

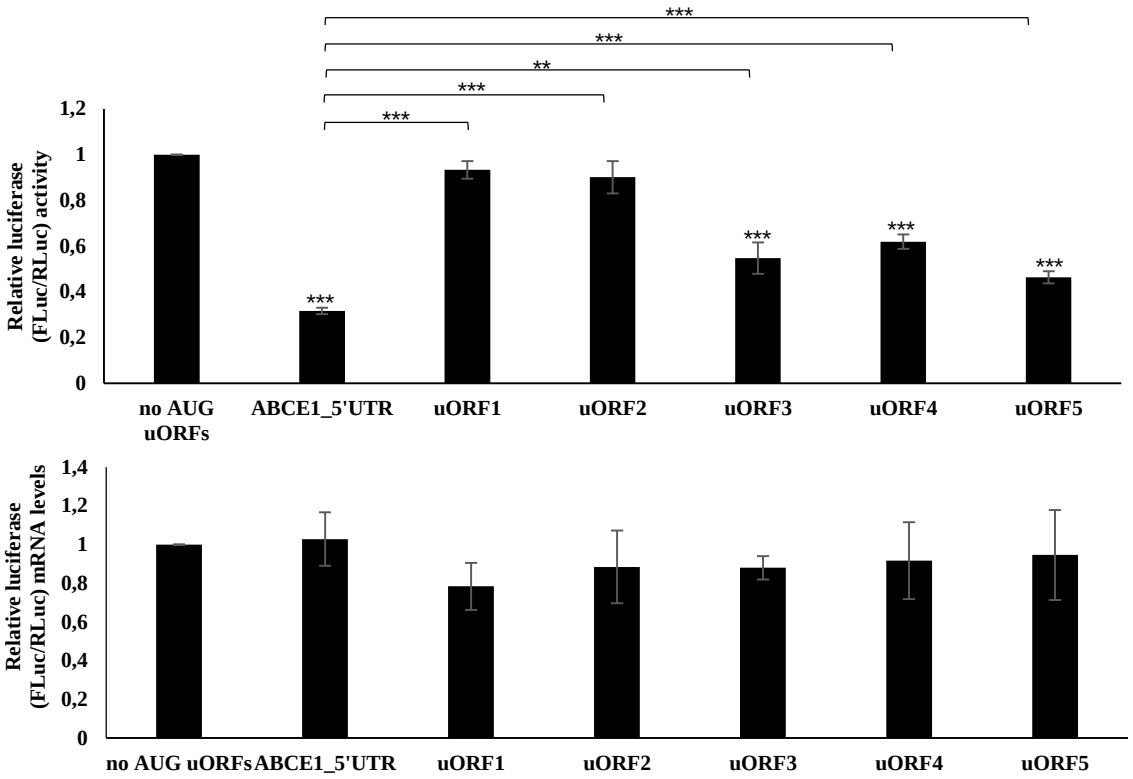


Figure 3.2 – The *ABCE1* AUG uORFs repress downstream mORF translation. (A) Schematic representation of the constructs used to study the translational regulatory function of the *ABCE1* AUG uORFs. The human *ABCE1* 5'-leader sequence was cloned into the pGL2-FLuc vector, upstream of the FLuc ORF (green box), obtaining the ABCE1_5'UTR construct. By site-directed mutagenesis, and using the ABCE1_5'UTR vector as template, each AUG initiation codon was mutated (in cDNA: ATG → TTG) creating the constructs with only one functional AUG uORF (uORF1-5, represented here by small boxes and only by its uORF number, 1 to 5, to simplify the scheme) or without functional AUG uORFs ("no AUG uORFs"). The uORFs represented have an UGA as termination codon. Non-AUG uORFs (represented in small boxes with the numbers 6 to 10) are mantained intact in these constructs; **(B)** Translational repression of FLuc reporter by the AUG uORFs within *ABCE1* 5'-leader sequence. HCT116 cells were transiently co-transfected with each one of the constructs described in (A) and the pRL-TK plasmid expressing RLuc, used as a transfection efficiency control. Twenty-four hours' post-transfection cells were harvested and lysed. Luciferase activity was measured by luminometry assays. The results are expressed as relative luciferase activity determined by normalyzing FLuc activity to that of RLuc and then by normalizing the FLuc/RLuc ratio from each transfected vector to the one of "no AUG uORFs" construct, arbitrarily defined as 1 (upper panel). In parallel, luciferase mRNA levels were determined by semi-quantitative RT-PCR and the results presented as relative mRNA levels determined by normalyzing FLuc mRNA levels to that of RLuc for each transfected construct followed by a normalization to the FLuc/RLuc obtained from the "no AUG uORFs" construct, also arbitrarily defined as 1 (lower panel). The results are presented as mean ± SEM from at least three independent experiments. Student's t-test was applied for statistical significance: $p<0.01$ (**) and $p<0.001$ (***).

However, none of the constructs carrying only one of the repressive AUG uORFs exhibited similar repression effect as the intact 5'-leader sequence in the construct ABCE1_5'UTR. The relative luciferase activity was increased in ~1.7-fold for uORF3, ~2-fold for uORF4, and ~1.5-fold for uORF5 relative to the luciferase expression induced by the ABCE1_5'UTR construct. This increase in relative luciferase activity points out for a synergistic/additive effect and/or a fail-safe mechanism of these three AUG uORFs in the repression of the mORF. We did not

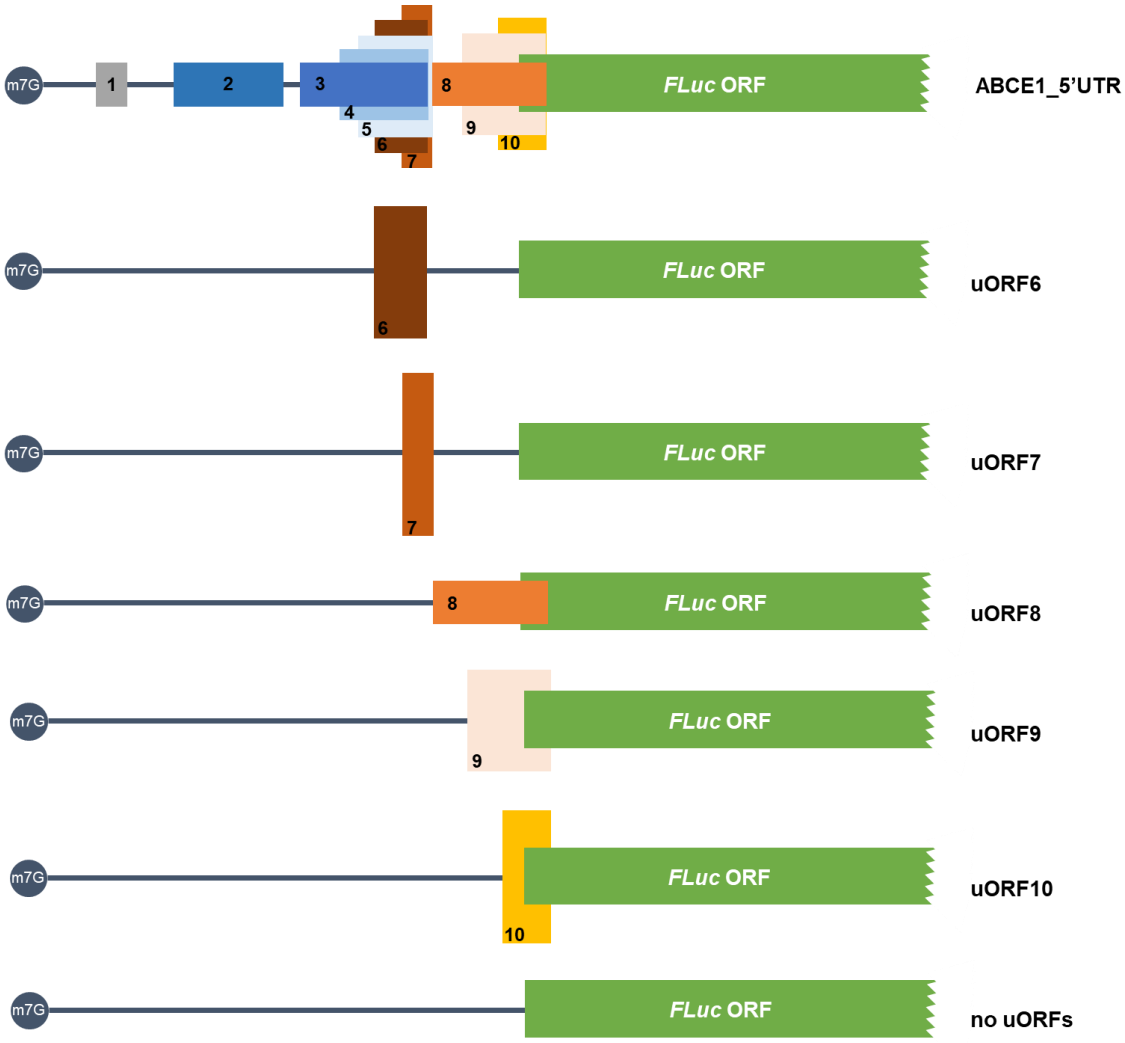
observe any significant differences in relative mRNA levels, showing that the effects reported in relative luciferase activity are due to a post-transcriptional regulation by the AUG uORFs and not due to alterations in the mRNA levels (Figure 3.2B; lower panel).

In summary, the 5'-leader sequence of the *ABCE1* transcript regulates mORF expression in HCT116 cells, and this regulatory function seems to be played by three AUG uORFs, the uORF3, uORF4 and uORF5, which may cooperate to reach a maximum repression at the mAUG.

3.3. The *ABCE1* non-AUG uORFs do not show any repressive function

To analyze the potential function of the *ABCE1* non-canonical uORFs, site-directed mutagenesis of each non-AUG start codon (in cDNA: CTG/GTG → GCG) was performed using the "no AUG uORFs" construct as template, to obtain constructs with no functional uORFs (without AUG nor non-AUG start codons, called "no uORFs" construct) and constructs carrying only one functional non-AUG uORF (called uORF6, uORF7, uORF8, uORF9 and uORF10) (Figure 3.3A). HCT116 cells were transiently co-transfected with these constructs and also with the control pRL-TK, as described above. The FLuc and RLuc activities were assessed by luminometry assays and relative luciferase activity was also determined by normalizing FLuc activity to that of RLuc and then normalized to that of the "no uORFs" construt, arbitrarily defined as 1. The FLuc and RLuc mRNA levels were quantified by semi-quantitative RT-PCR and relative mRNA levels determined by normalyzing FLuc mRNA to that of RLuc and then normalizing to the one of "no uORFs" construct (arbitrarily defined as 1), showing that the mRNA levels were expressed with no significant alterations in all tested conditions (Figure 3.3B).

A


m7G
1
2
3
4
5
6
7
8
9
10
FLuc ORF
ABCE1_5'UTR
uORF6
uORF7
uORF8
uORF9
uORF10
no uORFs

B

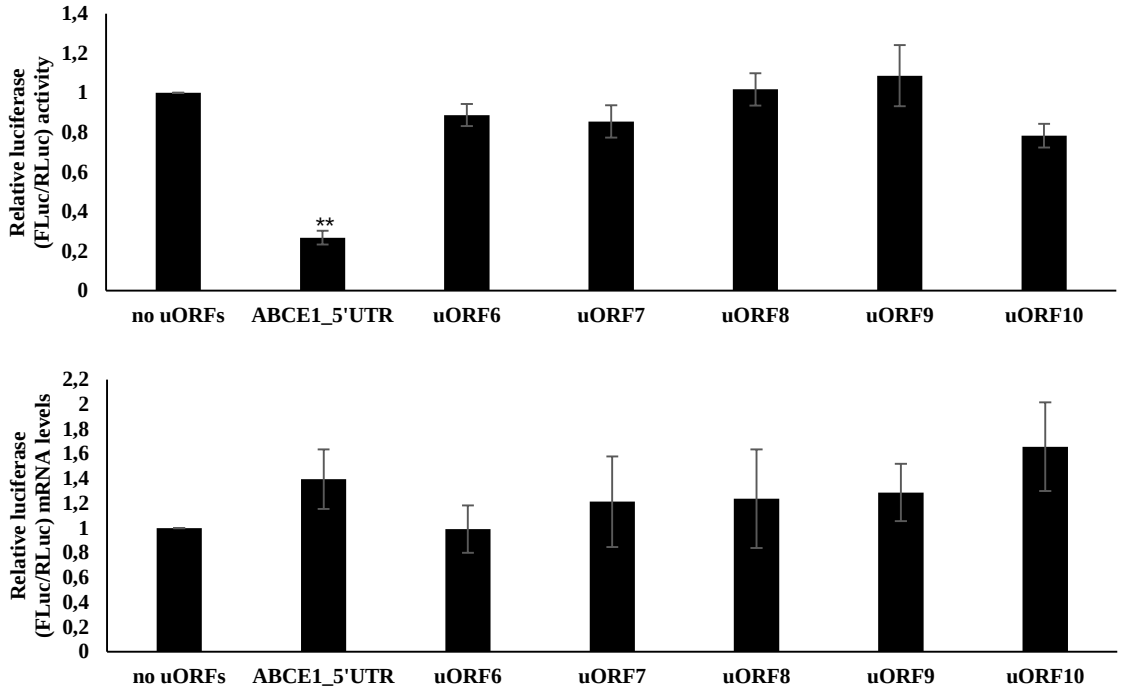


Figure 3.3 – Non-AUG uORFs within the *ABCE1* mRNA 5'-leader sequence do not repress translation of the mORF. (A) Schematic representation of the constructs used to study the translational regulatory function of the *ABCE1* non-AUG uORFs. The ABCE1_5'UTR, containing the intact *ABCE1* 5'-leader sequence, was already described in Figure 3.2A. By site-directed mutagenesis, and using the "no AUG uORFs" construct as template, each non-AUG initiation codon was mutated (in cDNA: CTG/GTG → GCG) creating the constructs with only one functional non-AUG uORF (uORF6-10, represented here by small boxes and only by its uORF number, 6 to 10, to simplify the scheme) or without any uORF (called "no uORFs"). uORF6-9 have a GTG as initiation codon and uORF10 contains a CTG start codon. The represented uORFs have an UGA as termination codon; **(B)** Lack of translational repression at FLuc mORF by the non-AUG uORFs within *ABCE1* 5'-leader sequence. HCT116 cells were transiently co-transfected with each one of the constructs described in (A) and the pRL-TK plasmid. Twenty-four hours' post-transfection, cells were harvested and lysed. Luciferase activity was measured by luminometry assays. The results are expressed as relative luciferase activity determined by normalyzing FLuc activity to that of RLuc and then by normalizing the FLuc/RLuc ratio from each transfected vector to the one of "no uORFs" construct, arbitrarily defined as 1 (upper panel). In parallel, luciferase mRNA levels were determined by semi-quantitative RT-PCR and the results presented as relative mRNA levels determined by normalyzing FLuc mRNA levels to that of RLuc for each transfected construct followed by a normalization to the FLuc/RLuc obtained from the "no uORFs" construct, also arbitrarily defined as 1 (lower panel). The results are presented as mean ± SEM from at least three independent experiments. Student's t-test was applied for statistical significance: $p<0.01$ (**).

As shown in Figure 3.3B, a significant decrease in the relative luciferase activity for ABCE1_5'UTR construct was observed again, representing a ~3.7-fold repression when compared to the relative luciferase activity obtained by the "no uORFs" construct. However, none of the non-functional uORFs has a significant inhibitory effect when compared to the expression of the "no uORFs" construct (Figure 3.3B).

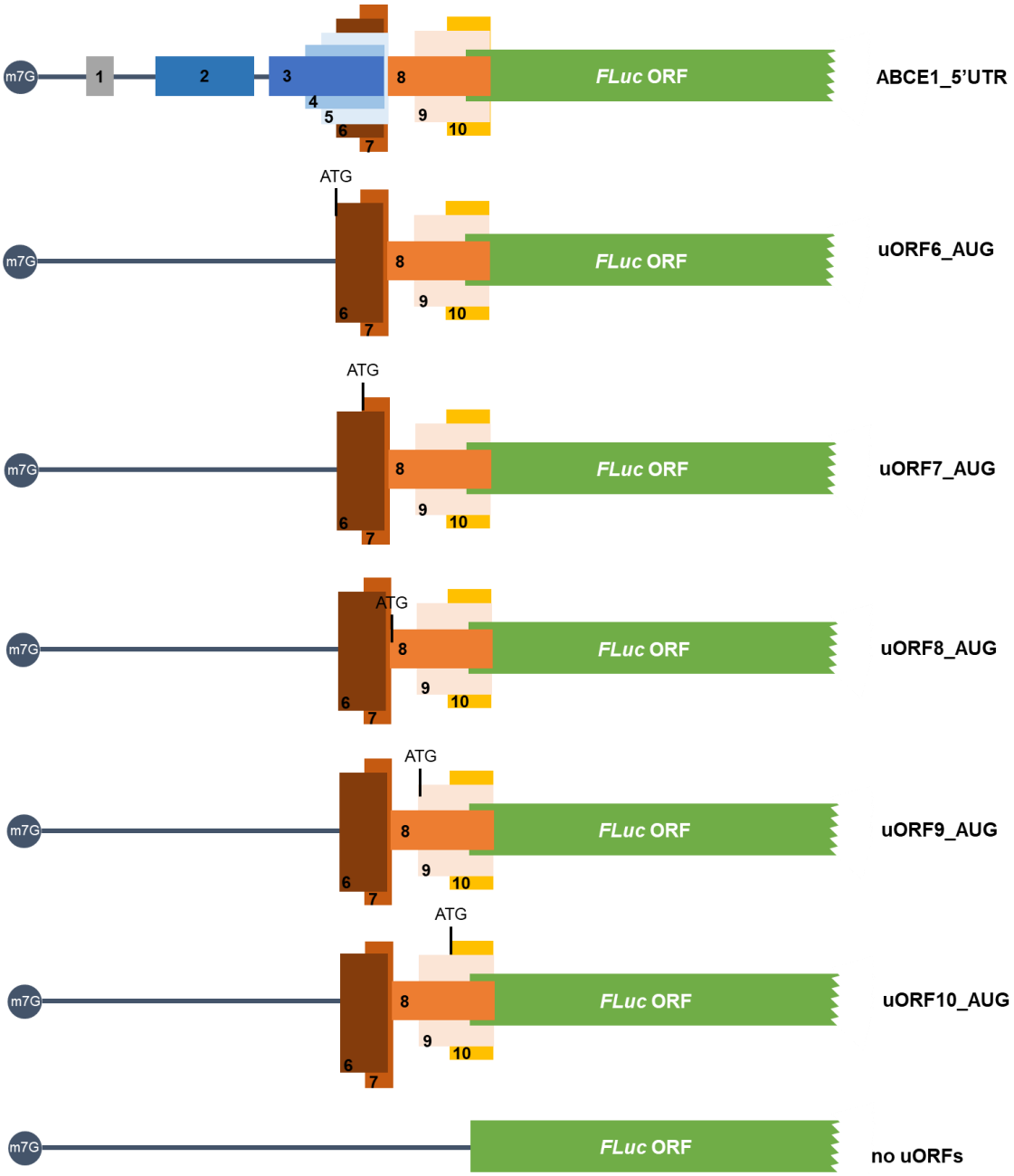
In addition, to understand if the low inhibitory activity of the non-AUG uORFs in the *ABCE1* 5'-leader sequence is due to their frequent ribosomal bypass as a consequence of their non-canonical nature, we mutated the start codon of each one of the non-AUG uORFs from CTG/GTG to an ATG, using as template the construct "no AUG uORFs", resulting the

constructs uORF6_AUG, uORF7_AUG, uORF8_AUG, uORF9_AUG and uORF10_AUG (Figure 3.4A). This is based on the fact that translation initiation in AUG start codons is more efficient when compared to the non-canonical ones [3]. Once more, those constructs or the controls ABCE1_5'UTR and "no uORFs" were co-transfected with pRL-TK into HCT116 cells. FLuc/RLuc activity was measured by luminometry assays and the corresponding mRNA levels by semi-quantitative RT-PCR.

