U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

INSTITUTO DE INVESTIGAÇÃO
E INOVAÇÃO EM SAÚDE
UNIVERSIDADE DO PORTO

Functional relevance of *MCL1* alternative 3'UTR mRNA isoforms in human cells

Beatriz Carreiro Garcia

Mestrado em Biologia Celular e Molecular

Departamento de Biologia

2019

Orientadora

Doutora Isabel Pereira-Castro, Investigadora Júnior, i3S/IBMC

Coorientadora

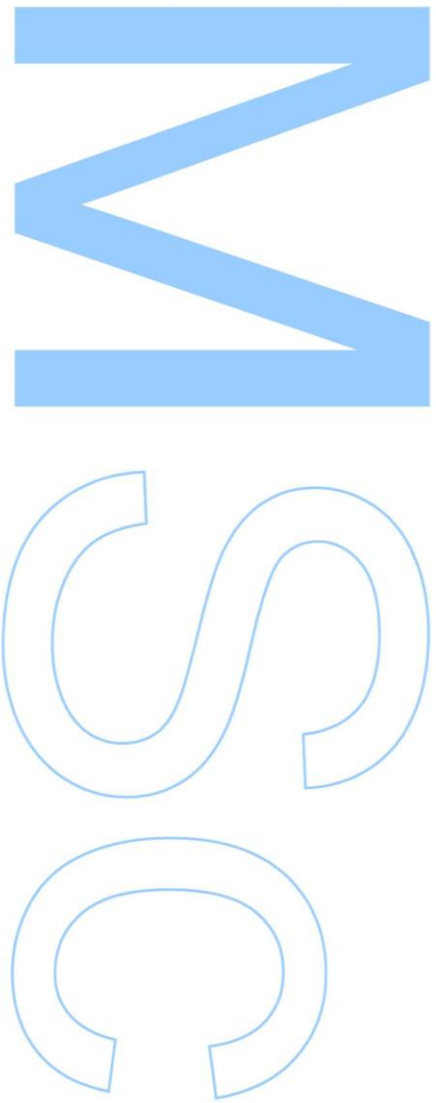
Doutora Alexandra Moreira, Investigadora Principal, ICBAS, UP

U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, _____/_____/_____



Beatriz Carreiro Garcia

Faculdade de Ciências da Universidade do Porto

Rua do Campo Alegre 1021/1055, 4169-007 Porto, Portugal

up201406231@fc.up.pt / biacgarcia13@gmail.com

914608746

Isabel Pereira-Castro (Orientadora)

Investigadora Júnior

Gene Regulation Group, IBMC- Instituto de Biologia Celular e Molecular

i3S- Instituto de Investigação e Inovação em Saúde, Universidade do Porto

Rua Alfredo Allen, 208

4200-135 Porto

isabel.castro@ibmc.up.pt

220408800

Maria Alexandra Marques Moreira Mourão do Carmo (Coorientadora)

Investigadora Principal

ICBAS- Instituto de Ciências Biológicas de Abel Salazar

Gene Regulation Group, IBMC- Instituto de Biologia Celular e Molecular

i3S- Instituto de Investigação e Inovação em Saúde, Universidade do Porto

Rua Alfredo Allen, 208

4200-135 Porto

alexandra.moreira@i3s.up.pt

220408800

Declaração

Eu, Beatriz Carreiro Garcia, aluna com o número 201406231 do Mestrado em Biologia Celular e Molecular da edição de 2018/2019, declaro por minha honra que sou a autora da totalidade do texto apresentado, não apresento texto plagiado, e tomei conhecimento das consequências de uma situação de plágio.

Porto, Setembro 2019

Beatriz Carreiro Garcia

Agradecimentos

Em primeiro lugar, quero agradecer à minha orientadora Doutora Isabel Pereira-Castro e à minha co-orientadora Doutora Alexandra Moreira por me terem recebido como aluna de mestrado no grupo Gene Regulation e por me terem proporcionado esta experiência incrível. Foi uma honra poder fazer parte de um grupo fantástico e poder desenvolver um projeto com a Isabel e a Alexandra. Também quero agradecer aos restantes membros do grupo, Ana Jesus, Joana Wilton e Marta Oliveira. Graças a todos vocês, a minha paixão pela ciência tornou-se ainda maior. Isabel, quero te agradecer por teres sido, não só, a melhor mentora que podia ter, mas também, uma excelente amiga e conselheira, e por saberes sempre lidar comigo (eu sei que sou chata). Desde sempre que foi notório o teu empenho e dedicação no sentido de me orientar da melhor maneira possível. Muito obrigada por tudo o que me ensinaste, por estares sempre presente e disponível e por seres um exemplo a seguir. Sou uma sortuda por te ter conhecido. Não há palavras que descrevam o quão grata estou por ter sido tua aluna. Joana, my sister from another mother, és uma pessoa extraordinária que marcou a minha vida de modo estatisticamente significativo. Rapidamente te tornaste uma amiga muito especial. Fizeste me crescer muito. Obrigada por tudo. Quero também agradecer aos meus pais, as duas pessoas responsáveis por me proporcionar uma vida e um futuro melhor. Foi graças a vocês e ao vosso apoio que alcancei os meus objetivos académicos e cheguei aqui. Muito obrigada por acreditarem sempre em mim e por lutarem sempre para a concretização dos meus sonhos. Quero agradecer especialmente à minha mãe por me ter ensinado a lutar sempre pelos meus objetivos, por mais difíceis que fossem, por ser sempre o meu alicerce e por toda a força e motivação que nunca deixou de me transmitir. Às minhas avós, vocês são incríveis e serão sempre um grande exemplo a seguir. À minha madrinha, obrigada por me ajudares sempre que precisei e por quereres sempre o melhor para mim. À minha restante família, quero também agradecer por estarem sempre a torcer por mim. Também não posso deixar de agradecer a todos os meus amigos por todo o apoio, apesar da distância. Muito obrigada por não me deixarem estar sempre a estudar, por tornarem a minha vida muito mais divertida e por todas as visitas que me fizeram. Nunca gostei tanto de ter um acampamento na sala da minha casa. À minha pequenina, Inês Costa, muito obrigada por existires. Foste a melhor amiga que a faculdade me deu. Esta etapa não teria sido igual se não a tivesse feito contigo ao meu lado. Foste incansável e estiveste sempre comigo quando precisei. Nunca te esqueças do teu DNA mitocondrial e que o declive é a resposta para tudo. Espero estar a fazer-me entender e não estar a ser muito confusa.

Abstract

Polyadenylation consists of the polymerization of an adenosine tail at the 3' end of mRNAs, due to the recognition of a hexamer sequence, known as the polyA signal. This final step of pre-mRNA processing is accountable for completing the maturation of nearly all eukaryotic mRNAs. Approximately 70% of mammalian genes contain more than one polyA signal, meaning they can generate different mRNA isoforms through the selection of different signals, a process that is designated by alternative polyadenylation (APA). In the human genome, about 54% of genes contain multiple polyA signals downstream of the stop codon. By 3' untranslated region (3'UTR)-APA, these genes are capable of producing different mRNA isoforms that only differ in the 3'UTR length, encoding the same protein. It is well stablished that the 3'UTR plays a major role in gene expression regulation, since it influences mRNA nuclear export, stability, localization, and translation efficiency. Furthermore, it has also been demonstrated that alternative 3'UTRs regulate protein function and localization. Myeloid cell leukemia 1 (Mcl-1) is an anti-apoptotic member of the Bcl-2 protein family with an essential role in cell survival. Although *MCL1* gene contain multiple polyA signals, there are no studies performed so far concerning *MCL1* regulation by alternative polyadenylation. Our laboratory has previously identified two *MCL1* mRNA isoforms generated by usage of the polyA signal 1 or polyA signal 2, entitled pA1 and pA2, which only differ in their 3'UTR length. Additionally, our group found that the protein levels and subcellular localization of Mcl-1 encoded by pA1 and pA2 are remarkably different, although the correspondent mRNAs code for the same Mcl-1 protein. Since the functional role of these two *MCL1* APA-derived mRNA isoforms has not yet been clarified, we focused on understanding their physiological effect on Mcl-1 function (cell survival) and their influence in Mcl-1 subcellular localization. To study the individual function of each *MCL1* mRNA isoforms, we applied CRISPR/Cas9 genome editing technology to either delete pA1 or pA2 isoforms in human cell lines. Genotyping techniques, such as PCR and sequencing, allowed us to determine that the genome editing of *MCL1* isoforms occurred successfully. By RT-qPCR and Western blot we validated that the CRISPR/Cas9 deletions indeed affected *MCL1* pA1 and pA2 mRNA expression and Mcl-1 protein levels. We then assessed the physiological effect of each *MCL1* APA isoform in Mcl-1 function in cell viability, quantifying the apoptosis levels in the edited cells by Annexin V-Propidium iodide staining. Our results demonstrated that both pA1 and pA2 mRNA isoforms contribute to the anti-apoptotic function of Mcl-1. Regarding the subcellular localization of Mcl-1, our group found that, while Mcl-1 protein encoded by pA1 mRNA isoform localizes in the mitochondria, which is the standard Mcl-1 localization, Mcl-1 protein encoded by pA2 mRNA isoform has a more diffuse

distribution throughout the cell. To explore the impact of the two *MCL1* APA-derived mRNA isoforms on Mcl-1 subcellular localization, we investigated whether delocalization of Mcl-1 pA2 encoded protein was owed to the low expression levels. To test this hypothesis, we decrease Mcl-1 protein expression by co-transfecting HeLa cells with a *MCL1*-specific Locked Nucleic Acid (LNA), together with EGFP-*MCL1*-pA1 (which has high Mcl-1 expression levels), and analyzed Mcl-1 protein localization by confocal microscopy. Notably, we observed that although Mcl-1 expression levels decreased due to the LNA, Mcl-1 protein derived from pA1 still localizes in the mitochondria, the standard Mcl-1 localization. These results indicate that Mcl-1 delocalization is not due to a question of a lower protein amount but rather due to features of the 3'UTR of the pA2 mRNA. This finding prompt us to analyze EGFP-*MCL1* fused constructs with different 3'UTRs lengths comprised between polyA signal 1 and polyA signal 2 (pEGFP-*MCL1*CDS-1920 and pEGFP-*MCL1*CDS-2371). Although pEGFP-*MCL1*CDS-1920 possess a 3'UTR sequence with about 500 nucleotides downstream of pA1, the Mcl-1 protein produced showed the same localization as the Mcl-1 protein encoded by pA1 mRNA. Interestingly, the Mcl-1 protein produced by pEGFP-*MCL1*CDS-2371 displayed the same localization as the Mcl-1 pA2 encoded protein. These results allowed us to define a region of approximately 450 nucleotides, contained between polyA signal 1 and polyA signal 2, which is responsible for altering the subcellular localization of the Mcl-1 protein. Additionally, we concluded that *MCL1* alternative 3'UTRs regulate Mcl-1 protein levels. Taken together, our findings demonstrate that both mRNA isoforms derived by 3'UTR-APA contribute to Mcl-1 anti-apoptotic function but play distinct roles on Mcl-1 protein expression and subcellular localization.