After determining that there were no significant alterations in the relative mRNA levels in each transfected condition, except for ABCE1_5'UTR (Figure 3.4B), we were able to show that the relative levels of luciferase activity were significantly reduced to approximately half or completely abolished when compared to those of the "no uORFs" construct (Figure 3.4B). Indeed, uORF6_AUG, uORF7_AUG and uORF9_AUG reduced, respectively, in ~1.7-fold, ~2.3-fold and ~4.3-fold the relative luciferase activity when in comparison to the one of the "no uORFs" control. In addition, uORF8_AUG and uORF10_AUG induced almost no relevant levels of luciferase activity when compared to results obtained from the "no uORFs" construct. The strong repression exherted by uORF8, uORF9 and uORF10 in this configuration can be in part associated with their overlapping nature. These results highlight that when a non-canonical uORF start codon is replaced by an AUG, the uORF start codon becomes strongly recognized by the scanning ribosomes. Thus, these results show that the non-canonical start codons of these uORFs are usually bypassed by the translation machinery due to their non-canonical nature and also due to their weak start codon Kozak context, leading instead to the translation of the mORF. This is in agreement to the low level of repression exherted by those non-canonical uORFs (Figure 3.3B).

In conclusion, the *ABCE1* uORFs bearing non-canonical initiation codons are not significantly recognized by the translational machinery, and thus do not play a role as translational repressors of the mORF.

A


m7G
1
2
3
4
5
6
7
8
9
10
FLuc ORF
ABCE1_5'UTR
ATG
uORF6_AUG
uORF7_AUG
uORF8_AUG
uORF9_AUG
uORF10_AUG
no uORFs

B

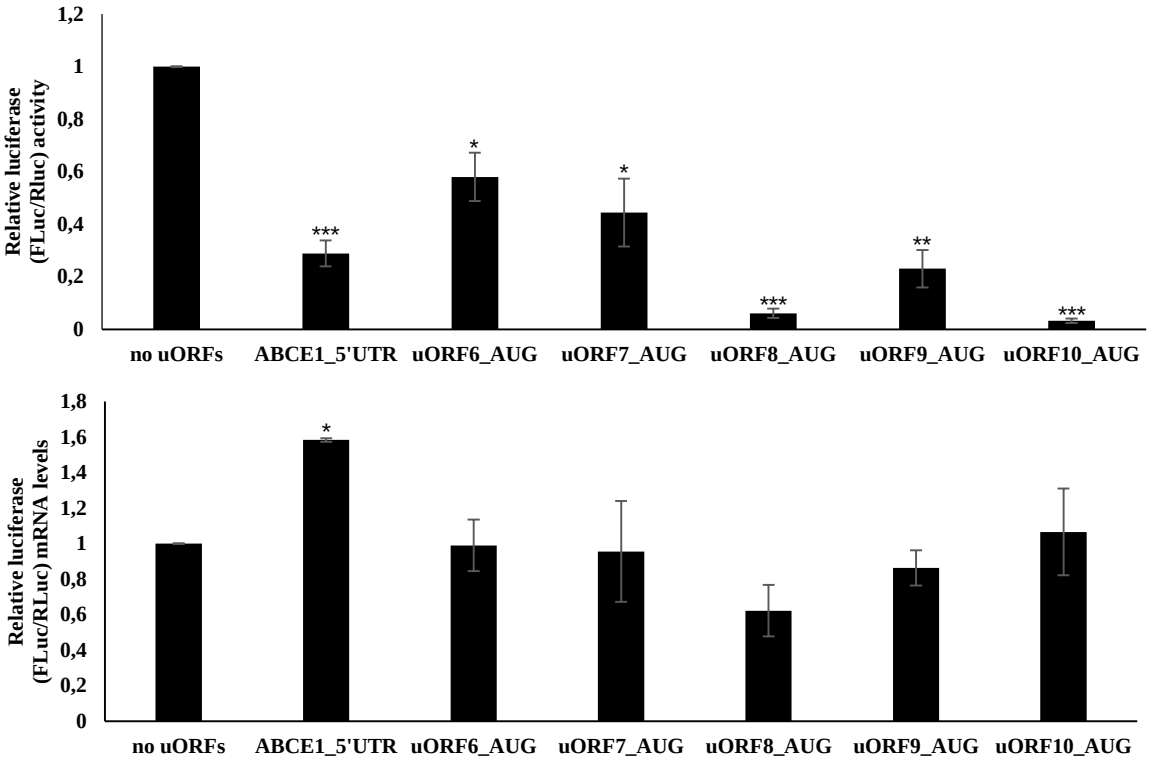


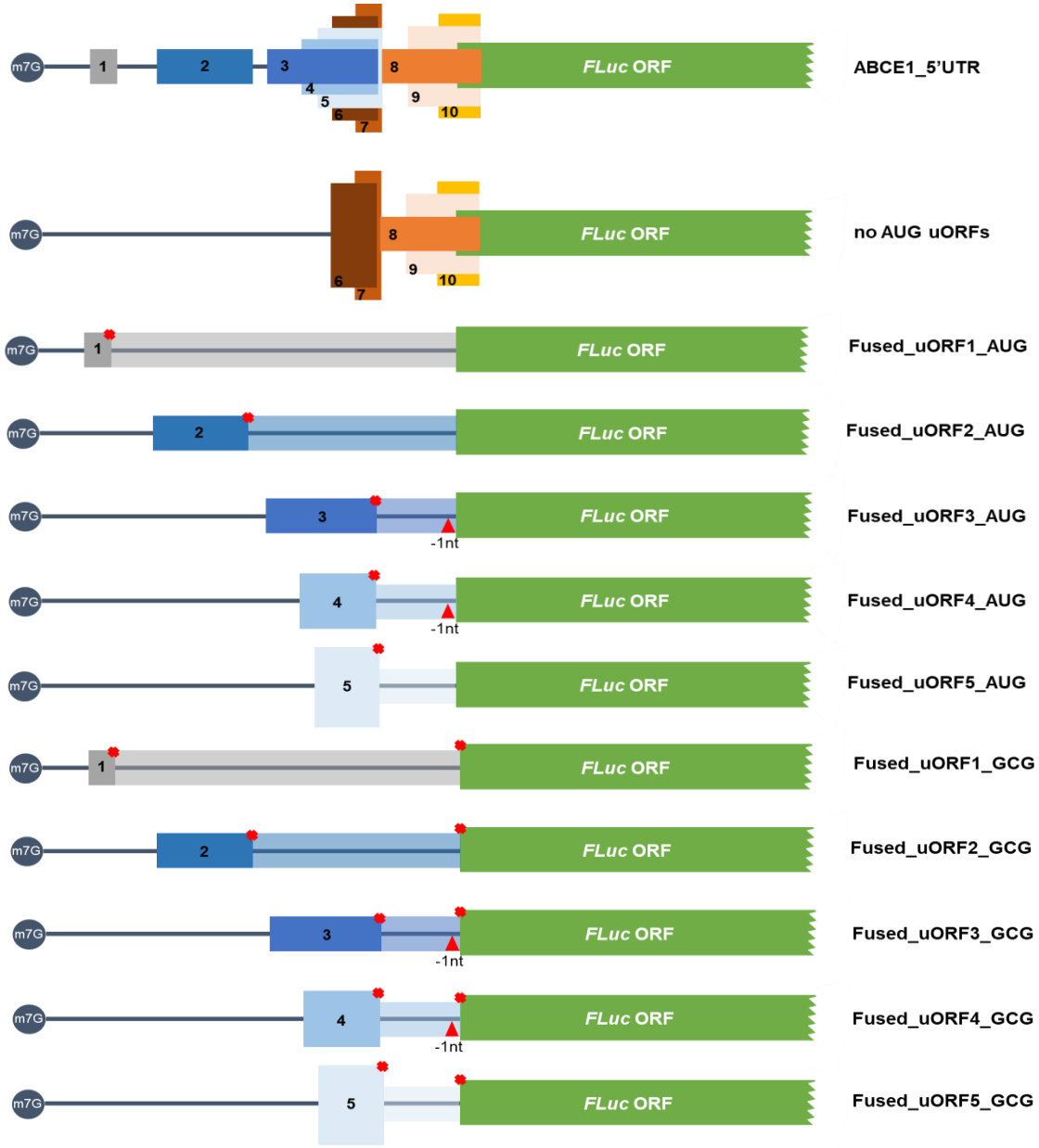
Figure 3.4 – The *ABCE1* non-AUG uORFs are not recognized by the scanning ribosome and thus, they do not affect mORF translation efficiency. (A) Schematic representation of the constructs used to study the ribossomal bypass of each one of the *ABCE1* non-AUG uORFs. The ABCE1_5'UTR construct, containing the intact *ABCE1* 5'-leader sequence, and the "no uORFs" construct, without any functional AUG and non-AUG uORFs, were already described in Figure 3.2A and Figure 3.3A, respectively. By site-directed mutagenesis, and using the construct "no AUG uORFs" (described in Figure 3.2A) as template, the start codon of each one of the non-AUG uORFs was mutated (in DNA: CTG/GTG → ATG), and the following constructs were obtained: uORF6_AUG, uORF7_AUG, uORF8_AUG, uORF9_AUG and uORF10_AUG. The uORFs represented have an UGA as termination codon; **(B)** The upstream non-canonical initiation codons within *ABCE1* 5'-leader sequence are bypassed by the scanning ribsomes and translation initiates at the AUG of the FLuc reporter. HCT116 cells were transiently co-transfected with each one of the constructs described in (A) and the pRL-TK plasmid. Twenty-four hours' post-transfection cells were harvested and lysed. Luciferase activity was assessed by luminometry assays (upper panel) and the corresponding mRNA levels by semi-quantitative RT-PCR (lower panel). Data were then analyzed as described in Figure 3.3B. The results are presented as mean ± SEM from at least three independent experiments. Student's t-test was applied for statistical significance: $p<0.05$ (*), $p<0.01$ (**) and $p<0.001$ (***).

3.4. The *ABCE1* AUG uORFs are recognized by the translation machinery

Taking into account the repressive activity of the AUG uORFs of the *ABCE1* transcript, we wanted to clearly understand if the start codon of each AUG uORF is recognized by the scanning ribosomes. For that, reporter vectors carrying each one of the AUG uORFs fused in-frame with the FLuc ORF were constructed (Figure 3.5A). To obtain those constructs, a site-directed mutagenesis strategy was applied to the constructs with only one functional AUG uORF. For uORF1, uORF2 and uORF5 (already in-frame with the FLuc ORF), the uORF stop codon and all the other in-frame stop codons were mutated to a sense codon (in cDNA: TGA → GGA). For uORF3 and uORF4, the common stop codon was also mutated (in cDNA: TGA

→ GGA) and, one nucleotide was deleted in the intercistronic region in order to make these two uORFs in-frame with the mORF. The resulting constructs are the Fused_uORF1_AUG, Fused_uORF2_AUG, Fused_uORF3_AUG, Fused_uORF4_AUG and Fused_uORF5_AUG. In parallel, we also mutated the AUG start codon of FLuc (in cDNA: ATG → GCG) using the above mentioned fused constructs as template, and thus impairing translation initiation at the FLuc initiation codon. The resulting constructs are Fused_uORF1_GCG, Fused_uORF2_GCG, Fused_uORF3_GCG, Fused_uORF4_GCG and Fused_uORF5_GCG. HCT116 cells were transiently transfected with the above mentioned constructs, or with the control constructs ABCE1_5'UTR, "no AUG uORFs" or the empty vector pGL2-FLuc (to monitor native FLuc expression). Native FLuc protein and N-extended uORF-FLuc isoforms were assessed by Western blot using a specific antibody for FLuc (Figure 3.5B).

A


m7G
1
2
3
4
5
6
7
8
9
10
FLuc ORF
ABCE1_5'UTR
no AUG uORFs
Fused_uORF1_AUG
Fused_uORF2_AUG
Fused_uORF3_AUG
Fused_uORF4_AUG
Fused_uORF5_AUG
Fused_uORF1_GCG
Fused_uORF2_GCG
Fused_uORF3_GCG
Fused_uORF4_GCG
Fused_uORF5_GCG
-1nt

B

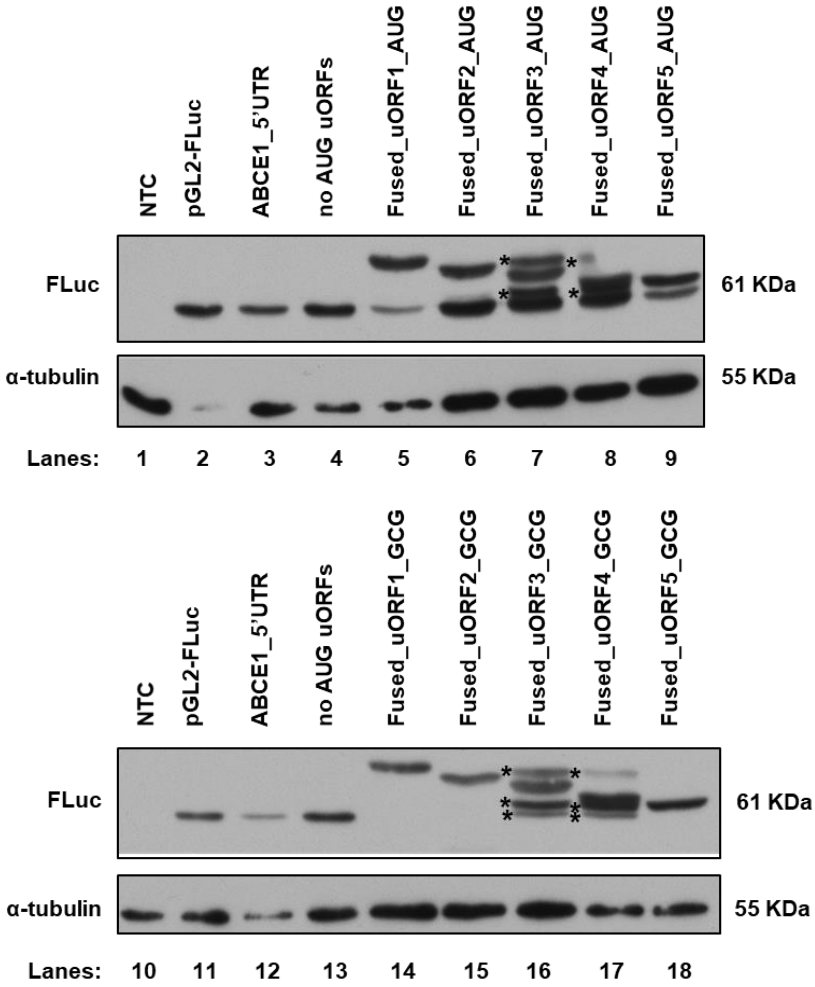


Figure 3.5 – Translation initiation can occur at the *ABCE1* uAUGs. (A) Schematic representation of the constructs used to analyze the translation initiation efficiency at the uAUGs in the *ABCE1* 5'-leader sequence. The ABCE1_5'UTR, containing the intact *ABCE1* 5'-leader sequence, and the "no AUG uORFs" construct, without AUG functional uORFs, were already described in Figure 3.2A. A site-directed mutagenesis approach was used to obtain constructs with each AUG uORF fused in-frame with the FLuc ORF. In the constructs uORF3 and uORF4, 1 nt (T) was deleted from the intercistronic space and, for the five AUG uORF-individual constructs (uORF1-5), each in-frame stop codon was mutated (in cDNA: TGA → GGA), resulting the following constructs: Fused_uORF1_AUG, Fused_uORF2_AUG, Fused_uORF3_AUG, Fused_uORF4_AUG and Fused_uORF5_AUG. The AUG of FLuc ORF in these constructs was also mutated (in cDNA: ATG → GCG), obtaining the constructs: Fused_uORF1_GCG, Fused_uORF2_GCG, Fused_uORF3_GCG, Fused_uORF4_GCG and Fused_uORF5_GCG. The red crosses represent the point mutations at the uORF's stop codons and in the FLuc initiation codon and the red triangle indicates the deletion of 1 nt. The five non-AUG uORFs are present in these constructs but not represented in the figure to simplify the scheme; **(B)** The uAUG initiation codons are recognized by the scanning ribosomes resulting in translation of the *ABCE1* AUG uORFs. HCT116 cells were transiently transfected with each one of the constructs described in (A) or with the empty vector pGL2-FLuc (to monitor native FLuc expression). Twenty-four hours' post-transfection, cells were harvested and lysed. Native FLuc protein and N-extended uORF-FLuc isoforms were assessed by Western blot using a specific antibody against FLuc. α-tubulin was used as loading control. These are representative immunoblots from at least three indepedent experiments. Note: the asterisks (*) represent translation initation at other initiation codons; NTC – non-template control.

As shown in the immunoblots of Figure 3.5B, the constructs ABCE1_5'UTR and "no AUG uORFs" (lanes 3 and 4, respectively) express a protein with a molecular weight of 61 kDa that corresponds to the native FLuc protein also observed in cells transfected with the empty vector

(lane 2). Of note, there is a decrease in the FLuc expression in the extracts of HCT116 cells transfected with ABCE1_5'UTR construct (lane 3) when compared to that from the "no AUG uORFs" construct (lane 4), which confirms the repression activity of the intact *ABCE1* mRNA 5'-leader sequence, as previously demonstrated (Figure 3.2.B). For all the fused constructs carrying the FLuc AUG start codon (lanes 5 to 9), we observe bands with a molecular weight higher than the one expected for the native FLuc protein (61 kDa), corresponding to each one of the uORF-FLuc fused proteins (Figure 3.5B, lanes 5-9). These results demonstrate that all the AUG uORFs have the potential to be translated; however, the majority of the ribosomes that load onto the human *ABCE1* mRNA 5'-leader sequence are able to recognize the uORF1 start codon. Moreover, we observe additional bands in lane 7 and 8 (labeled with asterisks), that may correspond to N-terminally extended FLuc proteins due to translation initiation at other in-frame start codons. Also, all the fused constructs analyzed in lanes 5 to 9 express the native FLuc protein which indicates that there is some degree of ribosomal bypass of the five uORFs. Indeed, the poor Kozak context of the uORF2, uORF3 and uORF4 initiation codons can explain these ribosomal bypass events, as shown in lanes 6 to 8 of Figure 3.5B, where we observe a higher expression of the native FLuc protein compared to the corresponding uORF-FLuc fused isoforms. One the other hand, in uORF1 and uORF5 this bypass seems to be less efficient, as, in fact the expression of the native FLuc protein is lower when compared to that of the uORF-FLuc fused protein (lanes 5 and 9, respectively, in Figure 3.5B). This correlates well with the better start codon Kozak context of uORF1 and uORF5 (Table 3.1). When we analyzed by Western blot the expression of the fused constructs carrying a mutation at the FLuc AUG (lanes 14 to 18 in Figure 3.5B), we only observe the expression of the expected uORF-FLuc fused proteins. Once more, in lanes 16 and 17, we also observe the presence of additional bands. We speculate that those bands can be the result of translation initiation at non-canonical initiation sites in-frame with the FLuc ORF.

In conclusion, the five uAUGs present within the human *ABCE1* 5'-leader sequence can be recognized by the translating ribosomes, althougth the AUGs of uORF1 and uORF5 are recognized more efficiently, and this is a strong indication that they can be involved in the translational regulation of the mORF.

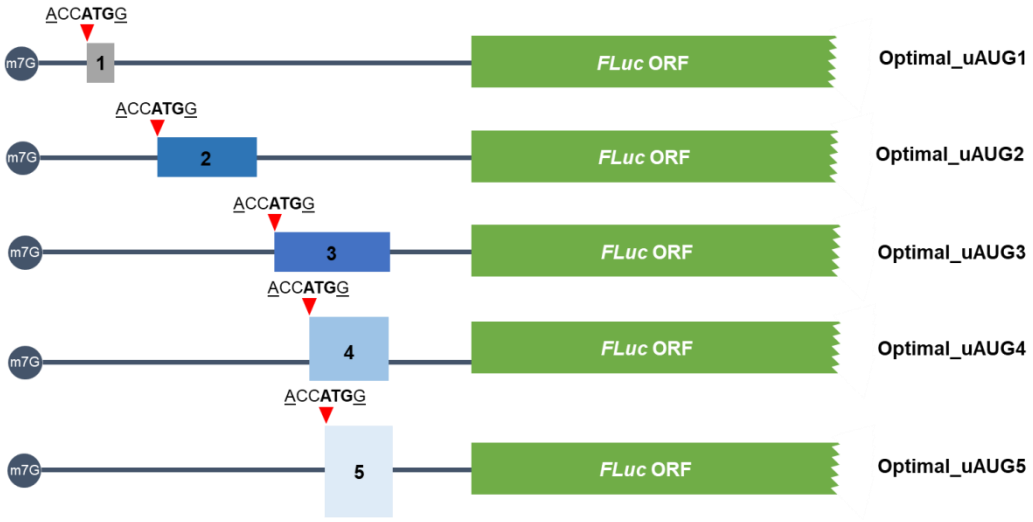
3.5. The uORFs present in the *ABCE1* 5'-leader sequence allow for translation reinitiation and ribosomal bypass

Taking into account the results obtained above in section 3.2 and 3.4, we speculate that due to the good Kozak context of uORF1 and uORF5 start codons, they can be frequently

recognized by the scanning ribosomes that load onto the *ABCE1* 5'-leader sequence, and after the corresponding uORF translation, the ribosomes can resume scanning and reinitiate translation at downstream start codons, including the mAUG. Although less frequently, the ribosomes may bypass the uORF1 and uORF5 start codons, initiating translation in downstream initiation codons. Conversely, the weak Kozak context of uORF2, uORF3 and uORF4 initiation codons may allow the scanning ribosomes to frequently bypass them, initiating at downstream start codons, including at the mAUG. However, we cannot rule out some degree of translation reinitiation at the mORF after translation of these uORFs, once the start codon of uORF2, uORF3 or uORF4 is recognized.