Key Words: Gene expression, 3' end processing, alternative polyadenylation, polyA signal, mRNA isoforms, cell cultures, CRISPR/Cas9 technology.

Resumo

A poliadenilação consiste na polimerização de uma cauda de adenosinas na extremidade 3' dos mRNAs, devido ao reconhecimento de uma sequência hexamérica, conhecida como o sinal poliA. Esta etapa final do processamento do pré-mRNA é necessária para completar a maturação de quase todos os mRNAs eucarióticos. Nos mamíferos, aproximadamente 70% dos genes contém mais que um sinal poliA, o que significa que podem gerar diferentes isoformas de mRNA através da seleção de sinais diferentes, um processo designado por poliadenilação alternativa (APA). No genoma humano, cerca de 54% dos genes, depois de transcritos no pré-mRNA, contém múltiplos sinais de poliA a jusante do codão stop. Por APA na região 3' não traduzida (3'UTR-APA), estes genes são capazes de produzir diferentes isoformas de mRNA que apenas diferem no comprimento do 3'UTR, codificando a mesma proteína. Está estabelecido que o 3'UTR desempenha um papel muito importante na regulação da expressão genética, visto que influencia a exportação nuclear, a estabilidade, a localização e a eficiência de tradução do mRNA. Para além disso, também já foi demonstrado que 3'UTRs alternativos regulam a função e a localização das proteínas. A proteína Myeloid cell leucemia 1 (Mcl-1) é um membro anti-apoptótico da família de proteínas Bcl-2 com um papel essencial na sobrevivência celular. Apesar do gene *MCL1* possuir múltiplos sinais de poliA, até ao momento, não foram realizados estudos sobre a regulação do *MCL1* por poliadenilação alternativa. Anteriormente, o nosso laboratório identificou duas isoformas de mRNA do *MCL1* geradas pelo uso do sinal poliA 1 ou 2, designadas pA1 e pA2, que apenas diferem no comprimento do seu 3'UTR. Adicionalmente, o nosso grupo descobriu que a localização subcelular e os níveis proteicos do Mcl-1 codificado por pA1 e pA2 são notoriamente diferentes, apesar dos mRNAs correspondentes codificarem a mesma proteína Mcl-1. Tendo em conta que ainda não se conhece a importância funcional das duas isoformas de mRNA do *MCL1* derivadas da APA, focamo-nos em estudar o efeito fisiológico de cada uma delas na função (sobrevivência celular) e na localização subcelular da proteína Mcl-1. Para estudar a função individual de cada isoforma de mRNA do *MCL1*, aplicou-se a tecnologia de edição genética CRISPR/Cas9 para deletar a isoforma pA1 ou a isoforma pA2 em linhas celulares humanas. Técnicas de genotipagem, como o PCR ou a sequenciação, permitiram determinar que a edição genética das isoformas do *MCL1* ocorreu com sucesso. Por RT-qPCR e western blot validou-se que as deleções afetaram os níveis de expressão das isoformas pA1 e pA2 e os níveis proteicos do Mcl-1. Depois avaliou-se o efeito fisiológico de cada isoforma na função do Mcl-1 na viabilidade celular, através da quantificação dos níveis de apoptose nas células editadas por marcação com Anexina

V-iodeto de propídio. Os nossos resultados demonstram que ambas as isoformas de mRNA, pA1 e pA2, contribuem para a função anti-apoptótica do Mcl-1. Em relação à localização subcelular do Mcl-1, o nosso grupo observou que, enquanto a proteína Mcl-1 codificada pela isoforma de mRNA pA1 localiza-se na mitocôndria, que é a típica localização do Mcl-1, a proteína Mcl-1 codificada pela isoforma de mRNA pA2 apresenta uma distribuição mais difusa por toda a célula. Para explorar o impacto das duas isoformas de mRNA do *MCL1* geradas por APA na localização subcelular do Mcl-1, investigamos se a deslocalização da proteína codificada pela isoforma pA2 era devido aos baixos níveis de expressão. Para testar esta hipótese, diminui-se a expressão da proteína Mcl-1 através de uma co-transfeção das células HeLa com um Locked Nucleic Acid (LNA) específico do *MCL1* juntamente com EGFP-*MCL1*-pA1 (que tem níveis mais altos de expressão do Mcl-1), e analisou-se a localização da proteína Mcl-1 por microscopia confocal. Notoriamente, observou-se que, apesar da diminuição dos níveis de expressão do Mcl-1 devido ao LNA, a proteína Mcl-1 codificada pela isoforma pA1 continua a apresentar a localização mitocondrial, característica do Mcl-1. Este resultado indica que a deslocalização do Mcl-1 não se deve a uma questão de menor quantidade proteica, mas sim devido a características do 3'UTR da isoforma de mRNA pA2. Este resultado incentivou-nos a analisar constructos de EGFP-*MCL1* com 3'UTRs de diferentes tamanhos compreendidos entre o sinal poliA 1 e 2 (pEGFP-*MCL1*CDS-1920 e pEGFP-*MCL1*CDS-2371). Apesar do pEGFP-*MCL1*CDS-1920 possuir uma sequência 3'UTR com cerca de 500 nucleótidos a jusante do pA1, a proteína Mcl-1 produzida apresentou a mesma localização que a proteína Mcl-1 codificada pela isoforma de mRNA pA1. Curiosamente, a proteína Mcl-1 produzida por pEGFP-*MCL1*CDS-2371 mostrou a mesma localização da proteína Mcl-1 codificada pela isoforma pA2. Estes resultados permitiram-nos definir uma região de cerca de 450 nucleótidos, contida entre os sinais de poliA 1 e 2, responsável por alterar a localização subcelular da proteína Mcl-1. Além disso, conclui-se que as 3'UTRs alternativas do *MCL1* regulam os níveis proteicos do Mcl-1. Em conjunto, os nossos resultados demonstram que ambas as isoformas de mRNA derivadas da 3'UTR-APA contribuem para a função anti-apoptótica do Mcl-1 mas desempenham papéis distintos na expressão e localização subcelular da proteína.

Palavras-chave: Expressão genética, processamento da extremidade 3', poliadenilação alternativa, sinal poliA, isoformas de mRNA, culturas celulares, tecnologia CRISPR/Cas9.

Index

List of Figures

Abbreviations

3'UTR | 3' untranslated region

APA | alternative polyadenylation

AREs | AU-rich elements

ASO | antisense oligonucleotides

aUTR | alternative UTR

Bak | Bcl-2 homologous antagonist killer 1

Bax | Bcl-2-associated X protein

Bcl-2 | B-cell lymphoma 2

BH | Bcl-2 homology

BH1 | Bcl-2 homology domain 1

BH2 | Bcl-2 homology domain 2

BH3 | Bcl-2 homology domain 3

BIRC3 | Baculoviral IAP repeated containing 3

Cas | CRISPR-associated protein

Cas3 | CRISPR-associated protein 3

Cas9 | CRISPR-associated protein 9

Cas10 | CRISPR-associated protein 10

CD47 | cluster of differentiation 47

Cdc42 | Cell division cycle 42

CDK-1 | cyclin dependent kinase 1

CDS | coding sequence

CFIm | cleavage factor Im

CFIIm | cleavage factor IIm

CHK-1 | checkpoint 1 protein

CLL | chronic lymphocytic leukemia

CPSF | cleavage and polyadenylation specificity factor

CPSF1 | cleavage and polyadenylation specificity factor subunit 1

CPSF2 | cleavage and polyadenylation specificity factor subunit 2

CPSF3 | cleavage and polyadenylation specificity factor subunit 3

CPSF4 | cleavage and polyadenylation specificity factor subunit 4

CR-APA | coding region-APA

CRISPR | clustered regularly interspaced short palindromic repeats

crRNA | CRISPR RNA

crRNP | CRISPR ribonucleoprotein

CstF | cleavage stimulation factor

CstF1 | cleavage stimulation factor subunit 1

CstF77 | cleavage stimulation factor subunit 77

CstF64 | cleavage stimulation factor subunit 64

CTD | carboxy-terminal domain

cUTR | constitutive UTR

DSBs | double-stranded breaks

DSE | downstream sequence elements

EGFP | enhanced green fluorescent protein

ELAVL1 | embryonic lethal abnormal vision-like 1

FIP1 | factor interacting with PAP

GTPases | guanosine triphosphatases

HDR | homology-directed repair

HEK | human embryonic kidney

Hip2 | Huntingtin interacting protein 2

hnRNPH | heterogeneous nuclear ribonucleoprotein H

HuR | human antigen R

IgM | immunoglobulin M

Indels | insertions or deletions

iPS | induced pluripotent stem

IQGAP1 | IQ motif containing GTPase activating protein 1

LNA | locked nucleic acid

Mcl-1 | Myeloid cell leukemia 1 protein

Mcl-1L | Myeloid cell leukemia 1 full length splice variant

Mcl-1S/ΔTM | Myeloid cell leukemia 1 short splice variant

Mcl-1ES | Myeloid cell leukemia 1 extra short splice variant

MCL1 | Myeloid cell leukemia 1 gene

MCL1 pA1 | short *MCL1* APA-derived mRNA isoform

MCL1 pA2 | long *MCL1* APA-derived mRNA isoform

miRNA | micro-RNA

MGEs | mobile genetic elements

mRNA | messenger RNA

NHEJ | non-homologous end joining

NOVA2 | neuro-oncological ventral antigen 2

OMM | outer mitochondria membrane

PABPN1 | polyA binding protein nuclear 1

PAM | protospacer adjacent motif

PAP | polyA polymerase

PAS | polyA signal

PAS1 | polyA signal 1

PAS2 | polyA signal 2

PBS | Phosphate Buffered Saline

PCNA | proliferating cell nuclear antigen

PEST | proline (P), glutamic acid (E), serine (S), threonine (T)

pre-crRNA | precursor crRNA

pre-mRNA | precursor mRNA

Primer F | primer forward

Primer R | primer reverse

PTBP1 | polypyrimidine tract binding protein

Rac1 | Rac family small GTPase 1

RALA | Ras like proto-oncogene A

RBP | RNA-binding protein

Rho | Ras homologue

RhoA | Ras homolog family member A

RNA | ribonucleic acid

RNAPII | RNA polymerase II

sgRNA | single-guide RNA

snRNP | small nuclear ribonucleoprotein

TM | transmembrane

T_{reg} | regulatory T

tracrRNA | transactivating crRNA

3'UDPL | 3'UTR-dependent protein localization

USE | upstream sequence elements

WDR33 | WD repeat domain 33

Introduction

1| Gene expression and pre-mRNA processing in eukaryotes

Gene expression is a process that begins with the transcription of DNA into a ribonucleotide sequence that forms the ribonucleic acid (RNA) molecule[1]. This process, called transcription, is performed by RNA polymerases and occurs in the nucleus. When the gene expression final product is a protein, we are in the presence of a protein-coding gene and its transcription product is designated messenger RNA (mRNA). Most protein-coding genes are transcribed by RNA polymerase II (RNAPII). After RNAPII transcription, mRNA molecules pass through the nuclear pores and reach the cytoplasm to be translated by ribosomes into proteins[1,2,3].