To study the potential of translation reinitiation at the mORF of each AUG uORF, the Kozak context of each uORF initiation codon was replaced by an optimal Kozak sequence context (in cDNA: ACC**ATG**G). This was achieved by site-directed mutagenesis in the constructs carrying only one functional AUG uORF (uORF1-5), resulting in the constructs Optimal_uAUG1, Optimal_uAUG2, Optimal_uAUG3, Optimal_uAUG4 and Optimal_uAUG5 (Figure 3.6A). This alteration increases the probability of the ribosomes to recognize the uORF initiation codon, and thus, FLuc expression will only take place if translation reinitiation at the mAUG occurs after uORF translation. To test this hypothesis, HCT116 cells were transiently co-transfected with these constructs or with the constructs with only one functional AUG uORF, the ABCE1_5'UTR or the "no AUG uORFs" construct, and with the pRT-TK control plasmid. mRNA levels were quantified by semi-quantitative RT-PCR and the relative mRNA levels determined as the FLuc/RLuc ratio from each transfected construct normalized to that of the "no AUG uORFs" construct. After to observe that the relative mRNA levels do not significantly change (Figure 3.6B), relative luciferase activity was quantified (Figure 3.6B). FLuc activity was measured by luminometry assays and normalized to that of RLuc in each case and the relative luciferase activity determined by normalizing to the results obtained from the "no AUG uORFs" construct. The relative luciferase activity of the optimal constructs is compared to that of the corresponding construct with only one functional AUG uORF.

A


ACCATGG
m7G
1
FLuc ORF
Optimal_uAUG1
2
Optimal_uAUG2
3
Optimal_uAUG3
4
Optimal_uAUG4
5
Optimal_uAUG5

B

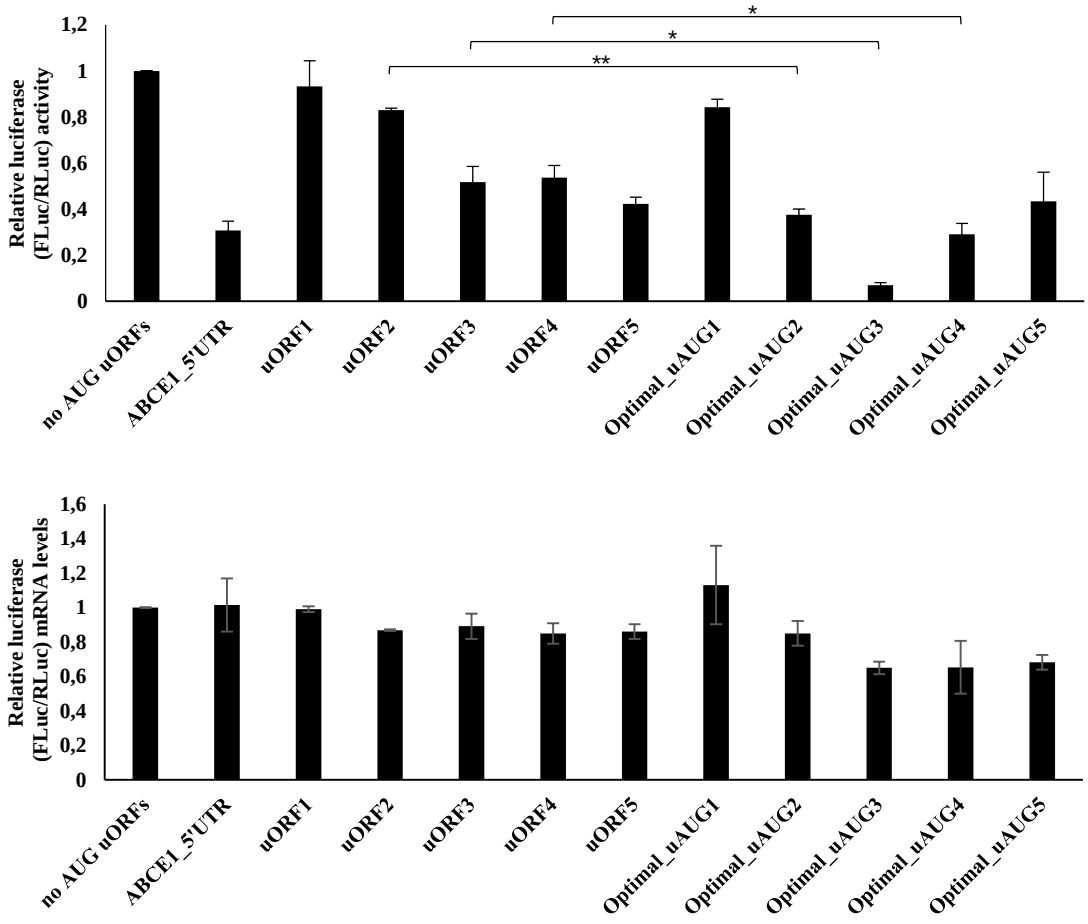
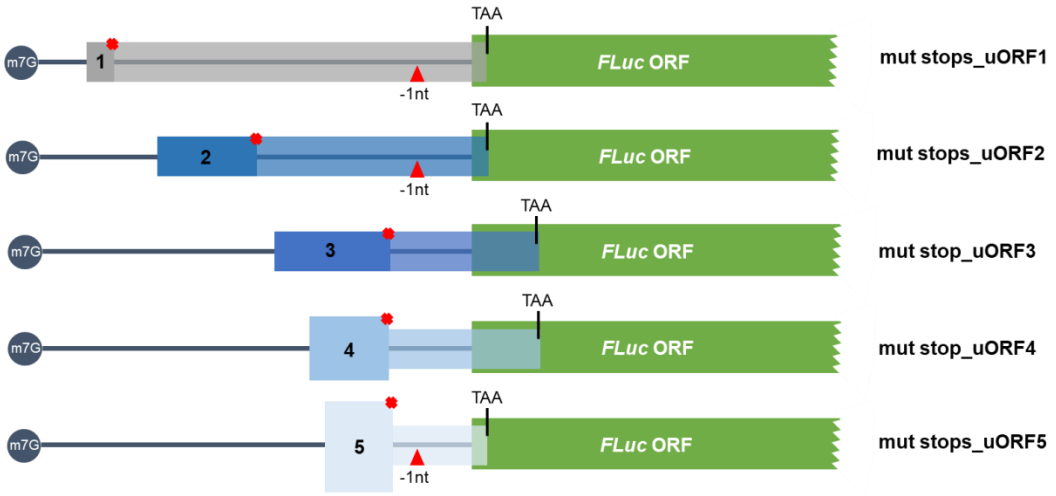

Relative luciferase (FLuc/RLuc) activity
1,2
1
0,8
0,6
0,4
0,2
0
*
*
**
no AUG uORFs
ABCE1_5'UTR
uORF1
uORF2
uORF3
uORF4
uORF5
Optimal_uAUG1
Optimal_uAUG2
Optimal_uAUG3
Optimal_uAUG4
Optimal_uAUG5
Relative luciferase (FLuc/RLuc) mRNA levels
1,6
1,4
1,2
1
0,8
0,6
0,4
0,2
0

Figure 3.6 – Translation reinitiation at the mORF is allowed after *ABCE1* AUG uORFs translation. (A) Schematic representation of the constructs used to study the occurence of translation reinitiation at the mORF after each one of the *ABCE1* AUG uORFs translation. The Kozak context of each uORF initiation codon was replaced by an optimal Kozak sequence context (in cDNA: ACCATGG; indicated by the red triangle) by applying a site-directed mutagenesis in the constructs carrying only one functional AUG uORF (uORF1-5), obtaining the constructs Optimal_uAUG1, Optimal_uAUG2, Optimal_uAUG3, Optimal_uAUG4 and Optimal_uAUG5. The uORFs represented have an UGA as termination codon. The five non-AUG uORFs are present in these constructs but not represented in the figure to simplify the scheme; **(B)** Translation reinitiation at the AUG of the FLuc reporter after *ABCE1* AUG uORFs translation. HCT116 cells were transiently transfected with each one of the constructs described in (A) or the ABCE1_5'UTR, "no AUG uORFs" or the constructs with only one functional AUG uORF (uORF1-5) already described in Figure 3.2A, and the pRL-TK plasmid. Twenty-four hours' post-transfection cells were harvested and lysed. Luciferase activity was assessed by luminometry assays (upper panel) and the corresponding mRNA levels by semi-quantitative RT-PCR (lower panel). Data were then analyzed as described in Figure 3.2B. The results are presented as mean ± SEM from at least three independent experiments. Student's t-test was applied for statistical significance: $p<0.05$ (*) and $p<0.01$ (**).

Results show that the relative luciferase activity of the Optimal_uAUG1 construct is expressed at comparable levels to those obtained from the uORF1 construct (Figure 3.6B). Thus, uORF1 start codon is extremely recognized, and only after its translation, the mORF start codon is reached by the translating ribosomes. This result demonstrates that uORF1 allows for reinitiation at the mORF initiation codon. We can not rule out events of translation reinitiation at other start codons upstream of the mORF, that are not assessed with these constructions. The same occurs for Optimal_uAUG5, which is expressed at levels comparable to the ones obtained with the uORF5 construct (Figure 3.6B), showing that reinitiation at the mAUG may occur after uORF5 translation. Both results are in agreement with the presence of a good Kozak context for uORF1 and uORF5 initiation codons. However, we observe that translation reinitiation after translation of uORF5 is not as efficient as the one that occurs after translation of uORF1, which might be correlated with the shorter intercistronic distance from the uORF5 stop codon to the mAUG. On the other hand, the Optimal_uAUG2 induces a relative luciferase activity that is ~2.2-fold smaller than the one obtained from the uORF2 (Figure 3.6B), demonstrating that reinitiation at the mAUG after uORF2 translation only accounts for approximatedly half of the relative luciferase activity obtained by the uORF2 (Figure 3.6B). This result points out that the scanning ribosomes can frequently bypass uORF2 start codon and initiate translation at the mAUG. This seems also to be the case at uORF4, since relative luciferase expressed by the Optimal_uAUG4 is ~1.8-fold smaller than the one induced by the construct uORF4 (Figure 3.6B). The relative luciferase activity of Optimal_uAUG3 construct is almost undetectable, indicating that there is no considerable translation reinitiation after uORF3 translation (Figure 3.6B). Thus, these results are in agreement with a frequent ribosomal bypass of the uORF3 initiation codon due to its weak Kozak context. Together, these results indicate that uORFs2-4 are frequently bypassed.

To unequivocally understand how much each uORF is recognized or bypassed by the translating ribosomes, we analyzed the luciferase expression of the constructs carrying each one of the uORFs out-of-frame and overlapped with the mORF. In the case of uORF1, uORF2 and uORF5, the uORF stop codon and all the other existing in-frame stop codons were mutated to a sense codon (in cDNA: TGA → GGA) and one nucleotide was deleted in the intercistronic space in order to make each one of these three uORFs out-of-frame and overlapped by 22 nts with the FLuc coding sequence (where a new in-frame stop codon is found), originating the constructs "mut stops_uORF1", "mut stops_uORF2" and "mut stops_uORF5" (Figure 3.7A). In the uORF3 and uORF4 constructs, only the common stop codon was mutated (in cDNA: TGA → GGA), obtaining out-of-frame uORFs overlapped by 83 nts with the FLuc coding sequence, originating the "mut stop_uORF3" and "mut stop_uORF4" constructs (Figure 3.7A). These constructs will only express FLuc if the uORF initiation codon is bypassed by the ribosome and the FLuc initiation codon is recognized, since translation reinitiation at the mORF AUG is completely abroggated. HCT116 cells were transiently co-transfected with each one of these constructs or with the constructs carrying only one functional AUG-uORF, the ABCE1_5'UTR or the "no AUG uORFs", and with the pRL-TK plasmid. FLuc and RLuc activity were measured by luminometry assays and the relative luciferase activity was quantified as previously described. The relative luciferase activity of these constructs is compared to that of the corresponding construct with only one functional AUG uORF. FLuc and RLuc mRNA levels were quantified by semi-quantitative RT-PCR and the relative mRNA levels determined as before, and it is shown that there were no considerable alterations in the relative luciferase mRNA levels when these constructs were transfected in HCT116 cells (Figure 3.7B).

A


TAA
m7G
1
-1nt
FLuc ORF
mut stops_uORF1
TAA
m7G
2
-1nt
FLuc ORF
mut stops_uORF2
TAA
m7G
3
FLuc ORF
mut stop_uORF3
TAA
m7G
4
FLuc ORF
mut stop_uORF4
TAA
m7G
5
-1nt
FLuc ORF
mut stops_uORF5

B

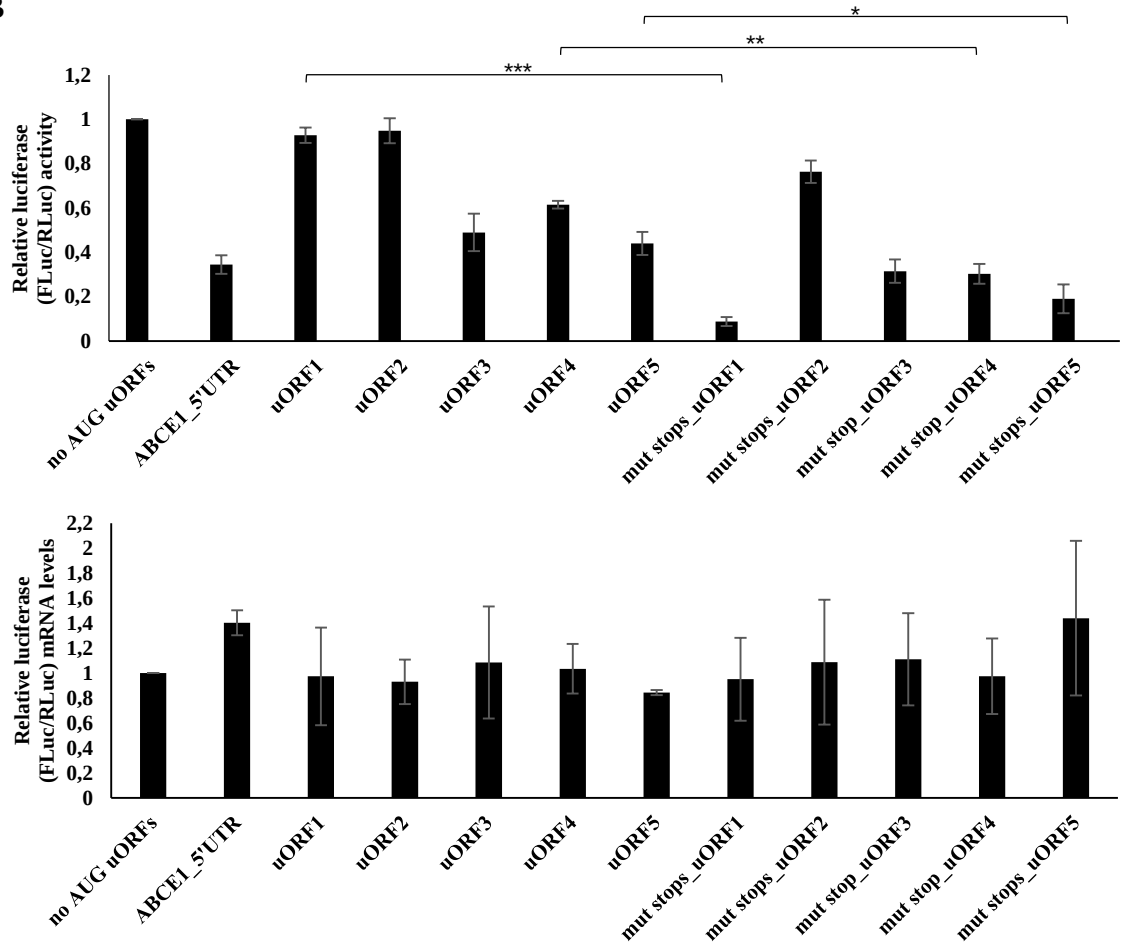

*
**
***
Relative luciferase (FLuc/RLuc) activity
1,2
1
0,8
0,6
0,4
0,2
0
no AUG uORFs
ABCE1_5'UTR
uORF1
uORF2
uORF3
uORF4
uORF5
mut stops_uORF1
mut stops_uORF2
mut stop_uORF3
mut stop_uORF4
mut stops_uORF5
Relative luciferase (FLuc/RLuc) mRNA levels
2,2
2
1,8
1,6
1,4
1,2
1
0,8
0,6
0,4
0,2
0
no AUG uORFs
ABCE1_5'UTR
uORF1
uORF2
uORF3
uORF4
uORF5
mut stops_uORF1
mut stops_uORF2
mut stop_uORF3
mut stop_uORF4
mut stops_uORF5

Figure 3.7 – Ribosomal bypass of the *ABCE1* AUG uORFs allows mORF translation. (A) Schematic representation of the constructs used to study the ribossomal bypass of each one of the *ABCE1* AUG uORFs. Series of sequential mutagenesis were performed to obtain out-of-frame and overlapped uORFs with the mORF. In the uORF1, uORF2 and uORF5 constructs, 1 nt (T) was deleted in the intercistronic space and, for all the five AUG uORF-individual constructs (uORF1-5), each in-frame stop codon was mutated (in cDNA: TGA → GGA), resulting the constructs "mut stops_uORF1", "mut stops_uORF2", "mut stop_uORF3", "mut stop_uORF4" and "mut stops_uORF5". The red crosses represent the point mutations at the uORF's stop codons and the red triangle indicates the deletion of 1 nt. The five non-AUG uORFs are present in these constructs but not represented in the figure to simplify the scheme; **(B)** The *ABCE1* uAUG initiation codons are bypassed by the scanning ribosomes and translation initiates at the AUG of the FLuc reporter. HCT116 cells were transiently transfected with each one of the constructs described in (A), or the ABCE1_5'UTR, "no AUG uORFs" or the constructs with only one functional AUG uORF (uORF1-5) already described in Figure 3.2A, and the pRL-TK plasmid. Twenty-four hours' post-transfection cells were harvested and lysed. Luciferase activity was assessed by luminometry assays (upper panel) and the corresponding mRNA levels by semi-quantitative RT-PCR (lower panel). Data were then analyzed as described in Figure 3.2B. The results are presented as mean ± SEM from at least three independent experiments. Student's t-test was applied for statistical significance: $p<0.05$ (*), $p<0.01$ (**) and $p<0.001$ (***).

Transfection of HCT116 cells with "mut stops_uORF1" construct induced almost no detectable relative luciferase activity when in comparison to the expression induced by uORF1 (Figure 3.7B), which demonstrates that native uORF1 is not usually bypassed by the translating ribosomes. This result demonstrates once more that the good context of uORF1 start codon makes it frequently recognized by the scanning ribosomes, thus allowing an efficiently translation of uORF1. This result is in agreement with the occurrence of translation reinitiation at the mORF after uORF1 translation as stated above (Figure 3.6B). Also, the "mut stops_uORF5" induces a lower relative luciferase activity (~2.3-fold) when comparing to uORF5 (Figure 3.7B), indicating that only a few ribosomes can bypass uORF5 start codon. This result demonstrates again that uORF5 initiation codon functions as a ribosomal barrier due to its good Kozak context that makes it extremely recognized, and that most of the translation at the mAUG will only occur by reinitiation after uORF5 translation, confirming the results obtained with the Optimal_uAUG5 construct (Figure 3.6B). On the other hand, the "mut stops_uORF2" is expressed at high levels and induces almost the same relative luciferase activity (~1.2-fold) when comparing to that from uORF2 (Figure 3.7B). This demonstrates that, in the uORF2 construct, approximately 80% of the relative luciferase activity is due to ribosomal bypass of the uAUG2. This is also observed in cells transfected with the Fused_uORF2_AUG, where the protein corresponding to the uORF-FLuc fusion is slightly less expressed than the native FLuc protein (lane 6 in Figure 3.5B). The relative luciferase activity levels obtained from the "mut stop_uORF3" is slightly reduced (~1.6-fold) comparing to the relative FLuc expressed by uORF3 (Figure 3.7B). This highlights that uORF3 is, although to a lesser extent than uORF2, bypassed by the scanning ribosomes, as a result of its weaker Kozak context. Yet, this mechanism is only responsible for approximately 64% of the relative luciferase activity induced by uORF3. The remaining FLuc activity might be due the little translation reinitiation at the mORF after uORF3 translation. Moreover, "mut stop_uORF4"

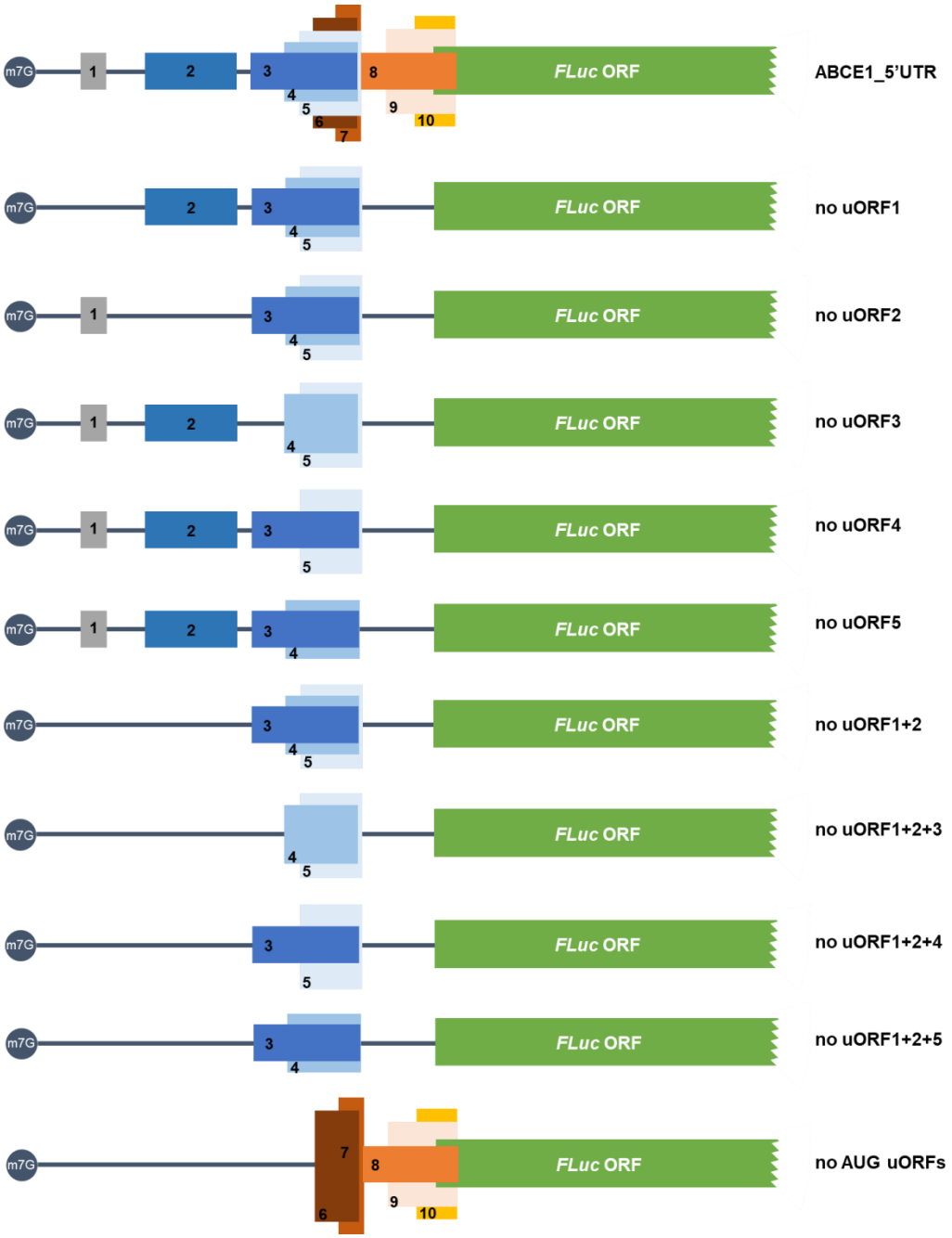
induces approximatedly half of the relative luciferase activity expressed by the uORF4 construct (Figure 3.7B). This indicates that about 50% of the scanning ribosomes can bypass uAUG4 and may initiate at the mORF start codon.

In summary, in the human *ABCE1* 5'-leader sequence there are five functional AUG uORFs. While uORF1 and uORF5 have the potential to be efficiently recognized and translated, uORF2 and uORF3 are most frequently bypassed by the scanning ribosomes and, in uORF4, both mechanisms (translation reinitiation and ribosomal bypass) can occur.

3.6. *ABCE1* uORF3 and uORF5 are competent translational repressors

To better understand how *ABCE1* AUG uORFs function together in the native conformation of the 5'-leader sequence, we used a site-directed mutagenesis strategy to mutate the ABCE1_5'UTR plasmid in a way to obtain only one AUG uORF start codon mutated (in cDNA: ATG → TTG). The clones obtained were named "no uORF1", "no uORF2", "no uORF3", "no uORF4" and "no uORF5" (Figure 3.8A). In addition, we also cloned constructs with different combinations of mutated and functional uORFs, where uORF1 and uORF2 initiation codons are always deleted ("no uORF1+2", "no uORF1+2+3", "no uORF1+2+4" and "no uORF1+2+5" constructs; Figure 3.8A). HCT116 cells were then transiently co-transfected with these constructs or with the controls ABCE1_5'UTR and "no AUG uORFs", and the pRL-TK control plasmid. Their regulatory effect was assessed by luminometry assays and the mRNA levels were quantified by semi-quantitative RT-PCR, as described before. No considerable changes in the relative luciferase mRNA levels were observed in HCT116 cells transfected with all these constructs (Figure 3.8B).

A


m7G
1
2
3
4
5
6
7
8
9
10
FLuc ORF
ABCE1_5'UTR
no uORF1
no uORF2
no uORF3
no uORF4
no uORF5
no uORF1+2
no uORF1+2+3
no uORF1+2+4
no uORF1+2+5
no AUG uORFs

B

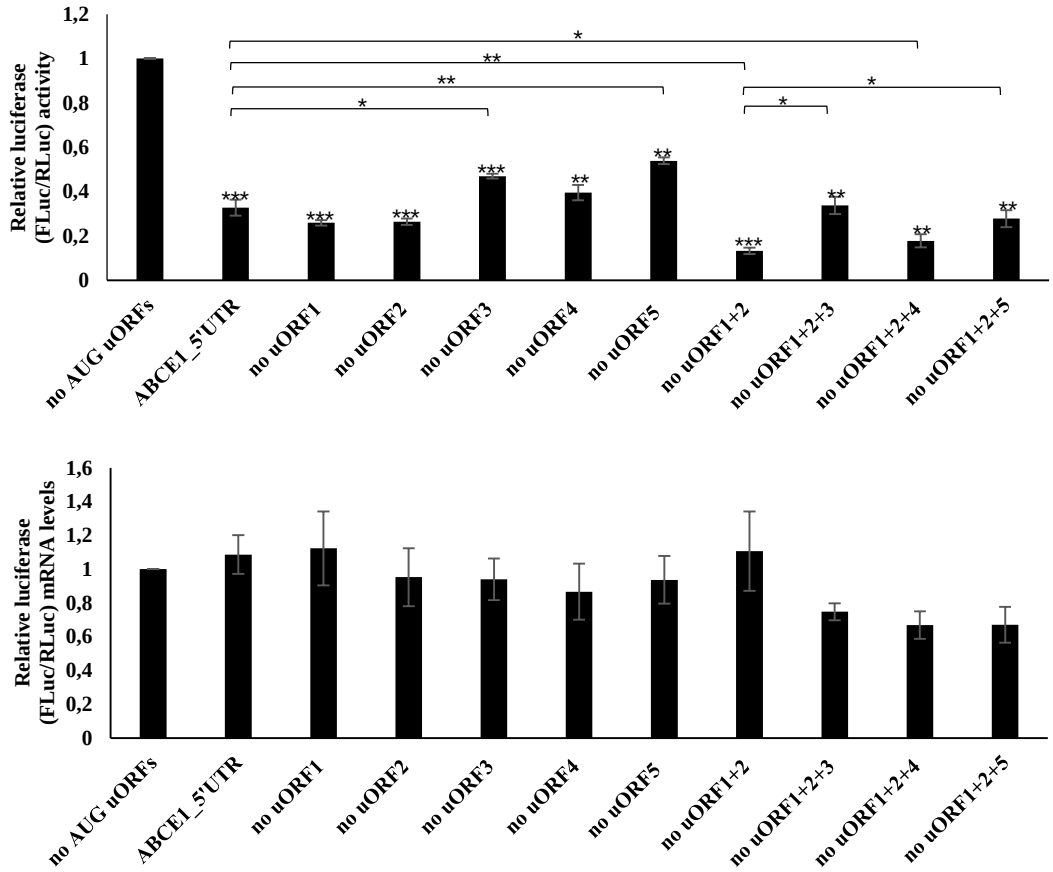


Figure 3.8 – The *ABCE1* uORF3 and uORF5 are strong translational repressors of the mORF expression. (A) Schematic representation of the constructs used to study how the AUG uORFs function in the native configuration of the 5'-leader sequence of the *ABCE1* mRNA. The ABCE1_5'UTR construct, containing the intact *ABCE1* 5'-leader sequence, and the "no AUG uORFs" construct, without functional AUG uORFs, were already described in Figure 3.2A. The construct ABCE1_5'UTR was used to mutate only one AUG start codon (in cDNA: ATG → TTG), obtaining the constructs "no uORF1", "no uORF2", "no uORF3", "no uORF4" and "no uORF5". Additionally, different combinations of mutated and functional uORFs were obtained in constructs where uORF1 and uORF2' initiation codons were always mutated, named "no uORF1+2", "no uORF1+2+3", "no uORF1+2+4" and "no uORF1+2+5" constructs. The uORFs represented have an UGA as termination codon. The five non-AUG uORFs are present in these constructs but not represented in the figure to simplify the scheme; **(B)** uORF3 and uORF5 significantly repress FLuc reporter expression. HCT116 cells were transiently transfected with each one of the constructs described in (A) and the pRL-TK plasmid. Twenty-four hours' post-transfection cells were harvested and lysed. Luciferase activity was assessed by luminometry assays (upper panel) and the corresponding mRNA levels by semi-quantitative RT-PCR (lower panel). Data were then analyzed as described in Figure 3.2B. The results are presented as mean ± SEM from at least three independent experiments. Student's t-test was applied for statistical significance: $p<0.05$ (*), $p<0.01$ (**) and $p<0.001$ (***).

In Figure 3.8B, a general significant decrease in relative luciferase activity is shown for all the constructs tested when compared to that of the "no AUG uORFs" control. In comparison to the ABCE1_5'UTR expression, the relative luciferase activity seems to be equally maintained in the constructs carrying only the mutated uAUG1, uAUG2 or uAUG4 start codon ("no uORF1", "no uORF2" or "no uORF4", respectively; Figure 3.8B). These results demonstrate that uORF1, uORF2 and uORF4, are devoid of inhibitory activity, or when absent, their

function is compensated by the remaining uORFs. The "no uORF3" and "no uORF5" constructs, however, present a modest, but significant, increase in relative luciferase activity (~1.4-fold and ~1.6-fold, respectively; Figure 3.8B). This shows that uORF3 and uORF5 are the most inhibitory uORFs of the *ABCE1* mRNA, as shown before (Figure 3.2B). Furthermore, the significant increase in relative luciferase activity observed when uORF5 start codon is inactive demonstrates that uORF3 is not able to totally overcome the lack of uORF5. If the uORF3 start codon is mutated, we also observe a significant increase of relative luciferase activity comparing to that of the ABCE1_5'UTR construct, indicating that uORF5 is not able to completely replace the function of uORF3. However, it needs to be noticed that even imparing the function of one of the inhibitory uORFs, the constructs still show a significant translational repression at the mAUG. These data indicate that those inhibitory uORFs actually function in a fail-safe manner to maintain the expression levels of their mORF at a minimum. Additionally, it seems that the level of repression achieved by the intact *ABCE1* 5'-leader sequence (evaluated by the ABCE1_5'UTR construct), is only accomplished if uORF3 and uORF5 are acting together (additive effect).

Moreover, when both uORF1 and uORF2 start codons are simultaneously eliminated (construct "no uORF1+2"), a ~2.5-fold decrease in relative luciferase activity is observed in HCT116 cells comparing to the effect of ABCE1_5'UTR (Figure 3.8B). Although devoid of inhibitory activity, we know from the results shown previously that the initiation codon of uORF1 is efficiently recognized by the translating ribosomes (Figure 3.5). The results now obtained from construct "no uORF1+2" are also compatible with the conclusion that uORF1 (and in a very less extent, uORF2) is frequently recognized by the translating ribosomes making it a barrier for ribosomal scanning. Thus, when both, uAUG1 and uAUG2 (but mainly the uAUG1) are not present, the ribosome no longer stops at these start codons and instead continue scanning until it reaches another start codon. We hypothesize that without uORF1 (and/or uORF2) start codon, more ribosomes can rapidly recognize the start codons of the inhibitory uORF3 and/or uORF5, and this event may enhance their repressive function, explaining this high level of translational repression induced by the "no uORF1+2" construct. The relative luciferase activity observed from the constructs "no uORF1+2+3" and "no uORF1+2+5" is, respectively, ~2.5-fold and ~2.1-fold higher than the one of the construct "no uORF1+2" (Figure 3.8B). This significant increase in luciferase level demonstrates once more that when one of the most inhibitory uORFs (uORF3 and uORF5) is eliminated the other one compensates its function, acting in a fail-safe manner, and maintaining the translational repression of the mORF. The construct "no uORF1+2+4" shows the same relative luciferase activity as the one

obtained from the "no uORF1+2" construct (Figure 3.8B), demonstrating again that uORF4 does not contribute significantly to the repressive effect promoted by the 5'-leader sequence of the *ABCE1* transcript.

Together, these results demonstrate that uORF1 and uORF2 are non-repressive uORFs that function as ribosomal barriers for the recognition of the inhibitory uORF start codons and that uORF3 and uORF5 are competent translational repressors, acting in combination to maintain ABCE1 protein at physiological levels.

3.7. *ABCE1* uORF-mediated translational regulation is maintained during thapsigargin-induced endoplasmic reticulum stress

It is well described the impact of uORFs in mediating translational regulation during stress conditions, where several transcripts escape the overall translational repression by a mechanism depending on uORFs in order to solve the stress and promote cell survival [1]. Thus, here our aim was to understand if the AUG uORFs are responsible for the regulation of *ABCE1* expression in conditions of global protein synthesis impairment. To pursue this goal, we chose thapsigargin (Tg), a well characterized inducer of ER stress. Thapsigargin induces ER stress by inhibiting the sarcoplasmic/ER Ca^{2+} ATPases (SERCA) leading to a decrease of the calcium storage in the ER lumen. Calcium-dependent ER chaperones will no longer function properly and consequently unfolded proteins start to accumulate, which in the end triggers the unfolded protein response (UPR) pathway to alleviate the stress [156, 157].

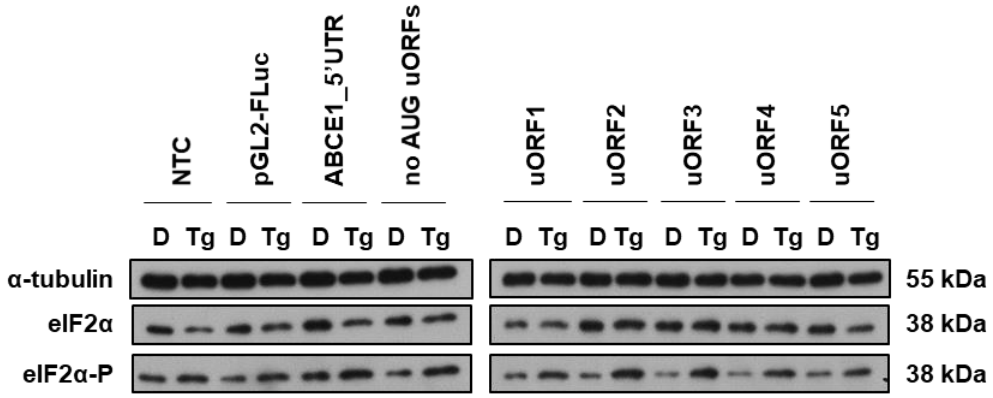
Thus, to study the uORF-mediated translational regulation in conditions of thapsigargin-induced stress, HCT116 cells were transiently co-transfected with the constructs "no AUG uORFs", ABCE1_5'UTR, or the reporter vectors carrying only one functional AUG start codon (uORF1-5) and the pRL-TK control plasmid (Figure 3.2A), and 4h after transfection, cells were treated with 4 μM of Tg, or the veichle control (DMSO; 'normal' condition), for 24h. Stress induction by thapsigargin was, first of all, assessed by immunoblotting of the eIF2α-P protein levels. In stress conditions, as for instance ER stress, eIF2α is phosphorylated by serine-threonine kinases, an event that impairs GDP to GTP exchange by eIF2B, and thus the formation of an active form of eIF2α necessary for the assembly of new ternary complexes, greatly inhibiting global translation [59–61]. In the Western blot of Figure 3.9A, eIF2α-P protein levels are increased in the condition of thapsigargin-treated cells compared to the cells treated with the vehicle, DMSO, indicating that cells are under stress. The effect of thapsigargin in the uORF-mediated translational regulation of *ABCE1* mRNA was then measured by luminometry assays and is shown as relative luciferase activity after normalizing to the effect

of the “no AUG uORFs” in the corresponding condition. FLuc mRNA levels were quantified by semi-quantitative RT-PCR, as previously described, and it was not shown a significant alteration in the relative mRNA levels in each condition.

In Figure 3.9B, it is observed that the relative luciferase activity obtained from the ABCE1_5’UTR construct, in comparison to the one obtained from the “no AUG uORFs”, is maintained without significant changes in thapsigargin- vs DMSO-treated cells. The same result was observed for the constructs uORF1 to uORF5, where the expression levels do not change under thapsigargin-induced ER stress conditions.

From these results we can conclude that the AUG uORFs within the 5’-leader sequence of *ABCE1* mRNA do not derepress mORF expression in conditions of global translation inhibition, at least towards ER stress induced by thapsigargin. Thus, it suggests that ABCE1 is not a stress-responsive protein, and instead probably it should be kept at constant levels, either in stress as in normal conditions.