In the nucleus, transcription of a protein-coding gene originates a precursor mRNA (pre-mRNA) that needs to be processed into a mature mRNA[1,2,3]. This pre-mRNA processing (Figure 1) consists of 5' capping, splicing and polyadenylation and regulates gene expression[4,5,6,7,8,9]. These three steps of pre-mRNA processing occur sequentially during transcription, and most of them are co-transcriptional[2].

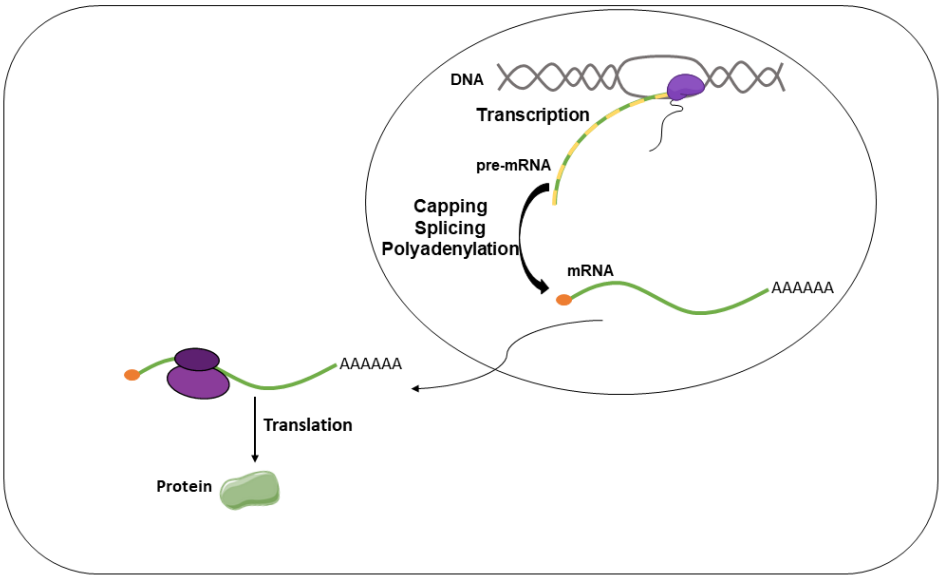


Figure 1: Eukaryotic pre-mRNA processing | In the nucleus, RNA Polymerase II is responsible for DNA transcription and generation of the pre-mRNA transcripts. These pre-mRNA molecules are processed by capping, splicing and polyadenylation giving rise to mature mRNA transcripts. Finally, mRNAs are exported to the cytoplasm allowing them to be translated into proteins by the ribosomes. Adapted from Alberts, Wilson et al. 2008.

As soon as the nascent pre-mRNA is being produced, its 5' end is modified by the addition of a modified guanosine nucleotide (a methylated guanosine). This 5' capping allows mRNAs to be discriminated from other RNA molecules in the cell, favors the transport of mRNAs into the cytoplasm, is responsible for protecting the 5' end of mRNAs from exonuclease degradation and has also an important role in translation

initiation[1,2,3]. Eukaryotic genes include small portions of coding sequences (exons) interspersed by large portions of non-coding sequences (introns), all of which are transcribed into the pre-mRNAs. Through RNA splicing, the intron sequences are removed from the pre-mRNA allowing binding of the exon sequences and mRNA formation[1,2,3]. It is known that splicing can increase transcriptome diversity in a process designated alternative splicing, in which the same pre-mRNA can be spliced in different ways, generating different mRNA isoforms that will encode different proteins[1]. The final step of pre-mRNA processing, polyadenylation, is detailed below.

2| Cleavage and Polyadenylation

Excluding histone replication-dependent transcripts[4,6], nearly all eukaryotic pre-mRNAs are subject to polyadenylation in order to complete its maturation[3,5,6,7,8,9,10,11]. Polyadenylation consists of the polymerization of an adenosine tail in the pre-mRNA 3' end after specific endonucleolytic cleavage of the polyA site[3,5]. This polyA site corresponds to the cleavage site, which is usually a CA dinucleotide, and is the place where adenosine residues are polymerized by the polyA polymerase (PAP) forming a polyA tail. The added adenines are not encoded in the gene and therefore their addition depends on the recognition of a polyA signal (PAS) and binding of the polyadenylation factors in the pre-mRNA[1,4,5,6,7,8,9,11,12,13]. This 3' end processing of pre-mRNAs has a significant function since the polyA tail is responsible for protecting the 3' end of mRNAs from exonucleases, and therefore favors mRNA stability, contributes to mRNA transport into cytoplasm and influences protein translation[2,4,5,6,7,8,9,11,13].

The PAS is located approximately 10-35 nucleotides upstream of the polyA site, and consists in a hexamer sequence. This core hexamer element could be the canonical sequence, AAUAAA, or a similar variant[4,5,6,7,10,11,12,13,14]. Normally, the PAS variant sequences are less efficient or weaker[7,11,14] than the canonical one and, when there are more than one PAS in the pre-mRNA sequence, weaker variants are located in more proximal positions[6,8,9]. Computational studies show that about 50% of human mRNAs possess the canonical PAS, whereas PAS variant sequences account for ~36% of all recognized PASs[13]. Only 10% of genes do not harbor a recognizable PAS.

3| RNA processing proteins

Several protein factors are required for the 3' end maturation of pre-mRNAs and, normally, these factors are conserved in eukaryotes[4]. In metazoans, the core polyadenylation machinery (Figure 2) includes the following protein complexes: cleavage and polyadenylation specificity factor (CPSF), cleavage stimulation factor (CstF), cleavage factor Im (CFIm), cleavage factor IIm (CFIIm), PAP, Symplekin and RNAPII, more specifically the carboxy-terminal domain (CTD)[4,6,11,15]. The CPSF complex can be subdivided into CPSF160 (also known as CPSF1), CPSF100 (CPSF2), CPSF73 (CPSF3), CPSF30 (CPSF4), factor interacting with PAP (FIP1) and WD repeat domain 33 (WDR33)[16]. The CstF is composed by the following subunits: CstF77, CstF50 (CstF1) and CstF64[8,9]. The CFIm complex acts as a heterodimer consisting of CFIm-25 and either CFIm-68 or CFIm-59. The CFIIm complex is constituted by PCF11 and CLP1 subunits[17].

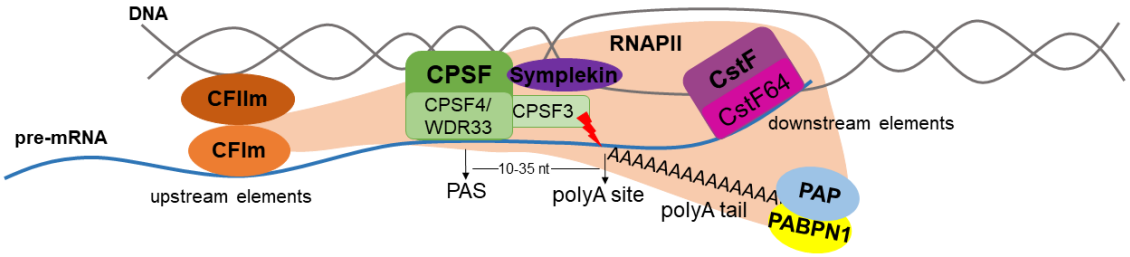


Figure 2: Core polyadenylation machinery | During polyadenylation, the polyA signal (PAS) is recognized by the cleavage and polyadenylation specificity factor (CPSF) subunits. CPSF3 subunit is responsible for the 3' end cleavage at the polyA site. The cleavage stimulation factor (CstF) subunits recognize downstream sequence elements. The cleavage factor IIm (CFIIm) and CFIm complexes are important for pre-mRNA recognition and polyA site selection, respectively. Symplekin displays a scaffolding function. The association between CPSF, polyA polymerase (PAP) and polyA binding (PAB) proteins results in the formation of the polyA tail. Adapted from Elkon, Ugalde et al. 2013.

In the pre-mRNA, the PAS (AAUAAA or variants) is recognized by the CPSF subunits, more specifically CPSF4 and WDR33[18,19,20] . The CstF64 subunit binds to U/GU-rich downstream sequence elements (DSE) situated 30 nucleotides downstream of the cleavage site and the cleavage reaction is catalyzed by the CPSF3[4,6,8,11,12,13,14]. It is also known that CFIm binds to the upstream UGUA element and plays a major role in the polyA site selection[17]. In fact, this core polyadenylation complex has also been shown to be important for the recognition of non-canonical PASs[21,22]. Regarding the core machinery, CFIIm has been poorly studied. However, a recent study demonstrated that this complex display high affinity for RNA, namely G-rich DSE, suggesting a role for CFIIm in pre-mRNA recognition[23]. Symplekin acts as a scaffolding protein that interacts with CPSF and CstF[17,24]. The polyA tail formation results from the cooperation between

CPSF, PAP and the polyA binding protein nuclear 1 (PABPN1). The addition of adenosine residues by the PAP is stimulated by PABPN1, which is also responsible for controlling the length of the tail by regulating interactions among the CPSF complex and PAP[4,9,12,14].

4| Alternative polyadenylation

When a gene possesses two or more PASs in its sequence, it can give rise to the formation of different mRNA isoforms through the choice of alternative PASs[4,5,6,7,9,25]. This process is designated alternative polyadenylation (APA) and plays an important role in transcriptome diversity, since one gene is capable of encoding numerous APA-derived mRNA isoforms with different 3' ends. Roughly, 70% of mammalian genes can be subject to APA since they hold multiple PASs in their sequences[26,27]. The relative strength of the polyadenylation factors modulate the use of alternative signals[6,9,11], but there have also been identified *cis*-auxiliary elements upstream and downstream of PASs suggesting that polyadenylation depends on the binding of several factors and small differences in their concentrations could affect the mRNA 3' end formation[5]. For instance, overexpression of 3' end processing factors may increase the use of proximal PASs[6]. Wherefore, the association of a wide variety of *trans*-acting protein factors and *cis*-acting sequence elements is probably responsible for the choice/use of a particular PAS[5,6,7,10,11].

Depending on the PAS localization (Figure 3), APA can be classified as coding region-APA (CR-APA) or 3' untranslated region-APA (3'UTR-APA)[6]. The CR-APA includes all APA events where the PASs are located upstream of the last exon. Since those events occur upstream of the stop codon, they give rise to mRNA isoforms with different coding regions that encode different proteins and, therefore, qualitatively affect gene expression[6,7]. The most common type of alternative polyadenylation is the 3'UTR-APA[9], which is characterized by the formation of mRNA isoforms with different 3'UTR lengths due to the existence of more than one PAS downstream the stop codon in the 3'-most exon[7,9]. In fact, this process happens in more than 50% of human genes[13]. Since PASs are located downstream of the stop codon, this type of APA does not affect the encoded protein but may quantitatively affect gene expression[4,5,6,7,10,11,14]. All 3'UTR-APA-derived mRNA isoforms share a region (Figure 4), the constitutive UTR (cUTR), located upstream of the proximal polyA site. While the region located downstream, which is designated as alternative UTR (aUTR), is only present in the longest isoform[8,9].