A



B

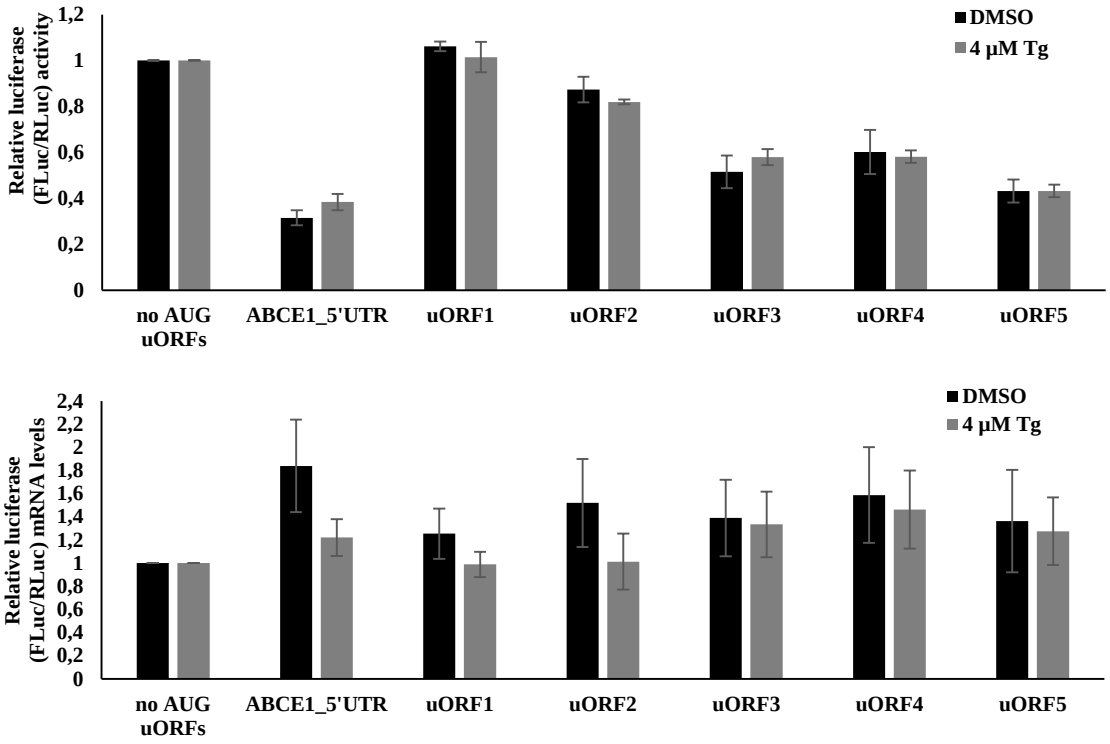


Figure 3.9 – *ABCE1* AUG uORF-mediated translational repression is maintained under ER stress induced by thapsigargin. HCT116 cells were transiently co-transfected with each one of the constructs described in Figure 3.2A and the pRL-TK plasmid. Four hours' post-transfection cells were treated with 4 µM of thapsigargin (Tg; grey bars) or DMSO (vehicle control of Tg; black bars) for 24h and then harvested and lysed; **(A)** The ER stress induced by Tg was assessed by Western blot using specific antibodies against eIF2α-P and the counterpart, eIF2α. α-tubulin was used as loading control. These are representative immunoblots from at least three indepedent experiments. NTC – non-template control; **(B)** Luciferase activity was assessed by luminometry assays (upper panel) and the corresponding mRNA levels by semi-quantitative RT-PCR (lower panel). Data were then analyzed as described in Figure 3.2B. The results are presented as mean ± SEM from at least three independent experiments.

3.8. The *ABCE1* transcript is not an NMD-target

We hypothesized that another possible role for the uORFs within the *ABCE1* 5'-leader sequence is to keep the mORF expression levels at a minimum by triggering NMD (Figure 1.2A). NMD is characterized as a mRNA surveillance mechanism that rapidly degrades transcripts with a PTC, that otherwise would lead to truncated proteins with a potential toxic effect for the cell [55–57]. In some cases, mRNAs containing uORFs fit this criterium, since uORF termination codons can be recognized as PTCs, thus reducing the stability of the transcript and therefore reducing the levels of the corresponding protein [2, 3, 55]. To study if the presence of uORFs within the 5'-leader sequence of *ABCE1* mRNA triggers NMD, the half-life of the *ABCE1* transcript was determined in control conditions or in NMD-inactivated cells, based on the knowledge that true NMD-targeted transcripts, including those with uORFs, are upregulated when NMD is inhibited [55]. Briefly, HeLa cells (a good model for NMD studies) were transfected with siRNAs for the central NMD factor *UPF1* (NMD-inactivated cells) and *LUC* (NMD-competent cells) and then treated with 60 µM of DRB to inhibit transcription, and cells were harvested at different time points (0 h, 0.5h, 1h and 2h). Then, total RNA was isolated and *ABCE1* mRNA levels were quantified by RT-qPCR in both tested conditions. Additionally, and knowing that the *GADD45A* mRNA is a natural NMD-target [158], its mRNA levels were also quantified to function as a positive control for NMD.

To control knockdown efficiency, the protein levels of UPF1 were assessed by Western blot. Results show an efficient knockdown of UPF1 in *UPF1* siRNA-transfected cells when compared to cells treated with the *LUC* siRNA (Figure 3.10A). In *UPF1* siRNA-knockdown cells, *ABCE1* transcript shows a shorter half-life than in *LUC* siRNA-knockdown cells (Figure 3.10B). Indeed, the half-life of the *ABCE1* transcript in *UPF1* siRNA-tranfected cells is 2.05h, which is half of that obtained in the *LUC* siRNA-treated cells (4.2h). On the contrary, the *GADD45A* transcript, in conditions of UPF1 knockdown has a half-life of 1.53h, which is significantly higher than in *LUC* siRNA-treated cells (0.89h; Figure 3.10B). Together, these results show that *GADD45A* transcript is an NMD-target, as expected, but *ABCE1* mRNA is not.

These results demonstrate that the termination codons of the *ABCE1* uORFs are not recognized as PTCs and thus the existence of these uORFs in the *ABCE1* 5'-leader sequence do not induce NMD.

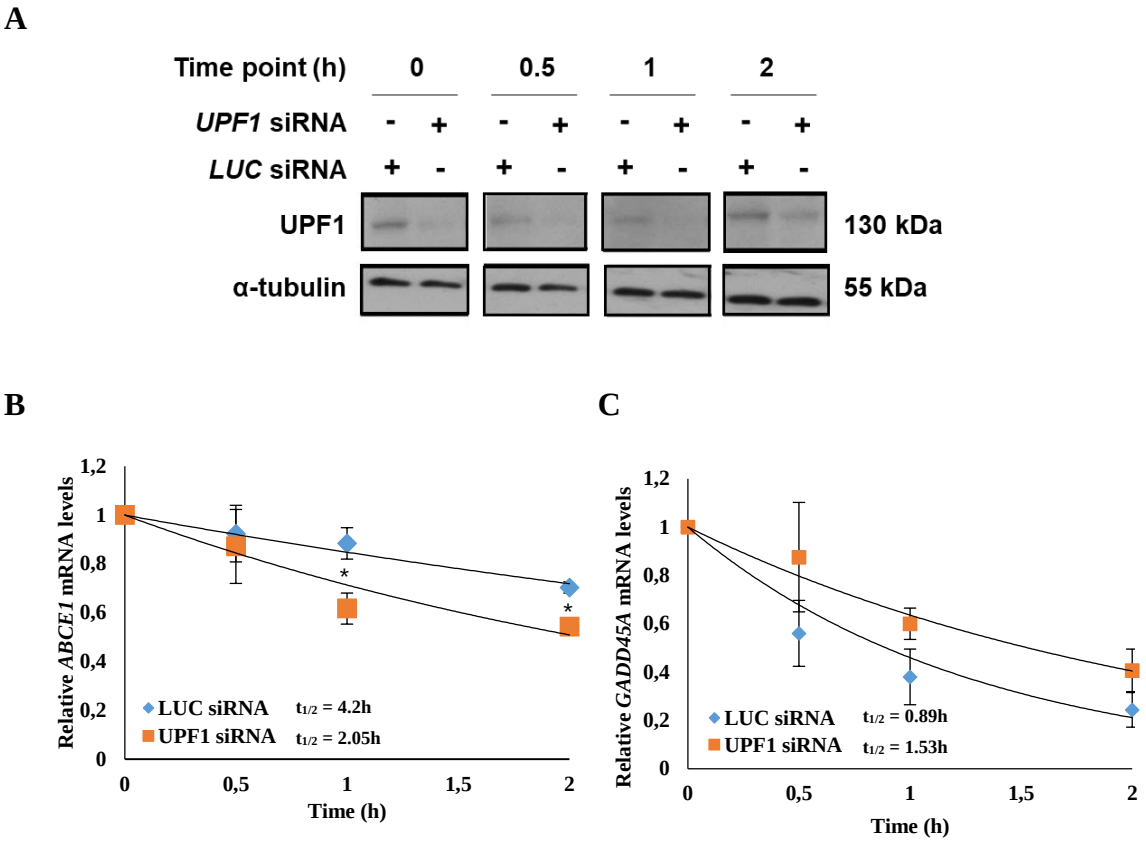


Figure 3.10 – The human *ABCE1* mRNA is not an NMD-target. HeLa cells were transfected with (+) or without (-) siRNAs for *UPF1* or *LUC* and treated with 60 µM of adenosine analogue 5,6-dichloro-1-3-D-ribofuranosylbenzimidazole (DRB) to inhibit transcription. Cells were then harvested at different time points (0, 0.5, 1 and 2h) and lysed. **(A)** The siRNA knockdown efficiency was assessed by Western blot using a specific antibody against UPF1. α-tubulin was used as loading control. These are representative immunoblots from at least three indepedent experiments; **(B)** mRNA levels were determined by RT-qPCR. The relative mRNA expression levels were calculated by applying the comparative Ct method (ΔΔCt): *ABCE1* and *GADD45A* (targets) mRNA levels were normalized to the ones of *GAPDH* at each time point for each transfected condition, then normalized to the ratio target/*GAPDH* of the time point 0h for *LUC* siRNA transfection (arbitrarely defined as 1). The mRNA half-life of *ABCE1* and *GADD45A* were calculated by an exponential defined from the relation between the relative mRNA levels of the target transcripts and the time exposed to DRB. The results are presented as mean ± SEM from at least three independent experiments. Student's t-test was applied for statistical significance: $p<0.05$ (*).

3.9. The *ABCE1* uORFs are devoid of oncogenic function

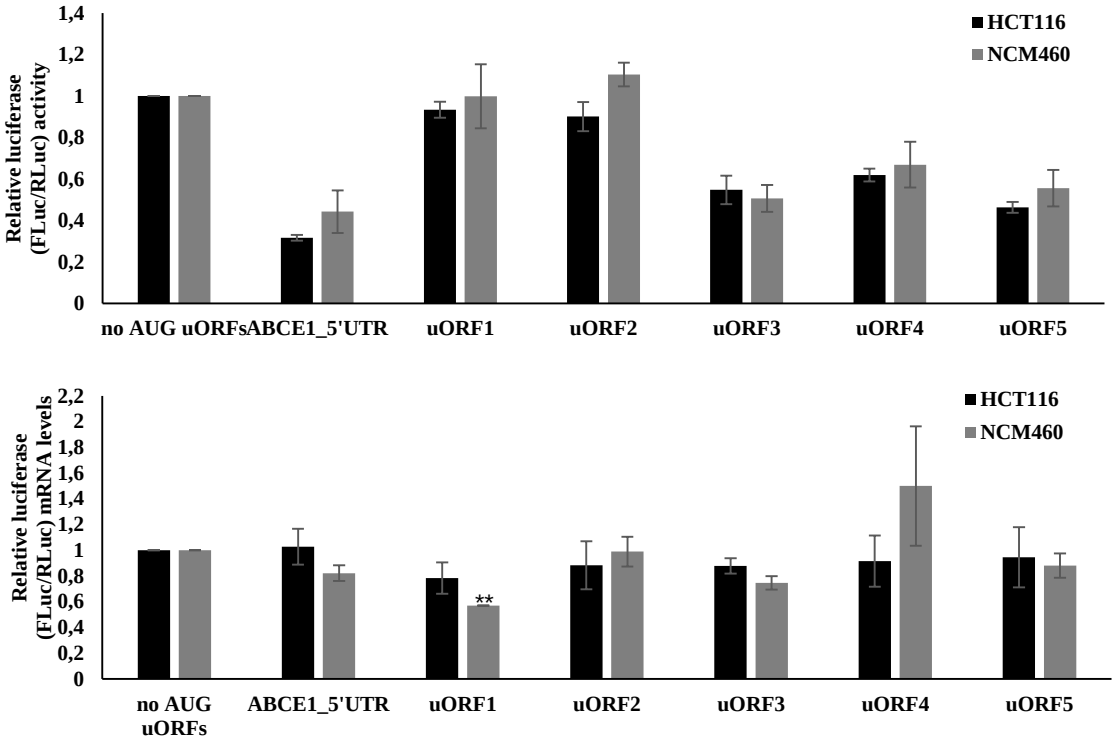
Taking into account the above results about the regulatory function of the *ABCE1* AUG uORFs in the colorectal cancer cell line HCT116, we wanted to investigate if the *ABCE1* uORF-mediated translational control plays a function in non-neoplasic vs cancerous colorectal cells, and thus imply those uORFs and also the ABCE1 transcript/protein in the tumorigenesis of colorectal cells. To accomplished that, we co-transfected the constructs "no AUG uORFs", ABCE1_5'UTR or the reporter vectors with only one functional AUG uORF (uORF1-5) and the pRL-TK control plasmid, (Figure 3.2A), in HCT116 cells and in a non-tumorigenic colorectal cell line, NCM460. Luminometry assays were used to quantify the FLuc and RLuc

activity and relative FLuc/RLuc activity of each transfection condition was normalized to that of the "no AUG uORFs" construct. The corresponding FLuc and RLuc mRNA levels were quantified by semi-quantitative RT-PCR. Relative FLuc/RLuc mRNA levels of each transfection condition was normalized to that of the "no AUG uORFs" construct and no considerable alterations in the relative mRNA levels were observed, except for uORF1 construct in NCM460 cells (Figure 3.11A).

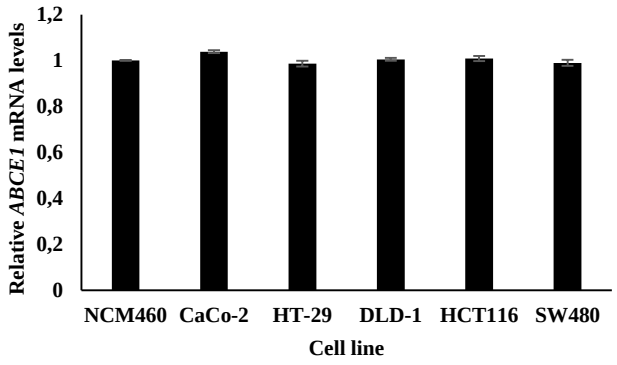
As shown in Figure 3.11A, in NCM460 cells, the ABCE1_5'UTR construct induces a ~2.3-fold repression in the relative luciferase activity when in comparison to the one obtained from the "no AUG uORFs" construct. This level of repression is comparable to the one already demonstrated in HCT116 cells (Figure 3.11A), thus showing a similar pattern of translational regulation exerted by the *ABCE1* uORFs independent on the cell type. Both uORF1 and uORF2 did not induce a significant decrease in the relative luciferase activity (Figure 3.11A), once again demonstrating that both uORFs are devoid of repression ability, as observed in HCT116 cells. On the contrary, a repression of ~2-fold by uORF3, ~1.5-fold by uORF4 and ~1.8-fold by uORF5 is observed. Results do not significantly change from that obtained in HCT116 cells (Figure 3.11A). These results demonstrate that the AUG uORFs of *ABCE1* mRNA equally regulate translation in non-tumorigenic and cancer colorectal cells, which is consistent with lack of oncogenic function.

We further investigated the lack of oncogenic activity of *ABCE1,* by determining its mRNA and protein levels by RT-qPCR and Western blot, respectively, in several colorectal cancer cell lines CaCo-2, HT-29, DLD-1, HCT116 and SW480, in comparison to the levels in the non-tumorigenic colorectal cell line NCM460. As shown in Figure 3.11B and C, respectively the *ABCE1* mRNA and protein levels are maintained unaffected in all the colorectal cancer cell lines tested, and are at comparable levels to those in NCM460 cells. These results discourage the association of ABCE1 to a functional role in the tumorigenic process of colorectal cancer.

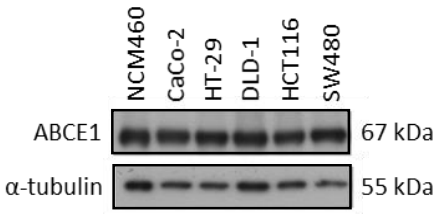
A


Relative luciferase (FLuc/RLuc) activity
1,4
1,2
1
0,8
0,6
0,4
0,2
0
HCT116
NCM460
no AUG uORFs
ABCE1_5'UTR
uORF1
uORF2
uORF3
uORF4
uORF5
Relative luciferase (FLuc/RLuc) mRNA levels
2,2
2
1,8
1,6
1,4
1,2
1
0,8
0,6
0,4
0,2
0
**
HCT116
NCM460
no AUG uORFs
ABCE1_5'UTR
uORF1
uORF2
uORF3
uORF4
uORF5

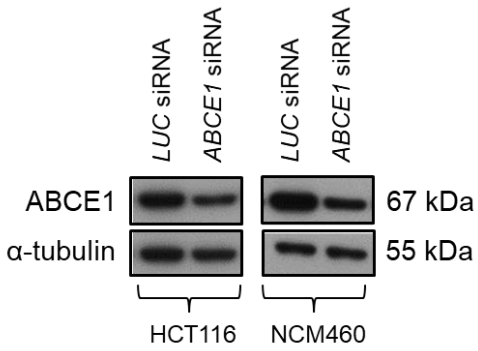
B


Relative ABCE1 mRNA levels
1,2
1
0,8
0,6
0,4
0,2
0
NCM460
CaCo-2
HT-29
DLD-1
HCT116
SW480
Cell line

C


NCM460
CaCo-2
HT-29
DLD-1
HCT116
SW480
ABCE1
67 kDa
α-tubulin
55 kDa

D


LUC siRNA
ABCE1 siRNA
LUC siRNA
ABCE1 siRNA
ABCE1
67 kDa
α-tubulin
55 kDa
HCT116
NCM460

E

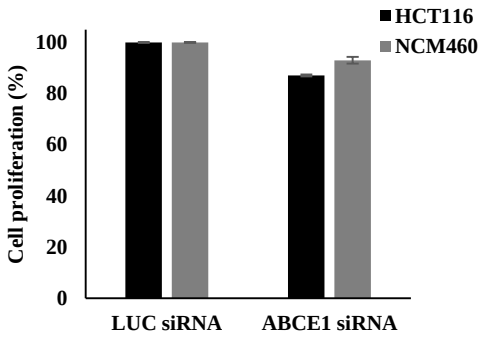

Cell proliferation (%)
100
80
60
40
20
0
HCT116
NCM460
LUC siRNA
ABCE1 siRNA

Figure 3.11 – ABCE1 and its AUG uORFs do not have an oncogenic role. (A) NMC460 and HCT116 cells were transiently co-transfected with each one of the constructs described in Figure 3.2A, and the pRL-TK plasmid. Twenty-four hours' post-transfection cells were harvested and lysed. Luciferase activity was assessed by luminometry assays (upper panel) and the corresponding mRNA levels by semi-quantitative RT-PCR (lower panel). Data were then analyzed as described in Figure 3.2B. The data from HCT116 cells are the one displaying in Figure 3.2B. The results are presented as mean ± SEM from at least three independent experiments. Student's t-test was applied for statistical significance: $p<0.01$ (**); The *ABCE1* mRNA levels **(B)** and the ABCE1 protein levels **(C**) were evaluated in the colorectal cancer cells lines CaCo-2, HT-29, DLD-1, HCT116 and SW480, as well as, in the non-tumorigenic cell line NCM460. The *ABCE1* mRNA levels were assayed by RT-qPCR and presented as the normalization of the mRNA levels of *ABCE1* to the ones of the internal control, *GAPDH*, from each colorectal cell line and nomalized to the ratio *ABCE1*/*GAPDH* from NCM460 cells. The results are presented as mean ± SEM from at least three independent experiments. The protein levels were obtained by Western blot using a specific antibody against ABCE1. α-tubulin was used as loading control. This is a representative immunoblot from at least three indepedent experiments; ABCE1 knockdown does not induce a decrease in cell viability in HCT116 and NCM460 cells: **(D)** The *ABCE1* siRNA knockdown efficiency was assessed by Western blot using a specific antibody against ABCE1. α-tubulin was used as loading control. These are representative immunoblots from at least three indepedent experiments; **(E)** Cell proliferation was measured using 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT). A linear relation between the absorbance signal and the number of viable cells was used to determine the rate of cell proliferation in *LUC* and *ABCE1* siRNA transfected cells. The rate of cell proliferation (%) was measured as the ratio of cell proliferation in *ABCE1* siRNA-transfected cells vs *LUC* siRNA-transfected cells for each cell line. The results are presented as mean ± SEM from at least three independent experiments.