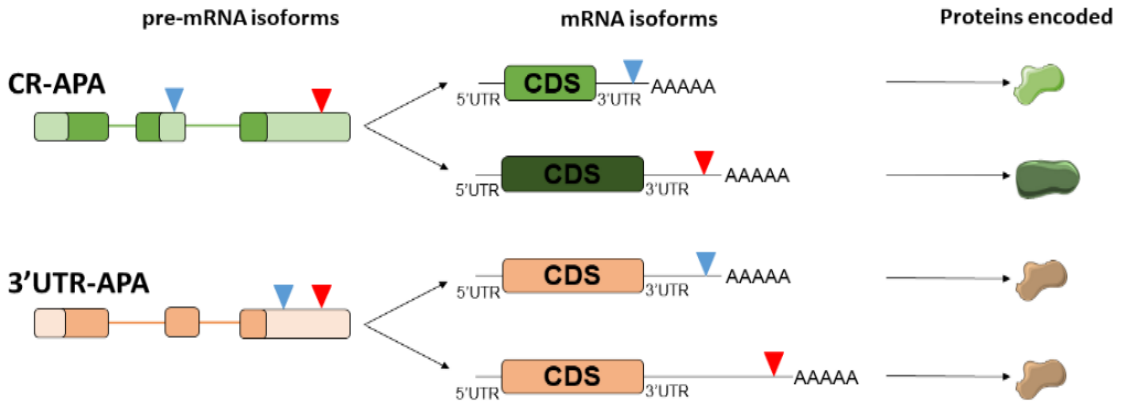


Figure 3: Different types of APA | CR-APA (coding region-alternative polyadenylation) is characterized by the presence of polyA signals (PASs) in upstream exons/introns, resulting in the expression of different proteins isoforms. Through the presence of PASs in the 3'-most exon, 3' untranslated region (3'UTR)-APA is responsible for the generation of mRNA isoforms that only differ in their 3'UTR. Dark boxes: exons. Light boxes: untranslated regions. Lines: introns. Blue head arrows: proximal PAS. Red head arrows: distal PAS. CDS: coding sequence. Adapted from Di Giammartino, Nishida et al. 2011.

5| 3' Untranslated Region-APA

The mRNA's 3'UTR can be very subject to regulation since they hold multiple *cis*-regulatory elements, such as AU-rich elements (AREs), RNA-binding protein (RBP) binding sites and micro-RNA (miRNA) binding sites[6,8,9,28]. Other important *cis*-regulatory elements present in the 3'UTR include: (1) upstream sequence elements (USE) such as the PAS, U-rich sequences and UGUA motifs; (2) DSE like U-rich sequences and GU-rich sequences[4,8,11]. It is well known that, mRNA isoforms with alternative 3'UTRs (Figure 4) may not only show different stability and localization but also different translation efficiencies[5,25]. Recently, it was also demonstrated that alternative 3'UTRs are capable of regulating protein function[29] and localization[28,30]. The stability, localization and translation of mRNAs can be affected by binding of RBPs to *cis* regulatory elements present in the 3'UTR[5,7,8,9,10,11,14,28,31]. Naturally, those *cis* elements are likely to be in higher amount on longer 3'UTRs[5]. Similarly, 3'UTRs with greater length are more susceptible to undergo negative regulation since they comprise more binding sites for miRNAs, which may degrade the mRNA or inhibit its translation[5,7,32]. For instance, the luciferase expression obtained by the longer transcript of Huntingtin Interacting Protein 2 (*Hip2*) was reestablished to the levels achieved by the shorter isoform after mutation of miR-21 and miR-155 target sites present in the longer 3'UTR[33]. Therefore, avoidance of miRNA regulation by shorter 3'UTRs allows them to yield larger protein quantities[5,7,9,14,28,33]. Thus, 3'UTR-APA influence the transcript outcome and, consequently, gene expression[5].

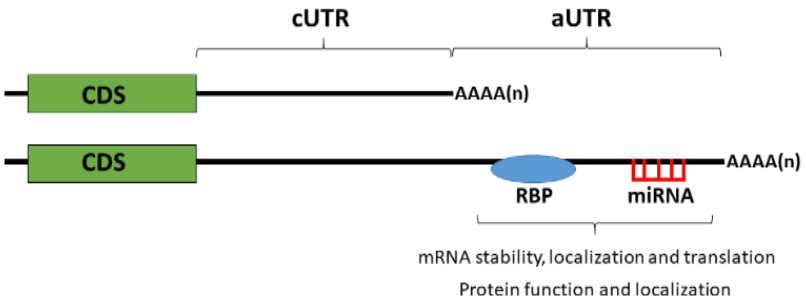


Figure 4: 3'UTR-APA impact on mRNA | 3'UTR-APA influences gene expression since longer APA-derived mRNA isoforms are more prone to suffer negative regulation. The constitutive UTR (cUTR) is the upstream region of the proximal polyA site and is present in the two isoforms. The region downstream is designated as alternative UTR (aUTR) and is exclusive to the long isoform. CDS, coding sequence; miRNA, microRNA; RBP, RNA-binding protein. Adapted from Tian and Manley 2013.

5.1| Specific patterns of 3'UTR-APA

Several studies have shown notorious tissue-specific patterns of APA in human genes[5,6,25,34]. In fact, APA regulation is expected to play a critical role during the different cellular states, like proliferation and differentiation, and even pathological events, such as cancer, since all these processes are highly dependent on the regulation of numerous genes[4,5,6,7,8,14].

The presence of shorter 3'UTRs is characteristic of some diseases, dedifferentiation programs and specific cell growth states[6,9,25]. Actually, 3'UTR shortening is deeply related to states of proliferation, probably because it leads to the increase of the protein amount[5,9,33]. In fact, the *Hip2* example previously mentioned was the first study that linked APA to proliferative cell states, since the choice of proximal PASs allowed an increase in protein production due to the avoidance of miRNA repression[9,33]. Accordingly, the choice of proximal PASs is strongly associated with activated T cells[33] and during spermatocytes development[35]. Concerning T cells, even though the shift from distal to proximal signals was well detected 48 h upon activation, it was not so obvious after 6 h of activation[33]. This observation indicates that the choice of the PAS does not occur straightaway, which is in agreement with the need to transcribe the polyadenylation factors[5]. Regarding cancer cell lines, it was already described that they are more likely to express mRNA molecules with shorter 3'UTRs[36]. This APA pattern of cancer cells was not surprising since it allows oncogenes with shorter 3'UTRs to avoid degradation by miRNAs[5,7] and/or destabilizing RBPs, such as ZFP36[28] (also known as TTP), Roquin-1 and Roquin-2[37].

In contrast, states of later development and differentiation are associated with the use of distal PASs generating transcripts with longer 3'UTRs[6,7,8,25]. In fact, this pattern of

3'UTR lengthening has already been described in developing mouse embryos[38], neurological tissues[39] and ovulated oocytes and zygotes[40]. Luciferase assays showed a decrease in protein production correlated with the use of distal PASs. This suggests the existence of negative posttranscriptional regulators downstream of proximal PASs and indicates that more differentiated tissues need to be more tightly regulated[5].

5.2| Cell and gene specific patterns of 3'UTR-APA

In a recent study of the Gene Regulation Group, the expression pattern of the APA-derived 3'UTR mRNA isoforms of classical Ras homologue (Rho) guanosine triphosphatases (GTPases) was analyzed[41]. For this, mRNA isoforms expression of Rac family small GTPase 1 (*Rac1*), Cell division cycle 42 (*Cdc42*) and Ras homolog family member A (*RhoA*) was assessed in many brain primary cell types, such as astrocytes, microglia, oligodendrocytes, cortical and hippocampal neurons[25]. These GTPases play a major role in numerous cellular processes, such as neuronal development, cell growth, motility and adhesion, among others. In this study it was showed that, during neurite growth of cortical neurons, the expression of the longest 3'UTR mRNA isoform of *Rac1* increases and this pattern does not happens on hippocampal neurons, suggesting that APA in *Rac1* is gene and cell-type specific[25]. They also observed that, contrary to hippocampal, in cortical neurons the longest *Rac1* mRNA localizes in the neurites, while the shortest isoform localizes in the soma. This result allowed concluding that, in cortical neurons, the longest isoform plays a crucial role in targeting *Rac1* mRNA to neurites[25]. Additionally, they observed that the long 3'UTR *Rac1* mRNA isoform is important for the neurites length and morphological features. Moreover, they suggest that *Rac1* longer 3'UTR mRNA has an essential physiological function in cortical neurons[25]. Overall, their findings indicate that APA is not only a tissue-specific mechanism but also gene- and cell- specific.

5.3| The role of 3'UTRs on protein function and localization

Mayr and Berkovits[30] demonstrated in human cell lines that protein localization could be regulated by 3'UTR-APA, regardless of mRNA localization. First, they observed that the cluster of differentiation 47 (CD47) protein encoded by the longest 3'UTR mRNA isoform localizes predominantly in the cell surface, while CD47 protein derived from the shorter 3'UTR mRNA isoform mainly localizes in the endoplasmic reticulum. Then, they found that binding of the RBP Embryonic Lethal Abnormal Vision-like 1 (ELAVL1, also

known as Human antigen R (HuR)) to the long 3'UTR of *CD47* enables the transport of the CD47 protein to the plasma membrane, in a post-translational way[30]. Furthermore, they concluded that HuR recruits SET to the long 3'UTR of *CD47* allowing local recruitment of SET to the translation site. After *CD47* mRNA translation, SET interacts with the C terminus driving CD47 protein to the cell surface through active RAC1[30]. With these results, the authors proposed a model (Figure 5) of 3'UTR-dependent protein localization suggesting that alternative 3'UTRs from APA-derived mRNA isoforms have an impact on protein subcellular localization. In addition, they also showed that *CD47* alternative 3'UTR isoforms display independent functions regarding cell migration and survival, meaning that 3'UTR-APA can also regulate protein function[30].

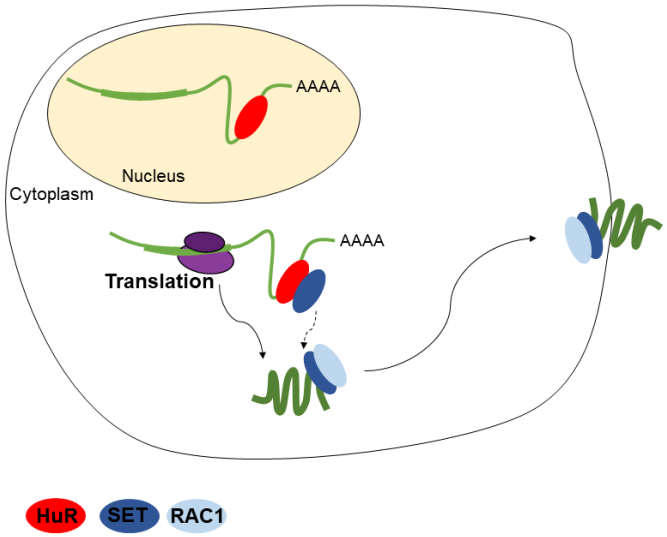


Figure 5: Model of 3'UTR-dependent protein localization | HuR binds to the long 3'UTR of *CD47* and recruits SET allowing local recruitment of this protein to the translation site. Upon translation, SET binds to CD47 protein C' terminal, leading it to the cell surface through interaction with RAC1. Adapted from Berkovits and Mayr 2015.

Additionally, the role of long 3'UTRs and 3'UTR-mediated protein-protein interactions were investigated in a more recent study[29]. Mayr et al have first observed an increase in the expression levels of the long Baculoviral IAP repeated containing 3 (*BIRC3*) 3'UTR isoform in malignant B cells derived from chronic lymphocytic leukemia (CLL). In order to understand the role of this long 3'UTR isoform, they analyzed 3'UTR-dependent protein interactors of BIRC3[29]. They concluded that the long 3'UTR is responsible of recruiting Ras like proto-oncogene A (RALA) and IQ motif containing GTPase activating protein 1 (IQGAP1) to the BIRC3 translation site. The results showed that the formation of a protein complex, consisting of BIRC3, RALA and IQGAP1, allows regulating B cell migration mediated by CXCR4, a crucial process for cell survival in

CLL[29]. Overall, they concluded that alternative 3'UTRs from APA-derived mRNA isoforms allow proteins to display additional functions in a 3'UTR-dependent manner[29].

6| Regulation of alternative polyadenylation

Genes that express APA isoforms are under the control of tight regulatory mechanisms[4]. It is well stablished that APA patterns can vary due to changes in the polyadenylation factors levels[4,6] or due to physiological cellular conditions[9], such as cellular proliferation and differentiation. In addition, APA is likely to be regulated by many other layers, from chromatin modifications to gene-specific RBPs[6].

Differential expression of the polyadenylation factors is an important regulatory APA mechanism. During B cell maturation, upregulation of CstF64 leads to a distal-to-proximal switch resulting in the shift of the membrane-bound to the secreted form of immunoglobulin M (IgM) heavy chain[42,43]. Furthermore, bioinformatics analysis revealed a global shortening of the 3'UTR lengths in seven types of cancer, with five of them displaying upregulation of CstF64 mRNA levels[44]. During induced pluripotent stem (iPS) cells generation of different cell types there is an increase in the expression levels of CstF and CPSF subunits, as well as the other core components of the polyadenylation machinery, resulting in 3'UTR shortening[45]. Moreover, in differentiated embryonic tissues it was observed a downregulation of the same factors and a global trend of 3'UTR lengthening[45].

Research data have shown that the chromatin structure plus its epigenetic marks are able to regulate APA and, therefore, the output of gene expression[9]. The presence of condensed DNA, also known as heterochromatin, causes RNAPII pausing during transcription, which leads to an increase in the use of proximal PASs[46]. However, this APA pattern is modulated according to the physiological cellular state. For example, proximal PAS selection increases in the presence of euchromatin during spermatogenesis[47]. Regarding nucleosomes distribution, a set of studies observed an extensive lack of nucleosomes near the polyA sites[6,8,9]. Furthermore, these studies also suggest that highly used polyA sites have higher nucleosome density in sequences downstream of the polyA site. Nevertheless, is not yet clarified whether the epigenetic marks stimulate specific APA patterns or whether it is the choice of the polyA site that prompts specific chromatin marks[6].

As a co-transcriptional mechanism, APA can be regulated by several factors of the transcription process[9]. Regarding this, the elongation rate of the RNAPII plays a

major role in the selection of the polyA site[6,8,9]. The expression of RNAPII with a slower elongation rate, in a mutant strain of *Drosophila melanogaster*, causes increased usage of the proximal polyA site of *polo* gene[48]. This pattern, also observed for other genes, happens because, with a slower RNAPII, the proximal PAS is more subjected to the 3' end processing machinery while the distal signal is not transcribed[6]. In agreement, the presence of *cis* elements that slow RNAPII elongation rate, like G-rich elements, favors the usage of the proximal polyA site[49]. Moreover, on highly expressed genes there is an abundance of shorter 3'UTR mRNA isoforms together with a paused RNAPII at proximal PAS[9].

Through interaction with polyadenylation factors, splicing factors are also able to regulate APA[6,9]. For instance, U1 small nuclear ribonucleoprotein (snRNP) interacts with CPSF1 and when U1A, the U1 snRNP-free form, binds to upstream auxiliary sequences, increases the association between the canonical PAS and CPSF[50]. It was first shown that U1 snRNP represses the usage of polyA sites as a result of PAP inhibition[51]. Later on, it was found that absence of U1 snRNP causes cryptic polyA site activation around transcription start sites[52] and with small amounts of this splicing factor there is an enhancement of proximal polyA site usage[53]. It is well stablished that the IgM heavy chain expresses two mRNA isoforms generated by CR-APA, with the longest one encoding a membrane-bound form of the antibody while the secreted form is encoded by the shortest isoform[9]. Binding of U1A to GU-rich sequences downstream of the proximal polyA site inhibits binding of the polyadenylation factor CstF64, preventing the shift of the membrane-bound to the secreted form of IgM during B cell activation[4,7].

During 3' end processing, the binding of RBPs regulates the expression of APA-derived mRNA isoforms. This happens because RBPs, depending on the biological conditions, are able to recruit 3' processing factors, which promote the polyA site usage, or compete with them, inhibiting the use of the polyA site[8,14]. For instance, the polypyrimidine tract binding protein (PTBP1 or hnRNPI) is able to bind to the pyrimidine-rich DSE and compete with Cst64[54], or it can increase cleavage, through the conscription of CstF and PAP, because it stimulates binding of the heterogeneous nuclear ribonucleoprotein H (hnRNPH) to G-rich elements[4,6]. Another RBP who is also implicated in the modulation of APA patterns is the neuro-oncological ventral antigen 2 (NOVA2). This splicing factor is capable of regulating the choice of the polyA site, in the brain, depending on the distance to which it binds to the PAS. In other words, when this RBP binds closely to the PAS inhibits polyadenylation, but when it binds far from the PAS enhances the use of the polyA site[6,8,9,14]. Analysis of twelve genes with alternative PASs

in their 3'UTRs revealed that, in nine of them, NOVA2 stimulated the usage of distal PASs, giving rise to isoforms with longer 3'UTRs[55].

Overall, APA is an extremely regulated process that accounts with the participation of several molecular players and layers of regulation, which varies in a gene, tissue and cell-type specific manner.

7| Apoptosis and the Bcl-2 protein family

There is an essential biological mechanism, regulated by many factors, through which cells suffer programmed cell death[56,57]. This fundamental process, called apoptosis, is accountable for the removal of misregulated or mutated cells and, consequently, for maintaining tissue homeostasis and allowing the establishment of the correct morphogenesis over embryonic development[56,58,59]. Apoptosis happens when caspases, from the cysteine proteases family, are activated resulting in the disposal of cellular components. Then, phagocytes are responsible for the removal of the apoptotic bodies constituted by the destroyed cell components[56]. Whenever cells are induced to apoptosis, caspases activation is regulated, intrinsic or extrinsically, by various mechanisms[56,58].

Regarding the intrinsic regulation of apoptosis, there is a very important group of proteins called B-cell lymphoma 2 (Bcl-2) family[60]. This protein family encompasses anti-apoptotic members, like Bcl-2 and Myeloid cell leukemia 1 (Mcl-1), and pro-apoptotic members, such as Bcl-2 homologous antagonist killer 1 (Bak) and Bcl-2-associated X protein (Bax). Those two antagonistic groups, more specifically their balance, are capable of inducing the activation of caspases through the regulation of cytochrome c released by the mitochondria[56,58,59,60,61,62].

7.1| Mcl-1 function and structure

Mcl-1 is one of the anti-apoptotic members from the Bcl-2 protein family, being involved in the regulation of intrinsic apoptosis[56,60,63]. As a pro-survival member, Mcl-1 promotes cell viability and this anti-apoptotic function is due to its ability to bind and sequestrate the pro-apoptotic members of the Bcl-2 family (Figure 6), Bak and Bax[56,58,64]. As pro-apoptotic proteins, Bak and Bax can induce pore formation in the outer mitochondria membrane (OMM) which enables the release of cytochrome c into the cytoplasm[56,58,61,65]. The binding of Mcl-1 to these two proteins blocks their pro-apoptotic

function. Indeed, it is known that cell viability can be increased by the presence of conditions that promote a slower Mcl-1 degradation[56,63]. In the presence of an apoptotic stimuli occurs the activation of a set of members of the Bcl-2 protein family, the BH3-only proteins, such as BCL-like protein 11 (BIM) and phorbol-12-myristate-13-acetate-induced protein 1 (NOXA). These proteins are able to bind the Bcl-2 pro-survival members and, thus, prevent them to interact with the pro-apoptotic proteins[66].

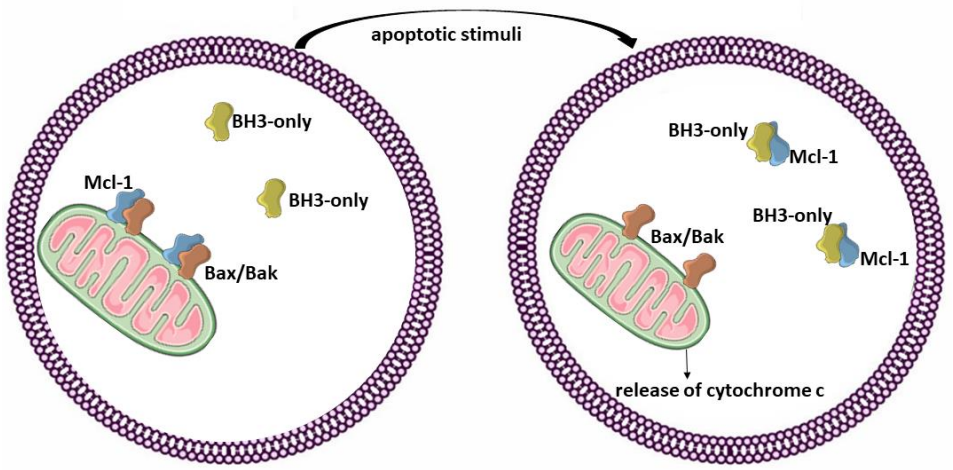


Figure 6: Apoptosis regulation by the Bcl-2 protein family | An apoptotic signal triggers the activation of BH3-only proteins that, in turn, bind to the pro-survival members. Pore formation in the outer mitochondria membrane by pro-apoptotic proteins, Bak and Bax, induces apoptosis through the release of cytochrome c into the cytoplasm. Apoptosis avoidance happens due to binding of the pro-survival members (such as Mcl-1 and Bcl-2) to Bak and Bax proteins. Adapted from Xiang, Yang et al. 2018.