We also evaluated the proliferation, and thus cell viability rate, of HCT116 and NCM460 cells after ABCE1 knockdown (using a siRNA for *LUC* as a transfected/reference control). For that, we used a well stablished method based on the reduction of a yellow tetrazolium salt known as MTT. This method implies that metabolic active cells (that are the viable cells) have the ability to reduce MTT by the action of dehydrogenase enzymes, resulting in the formation of intracellular purple water-insoluble formazan cristals. After solubilization in DMSO, the formazan dye is quantified in a spectrophotometer at 570 nm. The absorbance measured is directly proportional to the number of proliferating/viable cells, which makes a suitable assay for measuring alterations in the rate of cell proliferation. ABCE1 knockdown was able to downregulate in approximately 40% the ABCE1 expression levels in both cell lines as assessed by Western blot (Figure 3.11D). At these conditions, cell proliferation rate was then quantified after establishing a linear relation between the absorbance signal and the number of viable cells. The rate of cell proliferation (%) was measured as the ratio of cell proliferation in *ABCE1* siRNA-transfected cells vs *LUC* siRNA-transfected cells for each cell line. As shown in Figure 3.11E, *ABCE1* depletion in both cell lines does not significantly reduce cell proliferation levels when compared to the effect promoted by the *LUC* siRNA, maintaining about 90% of cell viability in cells depleted of *ABCE1*. The fact that there is no difference in the cell viability rate between HCT116 and NCM460 cells after *ABCE1* depletion, indicates that *ABCE1* does not have an oncogenic role in the colorectal cancer cell line HCT116.

In summary, *ABCE1* uORF-mediated translational regulation is maintained in non-tumorigenic and cancerous colorectal cells, which is consistent with a lack of oncogenic functions by ABCE1 in the colorectal cancer cell line tested.

4. Discussion and Future Perspectives

mRNA translation is tightly controlled during its initiation step by the presence of *cis*-acting regulatory elements within the 5'-leader sequence of the transcripts [2]. Out of these elements, we highlight the uORFs, which are potentially translated small ORFs present in about half of the human transcriptome [33, 35]. uORFs are typically described as repressors of the downstream mORF translation under physiological conditions [2, 33, 36, 46]. However, translation of a uORF has been reported to potentiate mORF translation during stress, in order to allow the translation of stress-responsive transcripts that can cope with stress [1]. Thus, knowing the mechanisms associated with uORF translational regulation is of the utmost importance, so that we can understand how its deregulation can lead to the onset of several diseases, including cancer.

The human *ABCE1* 5'-leader sequence is extremely complex due to the presence of ten uORFs — starting with either AUG or non-AUG initiation codons — of unknown function, and therefore we asked whether those uORFs play a role in *ABCE1* post-transcriptional regulation. Here, we show that the whole 5'-leader sequence of the *ABCE1* mRNA is able to significantly repress the downstream reporter mORF expression in the colorectal cancer cell line HCT116 (Figure 3.2B and Figure 3.3B). Indeed, the uORFs present in the *ABCE1* 5'-leader sequence promoted a translation repression of approximately 70%, which falls into the range of 30–80% repression announced by Calvo and co-workers when they studied 25 single uORF-containing transcripts [33]. The repression observed in the *ABCE1* translation occurs through three AUG-initiating uORFs — uORF3, uORF4 and uORF5 (Figure 3.2B) — that seem to be acting together (additive effect) to promote the maximum mORF repression: translational repression induced by constructs with only one functional AUG uORF is not as efficient as that induced by the fully functional 5'-leader sequence. Moreover, a fail-safe mechanism can also be hypothesized from our results, since each of the uORF3, uORF4 and/or uORF5 constructs can individually repress translation of the mORF, which means that when one of the three inhibitory uORFs is not functionally active the others compensate for it. Additionally, uORF1, uORF2, and the non-canonical uORFs within *ABCE1* 5'-leader sequence do not seem to repress mORF translation (Figure 3.2B and 3.3B).

The inhibitory activity of a uORF is usually positively correlated with: a strong Kozak uORF initiation codon context, a long distance from the transcript's 5'-end to the uORF initiation codon, the presence of multiple uORFs (additive effect), a long uORF, and/or a short intercistronic distance [33, 35, 46, 48, 49]. Thus, the repressive effect of uORF3, uORF4 and

uORF5 can be associated with them being long, the short intercistronic distance, and/or the fact that their initiation codons are located far from the 5'-end of the transcript. However, all the aforementioned uORFs, except uORF5, are not in a favorable Kozak context. Of note, uORF5 is the most repressive uORF, and is highly conserved among several species, which is an indicator of their important role in *ABCE1* translational regulation. On the other hand, the lack of repressive activity for both uORF1 and uORF2 can be associated with the following features: i) a suboptimal Kozak consensus sequence (although uORF1 has a better Kozak context than uORF2), ii) the close proximity of their AUGs to the 5'-end of the mRNA (especially for uORF1), and/or iii) the long intercistronic distance. Additionally, uORF1 is a short uORF (only 36 nts) when compared to uORF2 (132 nt-long), which is one of the longest uORFs within *ABCE1* 5'-leader sequence. However, a uORF's repression activity is a combination of some of the above-mentioned factors, and therefore, although uORF2 is long, it has other features that account for the ribosomes not being able to recognize it often.

In addition, as accounts for the AUG initiation codon, a weaker sequence context for the non-canonical start codons will decrease their recognition by the translation machinery [84, 88]. Regarding the non-AUG uORFs in the 5'-leader sequence of the human *ABCE1* mRNA, their initiation codons are generally in a weak Kozak context, which most likely turns them unrecognized by the ribosomes, thus explaining the low inhibitory activity associated with these uORFs (Figure 3.3B). Moreover, these non-AUG uORFs seem to be often bypassed by the scanning ribosomes. When the *ABCE1* non-AUG uORF start codons were individually mutated to an AUG we obtained a great inhibition of FLuc expression (Figure 3.4B). This is consistent with the fact that the AUG start codons are usually more efficiently recognized than the non-canonical ones [3]. This is the case, for instance, of the aforementioned *EPRS* transcript, which contains two uORFs within its 5'-leader sequence that start with non-canonical initiation codons. However, those uORFs are translational inhibitors of the mORF, contrary to the *ABCE1* non-AUG uORFs. Young and co-workers demonstrated that when the CUG and UUG start codons of uORF1 and uORF2, respectively, were mutated to an AUG in an optimal context, there was a downregulation of the mORF expression, which happens in physiological and stress conditions [82]. This once more demonstrates that the AUG start codon is usually more efficiently recognized than the non-AUG initiation codons, and that the latter are postulated to be frequently bypassed [3, 82]. Thus, the non-canonical nature of these *ABCE1* uORFs' start codons, allied to their poor Kozak sequence context, make them frequently bypassed by the scanning ribosomes. This increases the number of ribosomes at the mORF initiation codon, consistent with a non-repressive function. It is nowadays well recognized that RiboSeq studies

are contributing to the overall picture of the human proteome, by showing that non-AUG uORFs are in fact translated more frequently than expected [34, 42]. In addition, the RiboSeq data obtained in HCT116 cells revealed the presence of those non-AUG uORFs in the 5'-leader sequence of the human *ABCE1* transcript, which means that they are actually being translated. Thus, this demonstrates that, although less efficiently, the translating ribosomes recognize and tranlate these non-AUG uORFs as well as the *ABCE1* mORF. This is demonstrated for the already described example of *FRAT2* transcript, where it was detected a co-expression of the uORF-encoded peptide and the main protein in HEK293T cells [81]. Yet, since they do not seem to act as translational repressors of ABCE1 expression, the corresponding encoded peptides can be potentially associated with other important biological functions that need to be further investigated. For instance, it has been shown that the non-AUG-bearing uORFs within the 5'-untranslated region of the *BiP* mRNA encode small peptides that function as HLA-presented epitopes recognized by human T cells [86].

Since the *ABCE1* AUG-initiating uORFs have the potential to be translated (Figure 3.5B), and in fact, they have a major impact in the translational control of the mORF expression, we aimed to understand the mechanisms underlying uORFs' regulatory function and how they interact with each other to achieve such level of repression. As shown in Figure 3.6B and 3.7B, translation reinitiation and ribosomal bypass are two mechanisms by which ribosomes gain access to the mAUG. Regarding each one of the uORFs in an individual configuration at the *ABCE1* 5'-leader sequence, we show that translation reinitiation at the *ABCE1* mAUG is efficiently accomplished after uORF1 or uORF5 translation. In the case of uORF2, its leaky scanning accounts for almost the entire *ABCE1* mORF expression. uORF3 initiation codon is, however, less frequently bypassed by the scanning ribosomes when compared to uORF2, but when it occurs the ribosomes will initiate translation at the subsequent start codon, i.e., the mORF's or non-canonical uORFs' star codon. In addition, uORF4 is equally recognized and bypassed by the scanning ribosomes.

A short uORF and/or a long intercistronic distance are strongly associated with an efficient translation reinitiation at a downstream ORF initiation codon. A long intercistronic distance grants enough time for the formation of new active ternary complexes as the 40S small ribosomal subunit, still engaged to the mRNA after a short uORF translation, scans the downstream sequence, looking for the next initiation codon in a favorable context [3, 48, 64, 67]. For instance, being short, makes uORF1 a good candidate to promote translation reinitiation after its own translation (Figure 3.6B). This was also shown for the 3-aa-long uORF1 of the *ATF4* transcript, which is able to promote translation reinitiation at the following

downstream initiation codon, which is the *ATF4* second uORF start codon, in normal conditions [67]. Moreover, the long intercistronic distance between the uORF1 stop codon and the *ABCE1* mORF initiation codon, allows for efficient translation reinitiation at the mAUG. Thus, after uORF1 translation, the 40S small ribosomal subunit stays bound to the mRNA and will resume scanning until acquires a new active ternary complex and all the necessary initiator factors in time to start a new round of translation at a downstream initiation codon. Also, the Kozak context surrounding the uORF1 start codon is adequate for an efficient translation initiation, demonstrating that only after uORF1, the translating ribosomes gain access to the mORF initiation codon. Conversely, uORF5 does not show any features associated with an efficient translation reinitiation mechanism: it is located in close proximity to the mORF initiation codon, which reduces the intercistronic space, and it is a 78 nt-long uORF, thus with 26 codons, being longer than what is shown to be a critical uORF length for an efficient translation reinitiation [48]. However, despite lacking the above mentioned favorable-features, its initiation codon is in an adequate Kozak context increasing its recognition, which demonstrates that uORF5, acts as a potent barrier for the scanning ribosomes. This seems to indicate that in the native conformation of the *ABCE1* 5'-leader sequence, uORF5 only allows for a percentage of the ribosomes that loads onto the 5'-untranslated region to reach the mORF initiation codon, lowering its translation rate. Thus, reinitiation is still the major mechanism associated with the translational regulation of uORF5 (Figure 3.6B). In the case of uORF2, uORF3 and uORF4, the low levels of translation reinitiation at the mORF initiation codon can also be associated with them being lengthened uORFs. Moreover, the weak Kozak context of these uORF's initiation codons also corroborate this low translation reinitiation efficiency.

On the other hand, ribosomal bypass of the uORF initiation codon is promoted, to a greater extent, by its poor surrounding Kozak context, and to a lesser extent, by the close proximity of the uORF start codon to the 5'-end of the 5'-leader sequence of the transcripts [1, 3, 73, 74]. uORF2 in *ABCE1* 5'-leader sequence is frequently bypassed by the scanning ribosomes and this can be, in fact, strongly associated with its poor initiation codon's Kozak context. This ribosomal bypass of uORF2 initiation codon renders the entire mORF expression (Figure 3.7B). Of note, the uAUG2 is close to the 5'-end of the 5'-leader sequence in the *ABCE1* transcript, which is a feature that, although with minimal contribution, allows uORF2 bypass. Despite uORF1 being the closest uORF to the 5'-end cap of the *ABCE1* mRNA, its suitable uORF start codon context (together with other features mentioned above) makes it strongly recognized, preventing ribosomal bypass events. The low level of ribosomal bypass of uORF5 is also associated to its good initiation codon's Kozak context, that makes it greatly recognized by the

translation machinary. Additionally, but to a smaller extent then uORF2, uORF3 or uORF4 can also be bypassed by the translational machinery (Figure 3.7B), since their initiation codons are also in a weak Kozak context. Although the level of ribosomal bypass at uORF3 initiation codon accounts for almost the entire mORF expression, the ribosomes tend to recognize, although less efficiently, its initiation codon, explaining the low level of translation initiation at the *ABCE1* start codon. Thus, in the native configuration of *ABCE1* 5'-leader sequence we expect that after this uORF translation different outcomes can occur: i) the ribosomes can be dissociated and recycled; or ii) ribosomes can be stalled during the translation elongation or termination [1, 2].

Another possible explanation is that after translation of uORF1, in the native configuration of *ABCE1* 5'-leader sequence, the ribosome will reinitiate translation at other subsequent start codons placed upstream of the mORF initiation codon, as is the case of the inhibitory AUG-initiating uORFs in the *GCN4* transcript. In physiological conditions, *GCN4* AUG-initiating uORF1 is responsible for translation reinitiation at each of the three downstream inhibitory AUG-initiating uORFs, leading to GCN4 downregulation [36, 64]. However, ribosomes in the *ABCE1* uORF1, do not seem to act in the same way, since abolishing uORF1 translation, by mutating its initiation codon, results in a similar repression level to the one performed by the intact *ABCE1* 5'-leader sequence (Figure 3.8B). Additionally, impairing uORF2 translation has the same outcome as the intact 5'-leader sequence (Figure 3.8B). These results are thus consistent with a non-repressive function of uORF1 and uORF2. Yet, being translated, uORF1 (and to a smaller extent uORF2) can work as a barrier that prevents the ribosomes from promptly reaching the inhibitory uORFs.

Out of the inhibitory uORFs, uORF3 and uORF5 seem to cooperate for the maximum repression (addictive effect), as shown in Figure 3.8B. We can also postulate that the ribosomes that bypass uAUG3 can initiate translation at uAUG5, thus enhancing the repressive potential of each other. Additionally, uORF3 and uORF5 can operate in a fail-safe mechanism — each of which compensates for the other one's lack of repressive activity, whenever the translation machinery does not recognize one of them. This kind of fail-safe mechanism was already demonstrated for the two AUG uORFs of the human hemojuvelin (*HJV*) mRNA [38]: when uORF1, the strongest inhibitory uORF, is bypassed, it seems that the scanning ribosomes will initiate translation at the uORF2 initiation codon thus maintaining mORF expression at reduced levels.

During stress conditions, several transcripts, known as stress-responsive transcripts, escape the global translational impairment by a uORF-dependent mechanism, in order to deal with stress [1, 159]. This kind of regulation was already demonstrated for *ATF4* and *GCN4*

transcripts, in which the enhanced levels of eIF2α-P inhibit the formation of new and active ternary complexes in time to translate the inhibitory uORFs. This allows them being bypassed, and thus the 40S ribosomal subunit that remains attached to the mRNA after uORF1 translation, will resume scanning and reinitiate translation at the mORF [36, 64, 67]. Here, we show that the human *ABCE1* uORFs are not able to derepress mORF expression in cells treated with the ER-stressor thapsigargin (Figure 3.9B). The repression ability of the AUG uORFs is in fact maintained in normal and ER stress conditions. Indeed, the existence of uORF(s) is not a straight indicator that the corresponding transcripts will be more efficiently translated in stress conditions [51, 154]. uORFs are transversal to different classes of transcripts facing high levels of eIF2α-P during thapsigargin-induced stress, being distributed by three different classes of transcripts: i) transcripts that need to be induced, and thus named of "preferentially translated"; ii) transcripts that are repressed; and iii) transcripts that are resistant to eIF2α-P levels. In general, Baird and co-workers showed that the frequency, number and length of uORFs, are not significantly distinct between the three groups of mRNA defined to be differently regulated in stress conditions. However, it seems that the uORF's initiation context play an important role for the translational control in ER stress. They demonstrated a prevalence of a weak Kozak context surrounding the uORFs in the preferential and resistant group of transcripts, together with a strong Kozak context for the mORF, conversely to what happens for the repression group of transcripts [75]. Taking together, the fact that human ABCE1 expression needs to be kept at basal levels independent of the environmental conditions (normal or stressful), and its uORFs, in fact present a suboptimal Kozak context surrounding their initiation codons along with a strong Kozak context at the mAUG, makes *ABCE1* transcript part of the resistant group. The protein phosphatase 1 regulatory subunit 15B (PPP1R15b) — encoding a protein involved in the dephosphorylation of eIF2α-P, and therefore with an important function for ISR attenuation — was also seen to be maintained at the same expression levels in both thapsigargin and normal conditions, and is in fact listed as a resistant transcript [75, 160, 161]. Thus, we speculate that the need to maintain the low expression of ABCE1 during stress should be related to its function in the translational machinery, like accounts for PPR1R15b.

We also evaluated if *ABCE1* transcript is committed to NMD, based on the fact that the uORFs' stop codons can be recognized as PTCs [3]. We show that *ABCE1* mRNA is not an NMD-target, since in NMD-inhibition (knockdown of *UPF1* mRNA) conditions, *ABCE1* mRNA levels are downregulated (Figure 3.10B), which is contrary to what is expected for NMD-targeted transcripts [55], as demonstrated for *GADD45A* transcript (Figure 3.10C). The half-life of *ABCE1* in NMD-inactivated cells is practically half of the one obtained in NMD-

competent cells, demonstrating that *ABCE1* transcripts decay rapidly in these conditions. Conversely, as demonstrated by Mendell and co-workers, the half-life of several transcripts containing one or more uORFs in conditions of *UPF1*-depletion (NMD inactivation) is significantly higher than in *LUC*-depletion (without NMD inactivation) conditions in HeLa cells [55]. There are several examples of well-known transcripts bearing uORFs that are regulated by NMD, such as *ATF4* [55, 56], *CHOP* [56, 82], and more recently, the *STN1* transcript [58]. uORF-containing transcripts can escape NMD due to two main features: the uORF length, and the time the ribosome takes to translate it [2, 162]. Transcripts with AUG-proximal PTCs (and thus bearing short uORFs) can be NMD-resistant, as already shown for the human β-globin mRNA [162, 163]. This can be explained by the mRNA circularization happening along with uORF translation, where poly(A) binding protein cytoplasmic 1 (PABPC1) is in closer proximity to the uORF termination codon, allowing a proper uORF translation termination, which impairs NMD [164, 165]. On the other hand, transcripts containing long or slowly translatable uORFs usually are NMD-sensitive [77, 163, 166]. In the case of *ABCE1* 5'-leader sequence, there are ten uORFs, either short or long, that could make the *ABCE1* transcript resistant and sensitive, respectively, to NMD. However, since *ABCE1* uORFs regulate its mORF translation with minor impact at the mRNA levels, we exclude the possibility of *ABCE1* transcript being targeted to NMD due to their uORFs. Instead, it seems that the high translation efficiency of the uORF1, being a short uORF, dictates the fate of the transcript, and thus it behaves as NMD-resistant.

A correlation between translation of cancer-related transcripts and the translation of their own uORFs is associated with the initiation of the tumorigenic process [42]. Thus, we wonder if the uORFs present in the 5'-leader sequence of *ABCE1* mRNA can be associated with a potential function in the colorectal cancer tumorigenic process by regulating their mORF expression. We compared the translational control of the *ABCE1* AUG-initiating uORFs in the colorectal cancer cell line HCT116 to the one observed in the non-tumorigenic colorectal cell line NCM460. In Figure 3.11A, we observe that the AUG-initiating uORFs maintain the same regulation pattern both in non-tumorigenic and cancerous cell lines, which demonstrates that *ABCE1* AUG-initiating uORFs are not regulated through the involvement of cancer-specific factors. In addition, *ABCE1* AUG-initiating uORFs do not seem to contribute to the tumorigenic process in HCT116 cells. Moreover, there are not described any genetic alterations in the 5'-leader sequence able to disrupt, alter or create a uORF in the *ABCE1* transcript that can connect those uORF's deregulation to colorectal cancer. High-throughput studies focusing mainly in screening the 5'-leader sequence of the transcripts is of utmost importance to suppress such

lack of information. In addition, we need to study the impact of the uORF-synthesized peptides further, not only as *cis*- and *trans*-regulators of the mAUG expression, but also as cytotoxic molecules, towards colorectal cancer development.