Several studies have highlighted the main function of Mcl-1 on cell survival and many have shown that its distribution is tissue specific[56]. Despite both *MCL1* mRNA and protein display a short half-life[56,59], it is well known that during implantation, in mouse embryogenesis, *Mcl1* depletion is lethal[67]. Besides its importance for embryogenesis, it is also established that Mcl-1 is required for preservation and development of both T and B-lymphocytes[68]. Furthermore, Mcl-1 is important for neural progression[69] and for the regulation of apoptosis in macrophages[70] and neutrophils[71]. In addition, this protein also shows the ability to retard cell cycle progression when interacts with cyclin dependent kinase 1 (CDK-1)[72], proliferating cell nuclear antigen (PCNA)[73] and checkpoint 1 protein (CHK-1)[74].

In humans, *MCL1* gene comprises three exons and is located on chromosome 1q21[59,60]. During pre-mRNA processing and, more precisely, through splicing, *MCL1* loses his introns and generates the full length *MCL1* mRNA that possesses the three exons. This mRNA isoform, also known as Mcl-1L, which encodes the classic Mcl-1

protein, comprises 350 amino acids[59,60,75]. Mcl-1L contains similarity regions among Bcl-2 family members, which are designated as Bcl-2 homology (BH) domains. More specifically, Mcl-1L mRNA isoform-derived protein (Figure 7) is constituted by two PEST (peptide sequence rich in proline (P), glutamic acid (E), serine (S), threonine (T)) domains, one transmembrane (TM) domain and three BH (BH1-3) domains[59,60,75]. *MCL1* gene is also able to generate another two shortened mRNA isoforms, through alternative splicing, called Mcl-1S/ΔTM and Mcl-1ES[56] . Mcl-1S/ΔTM has exon 2 skipping and, therefore, encodes a shorter Mcl-1 protein that is only constituted by 271 amino acids[56,59,63,76]. Mcl-1 protein encoded by this shorter splicing-derived mRNA isoform (Figure 7) retains the PEST and BH3 domains but lacks the TM, BH1 and BH2 domains[56,59,60]. Mcl-1ES (Figure 7) lacks a small region of exon 1 resulting in the formation of a Mcl-1 protein (composed by 197 amino acids) without the PEST domains, but with the TM domain and the three BH domains[77]. As these two splicing-derived mRNA isoforms are inefficient in the sequestration of the pro-apoptotic Bcl-2 members, Bax and Bak, they have a pro-apoptotic function. In fact, they promote apoptosis through their capacity of binding to Mcl-1L isoform and causing its sequestration[56,76].

Figure 7: Splicing isoforms of Mcl-1 protein | Classical anti-apoptotic Mcl-1 protein comprises two PEST domains, three Bcl-1 homology (BH) domains (BH1-3) and one transmembrane (TM) domains. Mcl-1S/ΔTM lacks the BH1, BH2 and TM domains, while Mcl-1ES lacks PEST domains. Adapted from Thomas, Lam et al. 2010

7.2| The role of Mcl-1 in T cells and cancer

Through the intrinsic programmed cell death pathway, Bcl-2 family proteins are important for the regulation of T lymphocyte function and development at different stages[78]. Recent data suggest that, stimulation of T cell survival by the anti-apoptotic Bcl-2 proteins does not depend on single-member participation but on a collaboration among several members[65]. The anti-apoptotic Mcl-1 protein plays an important role in the survival and maintenance of naïve and activated T lymphocytes, as well as other T cell populations[78]. On early stages, Mcl-1 is essential for the maturation of T cells. In fact, loss of mature T cells is a consequence of induced deletion of Mcl-1[79]. Moreover, this protein is constitutively expressed during all developmental stages of T lymphocyte[78].

Regarding regulatory T (T_{reg}) lymphocytes, the expression of Mcl-1 is essential for the survival of these cells[80]. Indeed, Mcl-1 deletion results in loss of T_{reg} cells, which is associated with the onset of autoimmunity[80].

With such an important role for cell survival, it is not surprising that Mcl-1 deregulation results in numerous pathologies, such as cancer, rendering it a promising target for therapeutic manipulations[62,75]. Several studies have already established that Mcl-1 overexpression is a common feature detected in numerous human cancers[57,60,62], like pancreatic, skin, colon, renal, lung, breast, ovarian, prostate, head and neck cancer, as well hepatocellular carcinoma, multiple myeloma, leukemia and lymphoma[75]. In fact, it is already known that resistance to some anticancer therapies is induced by Mcl-1 overexpression in cancer cells[60,62], since it not only allows avoiding cell death, but also protects cancer cells against toxic factors[75]. Since this protein plays a major role in cancer maintenance and progression, several cancer therapeutics are based on Mcl-1 depletion in cancer cells. Nowadays, there are cancer therapies based on the combination of conventional therapies with Mcl-1 antisense oligonucleotides (ASO), since loss of Mcl-1 makes cancer cells sensitive to first-line treatments[60,75].

7.3| *MCL1* 3'UTR-APA

The Gene Regulation group has discovered that *MCL1* is regulated by APA in human T cells generating two mRNA isoforms (Figure 8), named pA1 (the shortest) and pA2 (the longest), independently of splicing. These two *MCL1* APA-derived mRNA isoforms are generated by usage of the PAS1 or PAS2 and only differ in their 3'UTR length, meaning that both encode the same protein. Our laboratory found that the longest pA2 mRNA isoform is typically more expressed than the shortest. However, it was demonstrated by luciferase assays that the *MCL1* shorter 3'UTR is more efficient in protein production than the longest 3'UTR. Additionally, recent results from the group have clearly showed that Mcl-1 protein encoded by pA1 mRNA isoform displays different subcellular localization when compared to Mcl-1 protein derived by the longest *MCL1* mRNA (pA2). However, the regulation, function and the mechanism(s) responsible for the localization of these two *MCL1* APA isoforms are still to be uncovered.

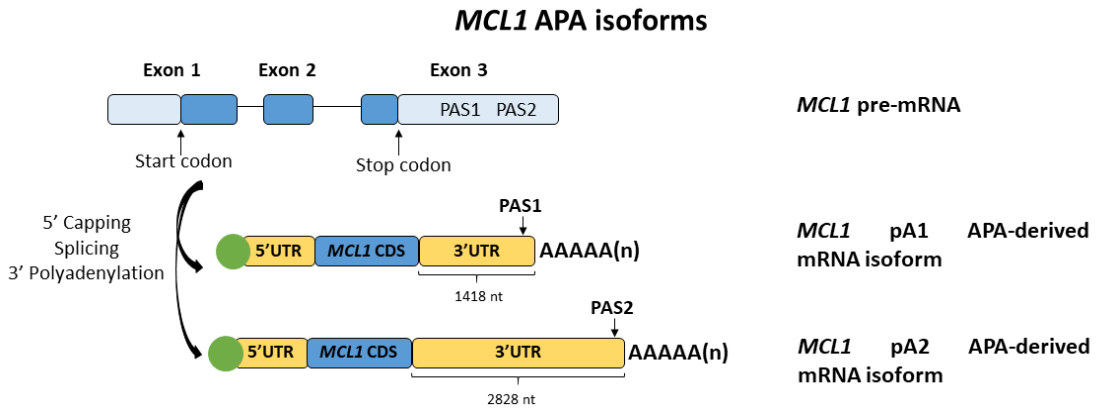


Figure 8: *MCL1* APA-derived mRNA isoforms | By alternative polyadenylation, *MCL1* gives rise to two mRNA isoforms that only differ in the length of the 3'UTR. The expression of the shortest isoform (pA1) occurs due to the choice of the proximal polyA signal (PAS1), whereas the use of the distal PAS (PAS2) generates the longest isoform (pA2). Both isoforms code for the same Mcl-1 protein. CDS, coding sequence; Light blue: untranslated regions (UTRs)

8| Genome editing with CRISPR/Cas9 technology

The clustered regularly interspaced short palindromic repeats/CRISPR-associated proteins (CRISPR/Cas) system is a defense mechanism that provides adaptive immunity to all archaea and approximately half of bacteria[81,82]. Through the existence of this CRISPR *loc*i, archaea and bacteria are able to inactivate and degrade plasmids and bacteriophages[83]. All CRISPR *loci* have the same structure pattern, which is characterized by the presence of short repeats alternated by spacer sequences, and, typically, CRISPR arrays are situated nearby the *cas* genes[81,82]. This system was identified as a prokaryotic immunity mechanism due to the finding that bacteria with CRISPR spacer sequences, equal to mobile genetic elements (MGEs) of viruses, are resistant to the corresponding virus[82]. It is well known that when a virus invades a bacterium, a virus-derived DNA fragment is integrated into the CRISPR *loci* creating a CRISPR spacer sequence. If the same virus invades the same bacteria over again, the transcription of the CRISPR spacers will result in the generation of a CRISPR RNA (crRNA) complementary to the invasive nucleic acid[82]. This target-specific interference allows the degradation of the foreign DNA and, therefore, is known as a RNA-guided adaptive immunity mechanism. The processes involved in this CRISPR-mediated immunity are divided into three stages[82]. The first one is the acquisition stage, where the foreign DNA is shattered into a new protospacer, which undergoes processing to give rise to a new spacer that will be integrated in the CRISPR locus[81,82]. The new spacers are detected due to the presence of a specific protospacer adjacent motif (PAM) and they are always integrated with a PAM-dependent orientation at the leader end of the

CRISPR locus[82], resulting in the formation of a timeline of invasive events[81,82]. The PAM sequence is specific of each CRISPR system. For instance, 5'-NGG is the specific PAM sequence of the CRISPR-Cas9 system from *Streptococcus pyogenes*[84]. Then the expression stage is followed, which is characterized by the transcription of the CRISPR locus and generation of crRNAs through precursor crRNA (pre-crRNA) transcript processing[81,82]. This pre-crRNA processing occurs through the endoribonucleolytic cleavage of the short repeats between the spacer sequences, leading to the separation of the individual crRNAs[82]. Finally, there is the interference stage where the crRNAs gather with the Cas proteins forming CRISPR ribonucleoprotein (crRNP) complexes that will recognize and cleave the foreign DNA complementary with the crRNA, degrading it[81,82].

There is a high diversity regarding CRISPR/Cas systems and they have been categorized into three types according to their specific Cas protein. The type I systems have the Cas3 nuclease-helicase, the type II contain the Cas9 nuclease and the type III have the Cas10[82]. Type I and III systems are characterized for containing crRNP with several Cas subunits, while the crRNPs from type II, known as Cas9 complexes, only comprise one Cas9 protein. The CRISPR locus from type II systems comprises a gene encoding for a transactivating crRNA (tracrRNA), apart from the CRISPR locus and the *cas* genes[82]. This tracrRNA holds a sequence complementary to the repeat sequence in the pre-crRNA, allowing them to form a hybrid and thus facilitating the processing of the pre-crRNA[84]. The ribonuclease RNase III is responsible for the pre-crRNA processing because it is able to recognize the double-stranded region, cleaving it[82]. Then, the mature crRNA-tracrRNA stay attached with the DNA endonuclease Cas9 protein that will cleave the target DNA[82,83], complementary to the crRNA, in a blunt manner and precisely three nucleotides from the 3' end of the target DNA[81,82,85]. It was described, in *Streptococcus pyogenes* type II system (Figure 9), the existence of crRNA fused to tracrRNA, which was designated as a single-guide RNA (sgRNA)[81,82,83,85]. Cas9-mediated genome editing occurs through the formation of double-stranded breaks (DSBs) at the target DNA followed by DNA damage repair mechanisms[84]. Regarding the mechanisms of DNA damage repair, there are two main pathways: non-homologous end joining (NHEJ) and homology-directed repair (HDR)[84]. The HDR pathway requires a repair template, exogenously introduced, to generate highly precise modifications such as mutations in a single nucleotide. This mechanism happens less often but features high fidelity. In the lack of a repair template, DSBs are rejoined via NHEJ repairing pathway resulting in random insertions or deletions (indels) of nucleotides[84,85]. When this

occurs in protein-coding regions can result in frameshift mutations and premature stop codons and, therefore cause gene knockouts.

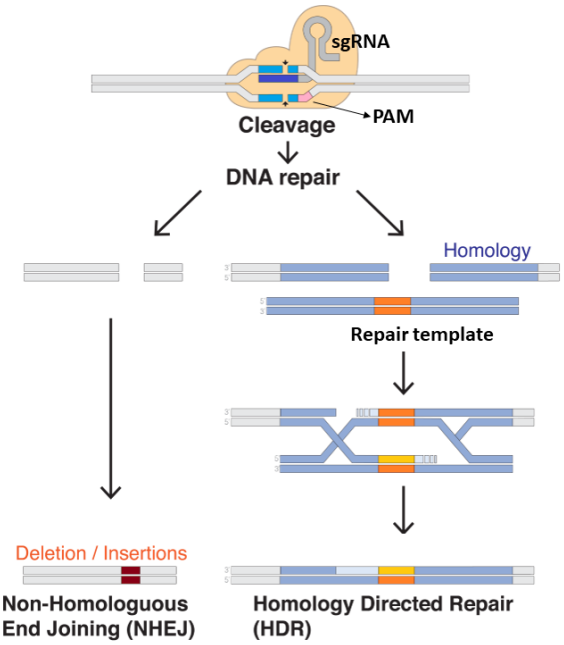


Figure 9: CRISPR/Cas9 genome editing | A single-guide RNA (sgRNA) drives the DNA endonuclease Cas9 protein to a complementary genomic sequence. Then, Cas9 protein cleaves the target DNA by detecting a specific protospacer adjacent motif (PAM) at the 3' end. Cas9-mediated genome editing results from the DNA damage repair mechanisms. The homology-directed repair (HDR) pathway generates highly precise modifications due to a repair template. The non-homologous end joining (NHEJ) mechanism repairs DNA without a repair template resulting in random insertions or deletions. Adapted from https://en.wikipedia.org/wiki/CRISPR_gene_editing; Author: Mariuswalter.

The discovery and application of the CRISPR/Cas9 system revolutionized the genetic engineering field since it is the most precise and sequence-specific technology to perform genome editing[82,85,86]. The RNA-guided Cas9 nuclease quickly become a powerful tool used to perform genome engineering in numerous cell types and organisms, as it is highly efficient, specific and, above all, easy to apply[84].

Objectives

Most of the studies conducted so far concerning *MCL1* gene expression regulation focused in both splicing and protein regulation. However, nothing was known about the role of the two *MCL1* mRNA isoforms generated by APA in Mcl-1 function and localization. Therefore, the aim of this study was to dissect the functional role of the two *MCL1* APA isoforms (pA1 and pA2) in human cells. In this regard, the specific aims of this work were to:

1) Study the physiological function of both pA1 and pA2 *MCL1* APA-derived mRNA isoforms.
2) Understand the influence, and associated molecular mechanism, of the *MCL1* alternative 3'UTRs on Mcl-1 protein localization.

Materials and Methods

1| sgRNA Design

For the sgRNAs design, we took advantage of available online tools listed in the Guide Design Resources of the Zhang Lab site (https://zlab.bio/guide-design-resources), namely, MIT's Zhang site, until it closes, and then CHOPCHOP site. The sgRNA score and the number of off-targets were the main criteria taken into account for the selection of each sgRNA.

2| CRISPR/Cas9 genome editing strategy

To delete *MCL1*-APA derived mRNA isoforms (*MCL1*ΔpA1 and *MCL1*ΔpA2), a two-sgRNA genome editing strategy was used. This approach consisted of using two sgRNAs flanking each of the PASs responsible for the generation of the two *MCL1* mRNA isoforms. Each sgRNA was obtained by the annealing of a forward (F) and a reverse (R) oligonucleotide, ordered at Sigma. Besides the desired sgRNA target sequence (20 bp), the F and R oligonucleotides also contained the *Bsm*BI restriction enzyme overhangs that allowed its cloning into the *Bsm*BI-digested all-in-one lentiviral vector with puromycin resistance (plentiCRISPRv2-Puro, a gift from Feng Zhang[87]; Addgene plasmid #52961) or with blasticidin resistance (plentiCRISPRv2-Blast, a gift from Mohan Babu; Addgene plasmid # 83480). Except for those cloned in plentiCRISPRv2-Blast, all cloned sgRNAs in plentiCRISPRv2-Puro were previously made.

For *MCL1*ΔpA1, one sgRNA (PAS1_sgRNA1) was designed to target a genomic sequence upstream of the PAS and the other sgRNA (PAS1_sgRNA2) was designed to target a downstream sequence, allowing to delete a ~70 nucleotide (nt) region that includes the proximal PAS (PAS1). For *MCL1*ΔpA2, two different pairs of sgRNAs flanking the distal PAS (PAS2) were used, the first pair (PAS2_sgRNA1 and PAS2_sgRNA2) generates a deletion of around 300 nt and the second pair (PAS2_sgRNA1 and PAS2_sgRNA3) deletes a ~380 nt region. In both cases, the deleted region includes the distal PAS.

The following oligos were used to obtain the sgRNAs : PAS1_gRNA1-F (5'-caccgGGCTTGCTTGTTACACACAC-3') and PAS1_gRNA1-R (5'-aaacGTGTGTGTAACAAGCAAGCCc-3'); PAS1_gRNA2-F (5'-caccgCAATGCAAAAACTTGCAACA-3') and PAS1_gRNA2-R (5'-aaacTGTTGCAAGTTTTTGCATTGc-3'); PAS2_gRNA1-F (5'-caccgGTTCGGGCAAATCCTCCAAA-3') and PAS2_gRNA1-R (5'-aaacTTTGGAGGATTTGCCCGAACc-3'); PAS2_gRNA2-F (5'-caccgTCAAATACAGGGTGTGATAT-3') and PAS2_gRNA2-R (5'-

aaacATATCACACCCTGTATTTGAc-3'); PAS2_gRNA3-F (5'-caccgTCAGCTGAGCAAATATGTAC-3') and PAS2_gRNA3-R (5'-aaacGTACATATTTGCTCAGCTGAc-3').

3| LentiCRISPRv2 Plasmid Digestion

Digestion of the plentiCRISPRv2-Puro and plentiCRISPRv2-Blast with *Bsm*BI restriction enzyme allowed the insertion of the desired sgRNA sequence. In a final volume of 50 µL, 3 µg of plentiCRISPRv2 were digested, during 2 h 30 min at 55ºC, with 1 µL of *Bsm*BI (10 000 U/mL; NEB®) in 1x NEBuffer™ 3.1 (NEB®). Thereafter, the digested lentiviral vector was loaded on a 0.8% agarose gel, visualized on the Gel Doc™ XR+ System (Bio-Rad), and the corresponding DNA band was cut and purified.

4| Agarose Gel Electrophoresis

For all electrophoresis, an agarose gel was prepared with 1x Tris-acetate-EDTA (TAE) buffer and SYBR® Safe DNA Gel Stain (Invitrogen). Whenever necessary, 6x loading dye (NEB) was added to the samples to monitor their migration during the electrophoresis. Ten µL of GeneRuler DNA Ladder Mix (Thermo Scientific) was used as a molecular weight size marker in all electrophoresis. All agarose gels were visualized on the Gel Doc™ XR+ System (Bio-Rad).

5| Sephadex DNA Purification

Previously, agarose gel bands should be extracted from an agarose gel, placed into Eppendorfs and stored at -80ºC for at least 20 min. Thereafter, 800 µL of Sephadex G-50 (GE Healthcare Life Sciences) resin solution were added to a column followed by a centrifugation of 4 min at 2200 g. Afterwards, the packed sephadex resin columns were placed into new Eppendorfs. The extracted agarose bands were heated during 4 min at 42ºC and the liquid fraction and the agarose were transferred into the column with the packed sephadex. After a centrifugation at 4400 g during 10 min, the liquid containing the DNA was collected in the Eppendorf tube and the columns with the sephadex were discarded. To precipitate DNA, 1 µL of glycogen, 1/3 of 10 M ammonium acetate ($NH_4CH_3CO_2$) and 10/3 of absolute ethanol were added to the samples followed by a centrifugation of 5 min at 14 000 g. The DNA pellet was washed with 1 mL of 70% ethanol. After centrifuging at 14 000 g during 5 min, the pellet was air-dried and eluted with the desired volume of nuclease-free water (HyClone™ HyPure Molecular Biology Grade Water; GE Healthcare Life Sciences).

6| LentiCRISPRv2-sgRNA Plasmids Cloning

For a final volume of 50 µL, 2 µM of each F and R oligo, with the desired sgRNA target sequence, were annealed in 10x T4 DNA Ligase Buffer (Thermo Fisher Scientific), at 95ºC during 5 min followed by a gradual cooling to room temperature with a rate of 5ºC/min. For the ligation, a mixture with 50 ng of *Bsm*BI-digested lentiCRISPRv2 vector and 1 µL of a 1:10 diluted annealed oligos (insert) solution was prepared. All ligation mixtures were prepared in a final volume of 20 µL. In addition to the insert and the vector, 1 µL of T4 DNA Ligase (5 U/µL; Thermo Fisher Scientific) enzyme and 10x T4 DNA Ligase Buffer was added to the Vector+Insert ligation mixture. Following an incubation period of 2 h at 22ºC, the enzyme was heat inactivated during 5 min at 70ºC. As a digestion control of the vector, a mixture without insert and T4 DNA Ligase enzyme (Vector without Ligase) was prepared. To control the binding of the digested vector with itself, a mixture only without insert (Vector with Ligase) was also made.