Additionally, we observed that both ABCE1 mRNA and protein levels are equivalent in different colorectal cancer cell lines, and when compared to NCM460 cells (Figure 3.11B and C). Shichijo and co-workers provided evidence that *ABCE1* mRNA is ubiquitously expressed in normal and in cancer colorectal cells, including HCT116 [149], which agrees with our mRNA expression data. Conversely, Hlavata and co-workers reported an up-regulation of *ABCE1* mRNA in a pool of colorectal cancers [150]. However, the mRNA expression is not a straight indication on the protein expression. In fact, we did not find in the literature any research data on the protein levels of ABCE1 in colorectal cancers. In The Human Protein Atlas database, ABCE1 protein was detected as highly expressed in colorectal cancer tissues; however, it was also present at high levels in normal tissues, which agrees with our data. On the other hand, ABCE1 protein has already been shown to be highly expressed in several tumors, such as in human lung adenocarcinoma tissues [141], esophageal carcinomas [145], and breast cancer [144]. Adding to this, depletion of *ABCE1* in different cancer cells substantially impairs cell viability, indicating that ABCE1 plays an oncogenic role in cancer, independent of the cell type. ABCE1 was demonstrated to play a role in cancer biology by inhibiting apoptosis and controlling malignant features, like tumor invasiveness and migration [143–145, 151]. This functional role of ABCE1 in cancer development and progression is sometimes associated with its inhibitory function towards RNase L, since the latter controls several cellular processes like cell proliferation and apoptosis, working as a tumor suppressor gene [141, 144, 145]. However, performing the *ABCE1* siRNA depletion on the colorectal cancer cell line HCTT16, and in the non-tumorigenic cell line NCM460, there was not a significant reduction in cell proliferation (Figure 3.11E), once again showing that ABCE1 seems to be devoid of an oncogenic role in HCT116 cells.

In conclusion, the uORFs within *ABCE1* 5'-leader sequence regulate ABCE1 expression by controlling its translation without changing considerably its mRNA levels, and thus without triggering NMD. ABCE1 is maintained under a certain basal expression levels in the colorectal cancer cells tested under normal conditions. This is accomplished mainly by two AUG uORFs (uORF3 and uORF5) that seem to act together and/or in a fail-safe manner. The same level of expression is maintained under stress conditions, showing that *ABCE1* is a resistant transcript whose functions are essential to be kept under conditions of global translation impairment. Additionally, *ABCE1* uORFs and also ABCE1 protein do not seem to be directly associated

with the tumorigenesis of HCT116 cells but probably the uORF-encoded peptides have some potential regulatory or cytotoxic effect in colorectal cancer cells that need to be further study. Moreover, potential genetic alterations in the 5'-leader sequence of the *ABCE1* mRNA could also have some role in the deregulation of its uORFs that may serve as a potential cancer-related feature. This points out for the importance of 5'-leader region screenings instead of looking only for the main coding sequence of the transcripts to find new cancer-related variants with potential for diagnostic and therapeutic tools.

5. References

1. Young SK, Wek RC (2016) Upstream open reading frames differentially regulate gene-specific translation in the integrated stress response. J Biol Chem 291:16927–16935.
2. Barbosa C, Peixeiro I, Romão L (2013) Gene expression regulation by upstream open reading frames and human disease. PLoS Genet 9:e1003529.
3. Somers J, Pöyry T, Willis AE (2013) A perspective on mammalian upstream open reading frame function. Int J Biochem Cell Biol 45:1690–1700.
4. Sonenberg N, Hinnebusch AG (2009) Regulation of translation initiation in eukaryotes: mechanisms and biological targets. Cell 136:731–745.
5. Jackson RJ, Hellen CUT, Pestova TV (2010) The mechanism of eukaryotic translation initiation and principles of its regulation. Nat Rev Mol Cell Biol 11:113–127.
6. Erickson FL, Hannig EM (1996) Ligand interactions with eukaryotic translation initiation factor 2: role of the y-subunit. EMBO J 15:6611–6320.
7. Pavitt GD, Yang W, Hinnebusch AG (1997) Homologous segments in three subunits of the guanine nucleotide exchange factor eIF2B mediate translational regulation by phosphorylation of eIF2. Mol Cell Biol 17:1298–1313.
8. Eliseev B, Yeramala L, Leitner A, Karuppasamy M, Raimondeau E, Huard K, Alkalaeva E, Aebersold R, Schaffitzel C (2018) Structure of a human cap-dependent 48S translation pre-initiation complex. Nucleic Acids Res 46:2678–2689.
9. Graff JR, Konicek BW, Carter JH, Marcusson EG (2008) Targeting the eukaryotic translation initiation factor 4E for cancer therapy. Cancer Res 68:631–634.
10. Algire MA, Maag D, Lorsch JR (2005) P_i Release from eIF2, not GTP hydrolysis, is the step controlled by start-site selection during eukaryotic translation initiation. Mol Cell 20:251–262.
11. Fringer JM, Acker MG, Fekete CA, Lorsch JR, Dever TE (2007) Coupled release of eukaryotic translation initiation factors 5B and 1A from 80S ribosomes following subunit joining. Mol Cell Biol 27:2384–2397.
12. Lee S, Liu B, Lee S, Huang S-X, Shen B, Qian S-B (2012) Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. Proc Natl Acad Sci U S A 109:E2424–E2432.
13. Kozak M (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell 44:283–292.
14. Pisarev AV, Kolupaeva VG, Pisareva VP, Merrick WC, Hellen CUT, Pestova TV (2006)

Specific functional interactions of nucleotides at key − 3 and + 4 positions flanking the initiation codon with components of the mammalian 48S translation initiation complex. Genes Dev 20:624–636.

15. Kozak M (1987) At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. J Mol Biol 196:947–950.
16. Dever TE, Green R (2012) The elongation, termination, and recycling phases of translation in eukaryotes. Cold Spring Harb Perspect Biol 4:a013706.
17. Gromadski KB, Schümmer T, Strømgaard A, Knudsen CR, Kinzy TG, Rodnina MV (2007) Kinetics of the interactions between yeast elongation factors 1A and 1Bα, guanine nucleotides, and aminoacyl-tRNA. J Biol Chem 282:35629–35637.
18. Agirrezabala X, Frank J (2010) From DNA to proteins via the ribosome: Structural insights into the workings of the translation machinery. Hum Genomics 4:226–237.
19. Beißel C, Neumann B, Uhse S, Hampe I, Karki P, Krebber H (2019) Translation termination depends on the sequential ribosomal entry of eRF1 and eRF3. Nucleic Acids Res 47:4798–4813.
20. McCaughan KK, Brown CM, Dalphin ME, Berry MJ, Tate WP (1995) Translational termination efficiency in mammals is influenced by the base following the stop codon. Proc Natl Acad Sci U S A 92:5431–5435.
21. Bertram G, Bell HA, Ritchie DW, Fullerton G, Stansfield I (2000) Terminating eukaryote translation: domain 1 of release factor eRF1 functions in stop codon recognition. RNA 6:1236–1247.
22. Preis A, Heuer A, Barrio-Garcia C, Hauser A, Eyler DE, Berninghausen O, Green R, Becker T, Beckmann R (2014) Cryoelectron microscopic structures of eukaryotic translation termination complexes containing eRF1-eRF3 or eRF1-ABCE1. Cell Rep 8:59–65.
23. Pisareva VP, Pisarev AV, Hellen CUT, Rodnina MV, Pestova TV (2006) Kinetic analysis of interaction of eukaryotic release factor 3 with guanine nucleotides. J Biol Chem 281:40224–40235.
24. Hellen CUT (2018) Translation termination and ribosome recycling in eukaryotes. Cold Spring Harb Perspect Biol 10:a032656.
25. Heuer A, Gerovac M, Schmidt C, Trowitzsch S, Preis A, Kötter P, Berninghausen O, Becker T, Beckmann R, Tampé R (2017) Structure of the 40S–ABCE1 post-splitting complex in ribosome recycling and translation initiation. Nat Struct Mol Biol 24:453–460.

26. Pisarev AV, Skabkin MA, Pisareva VP, Skabkina OV, Rakotondrafara AM, Hentze MW, Hellen CUT, Pestova TV (2010) The role of ABCE1 in eukaryotic post-termination ribosomal recycling. Mol Cell 37:196–210.
27. Kiosze-Becker K, Ori A, Gerovac M, Heuer A, Nürenberg-Goloub E, Rashid UJ, Becker T, Beckmann R, Beck M, Tampé R (2016) Structure of the ribosome post-recycling complex probed by chemical cross-linking and mass spectrometry. Nat Commun 7:13248.
28. Skabkin MA, Skabkina OV, Dhote V, Komar AA, Hellen CUT, Pestova TV (2010) Activities of Ligatin and MCT-1/DENR in eukaryotic translation initiation and ribosomal recycling. Genes Dev 24:1787–1801.
29. Mancera-Martínez E, Querido JB, Valasek LS, Simonetti A, Hashem Y (2017) ABCE1: A special factor that orchestrates translation at the crossroad between recycling and initiation. RNA Biol 14:1279–1285.
30. Leppek K, Das R, Barna M (2018) Functional 5' UTR mRNA structures in eukaryotic translation regulation and how to find them. Nat Rev Mol Cell Biol 19:158–174.
31. Lacerda R, Menezes J, Romão L (2017) More than just scanning: the importance of cap-independent mRNA translation initiation for cellular stress response and cancer. Cell Mol Life Sci 74:1659–1680.
32. Andrews SJ, Rothnagel JA (2014) Emerging evidence for functional peptides encoded by short open reading frames. Nat Rev Genet 15:193–204.
33. Calvo SE, Pagliarini DJ, Mootha VK (2009) 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.
34. Ingolia NT, Lareau LF, Weissman JS (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity of mammalian proteomes. Cell 147:789–802.
35. Ye Y, Liang Y, Yu Q, Hu L, Li H, Zhang Z, Xu X (2015) Analysis of human upstream open reading frames and impact on gene expression. Hum Genet 134:605–612.
36. Wethmar K, Smink JJ, Leutz A (2010) Upstream open reading frames: Molecular switches in (patho)physiology. Bioessays 32:885–893.
37. Morris DR, Geballe AP (2000) Upstream open reading frames as regulators of mRNA translation. Mol Cell Biol 20:8635–8642.
38. Onofre C, Tomé F, Barbosa C, Silva AL, Romão L (2015) Expression of human hemojuvelin (HJV) is tightly regulated by two upstream open reading frames in HJV mRNA that respond to iron overload in hepatic cells. Mol Cell Biol 35:1376–1389.

39. Sajjanar B, Deb R, Raina SK, Pawar S, Brahmane MP, Nirmale AV, Kurade NP, Manjunathareddy GB, Bal SK, Singh NP (2017) Untranslated regions (UTRs) orchestrate translation reprogramming in cellular stress responses. J Therm Biol 65:69–75.
40. Iacono M, Mignone F, Pesole G (2005) uAUG and uORFs in human and rodent 5′untranslated mRNAs. Gene 349:97–105.
41. Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324:218–223.
42. Sendoel A, Dunn JG, Rodriguez EH, Naik S, Gomez NC, Hurwitz B, Levorse J, Dill BD, Schramek D, Molina H, Weissman JS, Fuchs E (2017) Translation from unconventional 5′ start sites drives tumour initiation. Nature 541:494–499.
43. Tang L, Morris J, Wan J, Moore C, Fujita Y, Gillaspie S, Aube E, Nanda J, Marques M, Jangal M, Anderson A, Cox C, Hiraishi H, Dong L, Saito H, Singh CR, Witcher M, Topisirovic I, Qian S-B, Asano K (2017) Competition between translation initiation factor eIF5 and its mimic protein 5MP determines non-AUG initiation rate genome-wide. Nucleic Acids Res 45:11941–11953.
44. Kozak M (1990) Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. Proc Natl Acad Sci U S A 87:8301–8305.
45. Wethmar K, Barbosa-Silva A, Andrade-Navarro MA, Leutz A (2014) uORFdb—a comprehensive literature database on eukaryotic uORF biology. Nucleic Acids Res 42:D60–D67.
46. Wethmar K (2014) The regulatory potential of upstream open reading frames in eukaryotic gene expression. Wiley Interdiscip Rev RNA 5:765–768.
47. Johnstone TG, Bazzini AA, Giraldez AJ (2016) Upstream ORFs are prevalent translational repressors in vertebrates. EMBO J 35:706–723.
48. Kozak M (2001) Constraints on reinitiation of translation in mammals. Nucleic Acids Res 29:5226–5232.
49. Child SJ, Miller MK, Geballe AP (1999) Translational control by an upstream open reading frame in the HER-2/neu transcript. J Biol Chem 274:24335–24341.
50. Grant CM, Hinnebusch AG (1994) Effect of sequence context at stop codons on efficiency of reinitiation in GCN4 translational control. Mol Cell Biol 14:606–618.
51. Young SK, Willy JA, Wu C, Sachs MS, Wek RC (2015) Ribosome reinitiation directs gene-specific translation and regulates the integrated stress response. J Biol Chem

290:28257–28271.

52. Col B, Oltean S, Banerjee R (2007) Translational regulation of human methionine synthase by upstream open reading frames. Biochim Biophys Acta 1769:532–540.
53. Law GL, Raney A, Heusner C, Morris DR (2001) Polyamine regulation of ribosome pausing at the upstream open reading frame of S-adenosylmethionine decarboxylase. J Biol Chem 276:38036–38043.
54. Pöyry TAA, Kaminski A, Jackson RJ (2004) What determines whether mammalian ribosomes resume scanning after translation of a short upstream open reading frame? Genes Dev 18:62–75.
55. Mendell JT, Sharifi NA, Meyers JL, Martinez-Murillo F, Dietz HC (2004) Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. Nat Genet 36:1073–1078.
56. Gardner LB (2008) Hypoxic inhibition of nonsense-mediated RNA decay regulates gene expression and the integrated stress response. Mol Cell Biol 28:3729–3741.
57. Rebbapragada I, Lykke-Andersen J (2009) Execution of nonsense-mediated mRNA decay: what defines a substrate? Curr Opin Cell Biol 21:394–402.
58. Torrance V, Lydall D (2018) Overlapping open reading frames strongly reduce human and yeast STN1 gene expression and affect telomere function. PLoS Genet 14:e1007523.
59. Donnelly N, Gorman AM, Gupta S, Samali A (2013) The eIF2α kinases: their structures and functions. Cell Mol Life Sci 70:3493–3511.
60. Novoa I, Zeng H, Harding HP, Ron D (2001) Feedback inhibition of the unfolded protein response by GADD34 -mediated dephosphorylation of eIF2α. J Cell Biol 153:1011–1021.
61. Spilka R, Ernst C, Mehta AK, Haybaeck J (2013) Eukaryotic translation initiation factors in cancer development and progression. Cancer Lett 340:9–21.
62. Lee Y-Y, Cevallos RC, Jan E (2009) An upstream open reading frame regulates translation of GADD34 during cellular stresses that induce eIF2α phosphorylation. J Biol Chem 284:6661–6673.
63. Harding HP, Zhang Y, Zeng H, Novoa I, Lu PD, Calfon M, Sadri N, Yun C, Popko B, Paules R, Stojdl DF, Bell JC, Hettmann T, Leiden JM, Ron D (2003) An integrated stress response regulates amino acid metabolism and resistance to oxidative stress. Mol Cell 11:619–633.
64. Hinnebusch AG (2005) Translational regulation of GCN4 and the general amino acid

control of yeast. Annu Rev Microbiol 59:407–450.

65. Young SK, Palam LR, Wu C, Sachs MS, Wek RC (2016) Ribosome elongation stall directs gene-specific translation in the integrated stress response. J Biol Chem 291:6546–6558.
66. Skabkin MA, Skabkina OV, Hellen CUT, Pestova TV (2013) Reinitiation and other unconventional post-termination events during eukaryotic translation. Mol Cell 51:249–264.
67. Vattem KM, Wek RC (2004) Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. Proc Natl Acad Sci U S A 101:11269–11274.
68. Grant CM, Miller PF, Hinnebusch AG (1995) Sequences 5’ of the first upstream open reading frame in GCN4 mRNA are required for efficient translational reinitiation. Nucleic Acids Res 23:3980–3988.
69. Szamecz B, Rutkai E, Cuchalová L, Munzarová V, Herrmannová A, Nielsen KH, Burela L, Hinnebusch AG, Valášek L (2008) eIF3a cooperates with sequences 5' of uORF1 to promote resumption of scanning by post-termination ribosomes for reinitiation on GCN4 mRNA. Genes Dev 22:2414–2425.
70. Munzarová V, Pánek J, Gunišová S, Dányi I, Szamecz B, Valášek LS (2011) Translation reinitiation relies on the interaction between eIF3a/TIF32 and progressively folded cis-acting mRNA elements preceding short uORFs. PLoS Genet 7:e1002137.
71. Calkhoven CF, Müller C, Leutz A (2000) Translational control of C/EBPα and C/EBPβ isoform expression. Genes Dev 14:1920–1932.
72. Wethmar K, Bégay V, Smink JJ, Zaragoza K, Wiesenthal V, Dörken B, Calkhoven CF, Leutz A (2010) C/EBPβDeltauORF mice--a genetic model for uORF-mediated translational control in mammals. Genes Dev 24:15–20.
73. Kozak M (1991) Structural Features in eukaryotic mRNAs that modulate the initiation of translation. J Biol Chem 266:19867–19870.
74. Kozak M (1991) A short leader sequence impairs the fidelity of initiation by eukaryotic ribosomes. Gene Expr 1:111–115.
75. Baird TD, Palam LR, Fusakio ME, Willy JA, Davis CM, McClintick JN, Anthony TG, Wek RC (2014) Selective mRNA translation during eIF2 phosphorylation induces expression of IBTKα. Mol Biol Cell 25:1686–1697.
76. Palam LR, Baird TD, Wek RC (2011) Phosphorylation of eIF2 facilitates ribosomal bypass of an inhibitory upstream ORF to enhance CHOP translation. J Biol Chem 286:10939–10949.

77. Zhao C, Datta S, Mandal P, Xu S, Hamilton T (2010) Stress-sensitive regulation of IFRD1 mRNA decay is mediated by an upstream open reading frame. J Biol Chem 285:8552–8562.
78. Chen Y-J, Tan BC-M, Cheng Y-Y, Chen J-S, Lee S-C (2010) Differential regulation of CHOP translation by phosphorylated eIF4E under stress conditions. Nucleic Acids Res 38:764–777.
79. Rutkowski DT, Arnold SM, Miller CN, Wu J, Li J, Gunnison KM, Mori K, Akha AAS, Raden D, Kaufman RJ (2006) Adaptation to ER stress is mediated by differential stabilities of pro-survival and pro-apoptotic mRNAs and proteins. PLoS Biol 4:e374.
80. Starck SR, Ow Y, Jiang V, Tokuyama M, Rivera M, Qi X, Roberts RW, Shastri N (2008) A distinct translation initiation mechanism generates cryptic peptides for immune surveillance. PLoS ONE 3:e3460.
81. Slavoff SA, Mitchell AJ, Schwaid AG, Cabili MN, Ma J, Levin JZ, Karger AD, Budnik BA, Rinn JL, Saghatelian A (2013) Peptidomic discovery of short open reading frame–encoded peptides in human cells. Nat Chem Biol 9:59–64.
82. Young SK, Baird TD, Wek RC (2016) Translation regulation of the glutamyl-prolyl-tRNA synthetase gene EPRS through bypass of upstream open reading frames with noncanonical initiation codons. J Biol Chem 291:10824–10835.
83. Gao X, Wan J, Liu B, Ma M, Shen B, Qian S-B (2015) Quantitative profiling of initiating ribosomes in vivo. Nat Methods 12:147–153.
84. Schwab SR, Shugart JA, Horng T, Malarkannan S, Shastri N (2004) Unanticipated antigens: Translation initiation at CUG with leucine. PLoS Biol 2:e366.
85. Starck SR, Jiang V, Pavon-Eternod M, Prasad S, McCarthy B, Pan T, Shastri N (2012) Leucine-tRNA initiates at CUG start codons for protein synthesis and presentation by MHC class I. Science 336:1719–1723.
86. Starck SR, Tsai JC, Chen K, Shodiya M, Wang L, Yahiro K, Martins-Green M, Shastri N, Walter P (2016) Translation from the 5′ untranslated region shapes the integrated stress response. Science 351:aad3867.
87. Loughran G, Sachs MS, Atkins JF, Ivanov IP (2012) Stringency of start codon selection modulates autoregulation of translation initiation factor eIF5. Nucleic Acids Res 40:2898–2906.
88. Kozak M (1989) Context effects and inefficient initiation at non-AUG codons in eucaryotic cell-free translation systems. Mol Cell Biol 9:5073–5080.
89. Xue S, Barna M (2012) Specialized ribosomes: a new frontier in gene regulation and

organismal biology. Nat Rev Mol Cell Biol 13:355–369.

90. Sauert M, Temmel H, Moll I (2015) Heterogeneity of the translational machinery: Variations on a common theme. Biochimie 114:39–47.
91. Shi Z, Fujii K, Kovary KM, Genuth NR, Röst HL, Teruel MN, Barna M (2017) Heterogeneous ribosomes preferentially translate distinct subpools of mRNAs genome-wide. Mol Cell 67:71–83.
92. Oyama M, Kozuka-Hata H, Suzuki Y, Semba K, Yamamoto T, Sugano S (2007) Diversity of translation start sites may define increased complexity of the human short ORFeome. Mol Cell Proteomics 6:1000–1006.
93. Crappé J, Ndah E, Koch A, Steyaert S, Gawron D, De Keulenaer S, De Meester E, De Meyer T, Van Criekinge W, Van Damme P, Menschaert G (2015) PROTEOFORMER: deep proteome coverage through ribosome profiling and MS integration. Nucleic Acids Res 43:e29.
94. Raj A, Wang SH, Shim H, Harpak A, Li YI, Engelmann B, Stephens M, Gilad Y, Pritchard JK (2016) Thousands of novel translated open reading frames in humans inferred by ribosome footprint profiling. Elife 5:e13328.
95. Raney A, Law GL, Mize GJ, Morris DR (2002) Regulated translation termination at the upstream open reading frame in s-adenosylmethionine decarboxylase mRNA. J Biol Chem 277:5988–5994.
96. Pendleton LC, Goodwin BL, Solomonson LP, Eichler DC (2005) Regulation of endothelial argininosuccinate synthase expression and NO production by an upstream open reading frame. J Biol Chem 280:24252–24260.
97. Akimoto C, Sakashita E, Kasashima K, Kuroiwa K, Tominaga K, Hamamoto T, Endo H (2013) Translational repression of the McKusick-Kaufman syndrome transcript by unique upstream open reading frames encoding mitochondrial proteins with alternative polyadenylation sites. Biochim Biophys Acta 1830:2728–2738.
98. Besançon R, Valsesia-Wittmann S, Locher C, Delloye-Bourgeois C, Furhman L, Tutrone G, Bertrand C, Jallas A-C, Garin E, Puisieux A (2009) Upstream ORF affects MYCN translation depending on exon 1b alternative splicing. BMC Cancer 9:445.
99. Wethmar K, Schulz J, Muro EM, Talyan S, Andrade-Navarro MA, Leutz A (2016) Comprehensive translational control of tyrosine kinase expression by upstream open reading frames. Oncogene 35:1736–1742.
100. Bersano A, Ballabio E, Bresolin N, Candelise L (2008) Genetic polymorphisms for the study of multifactorial stroke. Hum Mutat 29:776–795.

101. Kanaji T, Okamura T, Osaki K, Kuroiwa M, Shimoda K, Hamasaki N, Niho Y (1998) A common genetic polymorphism (46 C to T substitution) in the 5'-untranslated region of the coagulation factor XII gene is associated with low translation efficiency and decrease in plasma factor XII level. Blood 91:2010–2014.
102. Kondo S, Schutte BC, Richardson RJ, Bjork BC, Knight AS, Watanabe Y, Howard E, De Lima RLLF, Daack-Hirsch S, Sander A, Mcdonald-Mcginn DM, Zackai EH, Lammer EJ, Aylsworth AS, Ardinger HH, Lidral AC, Pober BR, Moreno L, Arcos-Burgos M, Valencia C, Houdayer C, Bahuau M, Moretti-Ferreira D, Richieri-Costa A, Dixon MJ, Murray JC (2002) Mutations in IRF6 cause Van der Woude and popliteal pterygium syndromes. Nat Genet 32:285–289.
103. Poulat F, Desclozeaux M, Tuffery S, Jay P, Boizet B, Berta P (1998) Mutation in the 5' noncoding region of the SRY gene in an XY sex-reversed patient. Hum Mutat Suppl 1:S192–S194.
104. Witt H, Luck W, Hennies HC, Claßen M, Kage A, Laß U, Landt O, Becker M (2000) Mutations in the gene encoding the serine protease inhibitor, Kazal type 1 are associated with chronic pancreatitis. Nat Genet 25:213–216.
105. Silva J, Fernandes R, Romão L (2017) Gene expression regulation by upstream open reading frames in rare diseases. J Rare Dis Res Treat 2:33–38.
106. Nelson ND, Marcogliese A, Bergstrom K, Scheurer M, Mahoney D, Bertuch AA (2016) Thrombopoietin measurement as a key component in the evaluation of pediatric thrombocytosis. Pediatr Blood Cancer 63:1484–1487.
107. Ghilardi N, Wiestner A, Kikuchi M, Ohsaka A, Skoda RC (1999) Hereditary thrombocythaemia in a Japanese family is caused by a novel point mutation in the thrombopoietin gene. Br J Haematol 107:310–316.
108. Bisio A, Nasti S, Jordan JJ, Gargiulo S, Pastorino L, Provenzani A, Quattrone A, Queirolo P, Bianchi-Scarrà G, Ghiorzo P, Inga A (2010) Functional analysis of CDKN2A/p16^{INK4a} 5'-UTR variants predisposing to melanoma. Human Mol Genet 19:1479–1491.
109. Liu L, Dilworth D, Gao L, Monzon J, Summers A, Lassam N, Hogg D (1999) Mutation of the CDKN2A 5' UTR creates an aberrant initiation codon and predisposes to melanoma. Nat Genet 21:128–132.
110. Wen Y, Liu Y, Xu Y, Zhao Y, Hua R, Wang K, Sun M, Li Y, Yang S, Zhang X-J, Kruse R, Cichon S, Betz RC, Nöthen MM, van Steensel MAM, van Geel M, Steijlen PM, Hohl D, Huber M, Dunnill GS, Kennedy C, Messenger A, Munro CS, Terrinoni A, Hovnanian

A, Bodemer C, de Prost Y, Paller AS, Irvine AD, Sinclair R, Green J, Shang D, Liu Q, Luo Y, Jiang L, Chen H-D, Lo WH-Y, McLean WHI, He C-D, Zhang X (2009) Loss-of-function mutations of an inhibitory upstream ORF in the human hairless transcript cause Marie Unna hereditary hypotrichosis. Nat Genet 41:228–233.

111. Zhou Y, Koelling N, Fenwick AL, McGowan SJ, Calpena E, Wall SA, Smithson SF, Wilkie AOM, Twigg SRF (2018) Disruption of *TWIST1* translation by 5′ UTR variants in Saethre-Chotzen syndrome. Hum Mutat 39:1360–1365.
112. Kitano S, Kurasawa H, Aizawa Y (2018) Transposable elements shape the human proteome landscape via formation of *cis*-acting upstream open reading frames. Genes Cells 23:274–284.
113. McGillivray P, Ault R, Pawashe M, Kitchen R, Balasubramanian S, Gerstein M (2018) A comprehensive catalog of predicted functional upstream open reading frames in humans. Nuclei Acids Res 46:3326–3338.
114. Sherr CJ, Roberts JM (1999) CDK inhibitors: positive and negative regulators of G_1-phase progression. Genes Dev 13:1501–1512.
115. Somers J, Wilson LA, Kilday J-P, Horvilleur E, Cannell IG, Pöyry TAA, Cobbold LC, Kondrashov A, Knight JRP, Puget S, Grill J, Grundy RG, Bushell M, Willis AE (2015) A common polymorphism in the 5′ UTR of ERCC5 creates an upstream ORF that confers resistance to platinum-based chemotherapy. Genes Dev 29:1891–1896.
116. Occhi G, Regazzo D, Trivellin G, Boaretto F, Ciato D, Bobisse S, Ferasin S, Cetani F, Pardi E, Korbonits M, Pellegata NS, Sidarovich V, Quattrone A, Opocher G, Mantero F, Scaroni C (2013) A novel mutation in the upstream open reading frame of the CDKN1B gene causes a MEN4 phenotype. PLoS Genet 9:e1003350.
117. Jones L, Goode L, Davila E, Brown A, McCarthy DM, Sharma N, Bhide PG, Armata IA (2017) Translational effects and coding potential of an upstream open reading frame associated with DOPA Responsive Dystonia. Biochim Biophys Acta Mol Basis Dis 1863:1171–1182.
118. Schulz J, Mah N, Neuenschwander M, Kischka T, Ratei R, Schlag PM, Castaños-Vélez E, Fichtner I, Tunn P-U, Denkert C, Klaas O, Berdel WE, von Kries JP, Makalowski W, Andrade-Navarro MA, Leutz A, Wethmar K (2018) Loss-of-function uORF mutations in human malignancies. Sci Rep 8:2395.
119. Brown CY, Mize GJ, Pineda M, George DL, Morris DR (1999) Role of two upstream open reading frames in the translational control of oncogene mdm2. Oncogene 18:5631–5637.

120. Sobczak K, Krzyzosiak WJ (2002) Structural determinants of BRCA1 translational regulation. J Biol Chem 277:17349–17358.
121. Mehta A, Trotta CR, Peltz SW (2006) Derepression of the Her-2 uORF is mediated by a novel post-transcriptional control mechanism in cancer cells. Genes Dev 20:939–953.
122. Spevak CC, Park E-H, Geballe AP, Pelletier J, Sachs MS (2006) Her-2 upstream open reading frame effects on the use of downstream initiation codons. Biochem Biophys Res Commun 350:834–841.
123. Dean M, Rzhetsky A, Allikmets R (2001) The human ATP-binding cassette (ABC) transporter superfamily. Genome Res 11:1156–1166.
124. Tian Y, Han X, Tian D (2012) The biological regulation of ABCE1. IUBMB Life 64:795–800.
125. Dean M, Annilo T (2005) Evolution of the ATP-binding cassette (ABC) transporter superfamily in vertebrates. Annu Rev Genomics Hum Genet 6:123–142.
126. Bisbal C, Martinand C, Silhol M, Lebleu B, Salehzada T (1995) Cloning and characterization of a RNase L inhibitor. A new component of the interferon-regulated 2-5A pathway. J Biol Chem 270:13308–13317.
127. Toompuu M, Kärblane K, Pata P, Truve E, Sarmiento C (2016) ABCE1 is essential for S phase progression in human cells. Cell Cycle 15:1234–1247.
128. Chen Z, Dong J, Ishimura A, Daar I, Hinnebusch AG, Dean M (2006) The essential vertebrate ABCE1 protein interacts with eukaryotic initiation factors. J Biol Chem 281:7452–7457.
129. Karcher A, Schele A, Hopfner K-P (2008) X-ray structure of the complete ABC enzyme ABCE1 from Pyrococcus abyssi. J Biol Chem 283:7962–7971.
130. Li X-L, Blackford JA, Hassel BA (1998) RNase L mediates the antiviral effect of interferon through a selective reduction in viral RNA during encephalomyocarditis virus infection. J Virol 72:2752–2759.
131. Liang S-L, Quirk D, Zhou A (2006) RNase L: Its biological roles and regulation. IUBMB Life 58:508–514.
132. Martinand C, Montavon C, Salehzada T, Silhol M, Lebleu B, Bisbal C (1999) RNase L inhibitor is induced during human immunodeficiency virus type 1 infection and down regulates the 2-5A/RNase L pathway in human T cells. J Virol 73:290–296.
133. Zimmerman C, Klein KC, Kiser PK, Singh AR, Firestein BL, Riba SC, Lingappa JR (2002) Identification of a host protein essential for assembly of immature HIV-1 capsids. Nature 415:88–92.

134. Dooher JE, Schneider BL, Reed JC, Lingappa JR (2007) Host ABCE1 is at plasma membrane HIV assembly sites and its dissociation from Gag is linked to subsequent events of virus production. Traffic 8:195–211.
135. Kärblane K, Gerassimenko J, Nigul L, Piirsoo A, Smialowska A, Vinkel K, Kylsten P, Ekwall K, Swoboda P, Truve E, Sarmiento C (2015) ABCE1 is a highly conserved RNA silencing suppressor. PLoS ONE 10:e0116702.
136. Dong J, Lai R, Nielsen K, Fekete CA, Qiu H, Hinnebusch AG (2004) The essential ATP-binding cassette protein RLI1 functions in translation by promoting preinitiation complex assembly. J Biol Chem 279:42157–42168.
137. Pisareva VP, Skabkin MA, Hellen CUT, Pestova TV, Pisarev AV (2011) Dissociation by Pelota, Hbs1 and ABCE1 of mammalian vacant 80S ribosomes and stalled elongation complexes. EMBO J 30:1804–1817.
138. Kispal G, Sipos K, Lange H, Fekete Z, Bedekovics T, Janáky T, Bassler J, Netz DJA, Balk J, Rotte C, Lill R (2005) Biogenesis of cytosolic ribosomes requires the essential iron-sulphur protein Rli1p and mitochondria. EMBO J 24:589–598.
139. Al-Haj L, Blackshear PJ, Khabar KSA (2012) Regulation of p21/CIP1/WAF-1 mediated cell-cycle arrest by RNase L and tristetraprolin, and involvement of AU-rich elements. Nucleic Acids Res 40:7739–7752.
140. Diriong S, Salehzada T, Bisbal C, Martinand C, Taviaux S (1996) Localization of the ribonuclease L inhibitor gene (RNS4I), a new member of the interferon-regulated 2–5A pathway, to 4q31 by fluorescence in situ hybridization. Genomics 32:488–490.
141. Ren YI, Li Y, Tian D (2012) Role of the ABCE1 gene in human lung adenocarcinoma. Oncol Rep 27:965–970.
142. Liang Z, Yu Q, Ji H, Tian D (2019) Tip60-siRNA regulates ABCE1 acetylation to suppress lung cancer growth via activation of the apoptotic signaling pathway. Exp Ther Med 17:3195–3202.
143. Huang B, Gao Y, Tian D, Zheng M (2010) A small interfering ABCE1-targeting RNA inhibits the proliferation and invasiveness of small cell lung cancer. Int J Mol Med 25:687–693.
144. Huang B, Zhou H, Lang X, Liu Z (2014) siRNA-induced ABCE1 silencing inhibits proliferation and invasion of breast cancer cells. Mol Med Rep 10:1685–1690.
145. Huang B, Gong X, Zhou H, Xiong F, Wang S (2014) Depleting ABCE1 expression induces apoptosis and inhibits the ability of proliferation and migration of human esophageal carcinoma cells. Int J Clin Exp Pathol 7:584–592.

146. Seborova K, Vaclavikova R, Soucek P, Elsnerova K, Bartakova A, Cernaj P, Bouda J, Rob L, Hruda M, Dvorak P (2019) Association of ABC gene profiles with time to progression and resistance in ovarian cancer revealed by bioinformatics analyses. Cancer Med 8:606–616.
147. Zhang P, Chen X-B, Ding B-Q, Liu H-L, He T (2018) Down-regulation of ABCE1 inhibits temozolomide resistance in glioma through the PI3K/Akt/NF-κB signaling pathway. Biosci Rep 38:BSR20181711.
148. Yu Q, Han X, Tian D-L (2017) Deficiency of functional iron-sulfur domains in ABCE1 inhibits the proliferation and migration of lung adenocarcinomas by regulating the biogenesis of beta-actin in vitro. Cell Physiol Biochem 44:554–566.
149. Shichijo S, Ishihara Y, Azuma K, Komatsu N, Higashimoto N, Ito M, Nakamura T, Ueno T, Harada M, Itoh K (2005) ABCE1, a member of ATP-binding cassette transporter gene, encodes peptides capable of inducing HLA-A2-restricted and tumor-reactive cytotoxic T lymphocytes in colon cancer patients. Oncol Rep 13:907–913.
150. Hlavata I, Mohelnikova-Duchonova B, Vaclavikova R, Liska V, Pitule P, Novak P, Bruha J, Vycital O, Holubec L, Treska V, Vodicka P, Soucek P (2012) The role of ABC transporters in progression and clinical outcome of colorectal cancer. Mutagenesis 27:187–196.
151. Kara G, Tuncer S, Türk M, Denkbaş EB (2015) Downregulation of ABCE1 via siRNA affects the sensitivity of A549 cells against chemotherapeutic agents. Med Oncol 32:103.
152. Vanderperre B, Lucier J-F, Bissonnette C, Motard J, Tremblay G, Vanderperre S, Wisztorski M, Salzet M, Boisvert F-M, Roucou X (2013) Direct detection of alternative open reading frames translation products in human significantly expands the proteome. PLoS ONE 8:e70698.
153. Wurth L, Papasaikas P, Olmeda D, Bley N, Calvo GT, Guerrero S, Cerezo-Wallis D, Martinez-Useros J, García-Fernández M, Hüttelmaier S, Soengas MS, Gebauer F (2016) UNR/CSDE1 drives a post-transcriptional program to promote melanoma invasion and metastasis. Cancer Cell 30:694–707.
154. Andreev DE, O'Connor PB, Fahey C, Kenny EM, Terenin IM, Dmitriev SE, Cormican P, Morris DW, Shatsky IN, Baranov PV (2015) Translation of 5′ leaders is pervasive in genes resistant to eIF2 repression. Elife 4:e03971.
155. Barbosa C, Romão L (2014) Translation of the human erythropoietin transcript is regulated by an upstream open reading frame in response to hypoxia. RNA 20:594–608.
156. Abdullahi A, Stanojcic M, Parousis A, Patsouris D, Jeschke MG (2017) Modeling acute

ER stress in vivo and in vitro. Shock 47:506–513.

157. Oslowski CM, Urano F (2011) Measuring ER stress and the unfolded protein response using mammalian tissue culture system. Methods Enzymol 490:71–92.
158. Tani H, Imamachi N, Salam KA, Mizutani R, Ijiri K, Irie T, Yada T, Suzuki Y, Akimitsu N (2012) Identification of hundreds of novel UPF1 target transcripts by direct determination of whole transcriptome stability. RNA Biol 9:1370–1379.
159. Spriggs KA, Bushell M, Willis AE (2010) Translational regulation of gene expression during conditions of cell stress. Mol Cell 40:228–237.
160. Harding HP, Zhang Y, Scheuner D, Chen J-J, Kaufman RJ, Ron D (2009) Ppp1r15 gene knockout reveals an essential role for translation initiation factor 2 alpha (eIF2α) dephosphorylation in mammalian development. Proc Natl Acad Sci U S A 106:1832–1837.
161. Jousse C, Oyadomari S, Novoa I, Lu P, Zhang Y, Harding HP, Ron D (2003) Inhibition of a constitutive translation initiation factor 2 phosphatase, CReP , promotes survival of stressed cells. J Cell Biol 163:767–775.
162. Silva AL, Pereira FJC, Morgado A, Kong J, Martins R, Faustino P, Liebhaber SA, Romão L (2006) The canonical UPF1-dependent nonsense-mediated mRNA decay is inhibited in transcripts carrying a short open reading frame independent of sequence context. RNA 12:2160–2170.
163. Inácio A, Silva AL, Pinto J, Ji X, Morgado A, Almeida F, Faustino P, Lavinha J, Liebhaber SA, Romão L (2004) Nonsense mutations in close proximity to the initiation codon fail to trigger full nonsense-mediated mRNA decay. J Biol Chem 279:32170–32180.
164. Silva AL, Ribeiro P, Inácio A, Liebhaber SA, Romão L (2008) Proximity of the poly(A)-binding protein to a premature termination codon inhibits mammalian nonsense-mediated mRNA decay. RNA 14:563–576.
165. Peixeiro I, Inácio A, Barbosa C, Silva AL, Liebhaber SA, Romão L (2012) Interaction of PABPC1 with the translation initiation complex is critical to the NMD resistance of AUG-proximal nonsense mutations. Nucleic Acids Res 40:1160–1173.
166. Hurt JA, Robertson AD, Burge CB (2013) Global analyses of UPF1 binding and function reveal expanded scope of nonsense-mediated mRNA decay. Genome Res 23:1636–1650.