7| Competent Bacterial Transformation

Chemically competent bacteria were transformed with the obtained lentiCRISPRv2-sgRNA recombinant plasmids and individual bacterial colonies were selected in Luria-Bertani Agar (Miller's LB Agar; Conda) Petri dishes containing 100 µg/mL of ampicillin (NZYTech), the antibiotic resistance marker present on the lentiCRISPRv2 vectors. Briefly, 100 µL of TOP10 chemically competent *Escherichia coli* strain was mixed with the plasmid DNA, using a 1:10 ratio. After leaving the mixture on ice for 15 min, the heat-shock was performed at 42ºC for 1 min, and then the mixture was quickly placed on ice for 5 min. Thereafter, approximately 4 volumes of LB (Miller's LB Broth; Conda) without antibiotics was added to the mixture. After an incubation period of 60 min at 37ºC with shaking (200 rpm), 100 µL of the transformation mixture was placed on LB Agar Petri dishes with 100 µg/mL of ampicillin and incubated overnight at 37ºC.

We only proceed to Colony PCR when there was no growth of colonies in Vector without Ligase control plate and when Vector with Ligase control plate presented much less colonies than Vector+Insert plate.

8| Colony PCR

Positive colonies were screened by colony PCR of individual bacterial colonies. For each colony/sample, the reaction mix with a final volume of 10 µL contained: nuclease-free water, 2 µL of 5x Green GoTaq® Flexi buffer, 0.75 µL of $MgCl_2$ (25 mM), 0.5 µL of dNTPs mixture (dATP, dCTP, dGTP and dTTP, each at 2.5 mM), 0.5 µL of each primer (10 µM), lentiCRISPR_F (5'-TGGAAAGGACGAAACACCG-3') and

lentiCRISPR_R (5'- TGGCACCGAGTCGGTGCTT-3'), 0.15 µL of GoTaq® Flexi DNA Polymerase. For a correct track of colonies, a 96-well cell culture plate, with 100 µL of LB plus 100 µg/mL of ampicillin, was inoculated with the same colonies used in the PCR. For this, the isolated colonies were scraped with pipette tips and dipped in the PCR mixtures and then in the respective well of the 96-well plate. After a cell lysis step at 95ºC for 5 min, samples were amplified for 35 cycles at 95ºC for 60 s, 56ºC for 30 s and 72ºC for 30 s, followed by a final extension of 7 min at 72ºC. During the PCR reaction, the 96-well plate containing LB was incubated at 37ºC. The PCR total reaction volume was loaded onto a 1.2% agarose gel, at 120V during 30 min. Once the electrophoretic result was analyzed, 5 µL of four positive colonies (colonies with PCR amplification) of the 96-well plate were used to re-inoculate in 5 mL of LB, with 100 µg/mL of ampicillin, to perform plasmid DNA extraction. The LentiCRISPRv2-sgRNA recombinant plasmids were extracted using the ZR Plasmid Miniprep™ kit (ZYMO RESEARCH), according to the manufacturer's protocol.

9| Sanger Sequencing

Sanger sequencing was performed to verify the cloned sgRNA sequences. Briefly, sequencing reactions were performed using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) with 5x Sequencing Buffer (Applied Biosystems), 10 µM of primer, the purified PCR product or plasmid DNA and nuclease-free water. The sequencing products were purified with illustra ™ Sephadex ™ G-50 Fine DNA Grade and the automated capillary electrophoresis was accomplished on 3500 Genetic Analyzer sequencer (Applied Biosystems).

10| Cell Culture

HeLa and Human Embryonic Kidney (HEK) 293T cells were grown in Dulbecco's Modified Eagle Medium (DMEM) with GlutaMAX (Gibco, Thermo Fisher Scientific) supplemented with 10% fetal bovine serum (FBS; Gibco, Thermo Fisher Scientific) and 1% Penicillin-Streptomycin solution (Gibco, Thermo Fisher Scientific). HAP1 cells grew in Iscove's Modified Dulbecco's Medium (IMDM; Gibco, Thermo Fisher Scientific) with 10% FBS and 1% Penicillin-Streptomycin solution. HeLa, HEK 293T and HAP1 adherent cells in each passage were washed with 1x Phosphate Buffered Saline (PBS; Gibco, Thermo Fisher Scientific) and then trypsinized with TrypLE™ Express reagent (Gibco, Thermo Fisher Scientific). Thereafter, cells were resuspended in fresh complete media and transferred to a new cell culture flask in a 1:10 dilution.

Jurkat E6.1 suspension cell line was cultured and maintained in Roswell Park Memorial Institute (RPMI) 1640 medium with GlutaMAX (Gibco, Thermo Fisher

Scientific), supplemented with 10% FBS and 1% Penicillin-Streptomycin solution. At each passage, cells were centrifuged for 5 min at 300 g. Then, the cell pellet was resuspended in fresh complete RPMI with a 1:5 dilution and placed into a new cell culture flask.

All cell lines were maintained in a humidified incubator at 37ºC with 5% CO_2, and were splitted every 3-4 days. An inverted light microscope (Zeiss Axiovert 25) was used to observe cell morphology and for cell counting in a Neubauer improved bright-line chamber (Marienfeld Superior), with cells diluted 1:1 in Trypan blue (Gibco, Thermo Fisher Scientific).

11| Cell Line Transfections

For the lentiviral production, HEK 293T packaging cell line was co-transfected with three different plasmids: (1) lentiCRISPRv2-sgRNA recombinant plasmid, the all-in-one lentiviral vector with the respective sgRNA and the humanized *Streptococcus pyogenes* Cas9 (hSpCas9); (2) p8.91 (Addgene), expressing HIV Gag (Group antigens), Pol (Reverse transcriptase), Tat (Transactivator of transcription) and Rev (Regulator of expression of viral) proteins; (3) pMD.G (Addgene), VSV-G envelope expression plasmid. The lentiCRISPRv2 all-in-one lentiviral vector encodes the hSpCas9 and the puromycin resistance expression cassete (Puro). In the case of the plentiCRISPRv2-Blast, the resistance expression cassete was replaced by blasticidin (Blast). The procedure was done using the jetPRIME® transfection reagent (Polyplus transfection®), according to the manufacture's guidelines. Approximately 600 000 HEK 293T cells per well were seeded in a 6-well cell culture plate, in a final volume of 2 mL, to reach 60% to 80% confluence 24 h later. In the day of transfection, complete media was replaced by 2 mL of DMEM without 1 % Penicillin-Streptomycin solution, to avoid interference of the antibiotics in the transfection efficiency. To ensure an equal cell uptake, the amount of each of the three different plasmids used for the transfection was proportional to its size. This means that, in a final DNA amount of 2 µg we used 839 ng of each plentiCRISPRv2-Puro-sgRNA (~13 kb) or 831 ng of each plentiCRISPRv2-Blast-sgRNA (~12.8 kb), 774 ng of p8.91 (~12 kb) and 387 ng of pMD.G (~6 kb). We used one well for each lentiCRISPRv2-sgRNA recombinant plasmid. In addition to a non-transfected negative control, we also had a positive control of transfection using an enhanced green fluorescent protein plasmid (pEGFP). In this case we used 588 ng of pEGFP construct (~7.5 kb), 941 ng of p8.91 and 471 ng of pMD.G. In addition to the DNA, the transfection mixture also contained 200 µL of jetPRIME® buffer and 4 µL of jetPRIME® reagent. Four hours post-transfection, the medium of each well was replaced by 2 mL of complete DMEM, whenever we planned to infect Hela cells, complete IMDM, whenever we

envisioned to infect HAP1 cells, or complete RPMI, whenever we intended to infect Jurkat E6.1 cells. HEK 293T cells were allowed to release lentivirus particles in the supernatant for 48 h, which was then used to transduce HeLa, HAP1 and Jurkat cells.

To perform overexpression of the pEGFP-*MCL1* fused constructs, HeLa cells were transfected using the jetPRIME® transfection reagent. The day before transfection, approximately 80 000 cells per well were seeded in a 24-well cell culture plate with 13 mm coverslips (VWR®), in a final volume of 500 µL. Upon 24 h, the medium was replaced by DMEM without antibiotics, for the same reasons already stated above, and the procedure was done as described by the manufacturer. Cells were transfected with different pEGFP-*MCL1* fused constructs: pEGFP-*MCL1*CDS – with *MCL1* coding sequence (CDS); pEGFP-*MCL1*CDS pA1 – harboring *MCL1* CDS followed by the pA1 3'UTR sequence; pEGFP-*MCL1*CDS pA2 – with the proximal PAS mutated and the pA2 3'UTR sequence following *MCL1* CDS; pEGFP-*MCL1*CDS 1920 and pEGFP-*MCL1* CDS2371 – containing the *MCL1* CDS followed by 3'UTR sequences with different lengths (1920 and 2371 nt) varying between the short and long *MCL1* isoforms. Five hundred ng of each pEGFP-*MCL1* fused construct were used for the transfection. For the co-transfection, 500 ng of pEGFP-*MCL1*CDS pA1 and 25 nM of a chemically modified siRNA, locked nucleic acid (LNA), specific for *MCL1* (5'-TTCCTGATGCCACCTT-3'; Exiqon) were used following the guidelines of the DNA & siRNA jetPRIME® co-transfection protocol. A scramble-LNA (5'-GTGTAACACGTCTATACGCCCA-3'; Exiqon) was used in the LNA co-transfection as a negative control. In both transfection and co-transfection, 50 µL of jetPRIME® buffer and 1.5 µL of jetPRIME® reagent were added to the mixtures and cells were incubated for 48 h. Twenty-four hours post-transfection, the medium was replaced by complete DMEM and, 24 h later, the staining and fixation for microscopy was initiated.

12| Lentiviral Infection

For CRISPR/Cas9 genome editing, lentiviral infection of Jurkat, HeLa and HAP1 cells was implemented. Briefly, cells were infected with lentiviral particles produced by HEK 293T packaging cell line. Forty-eight hours post-transfection of HEK 293T, the medium with the lentiviral particles produced in each well was collected in a falcon tube and centrifuged for 5 min at 1770 g to remove any cells present. After the centrifugation, the supernatant was filtered into a new falcon tube, using a 0.45 µm filter, to ensure the absence of cells in the supernatant. This filtration detail is especially important whenever we infected cells that grow in the same medium as the 293T cells. Regardless of the cell type, the lentiviral infection was always performed using a final volume of 2 mL of lentiviral particles-containing medium.

For transduction of suspension cells, such as Jurkat, 1 000 000 cells were resuspended in 2 mL of medium with the lentiviral particles and placed into a T25 cell culture flask. To delete the two *MCL1* PASs (ΔPAS1 and ΔPAS2) with the two-sgRNA genome editing strategy, cells were infected with 1 mL of each lentiviral particles-containing medium with the desired lentiCRISPRv2-sgRNA recombinant plasmid. Three mL of complete RPMI was then added and cells were incubated with the lentivirus, plus 8 µg/mL of polybrene (Sigma), during 24 h. After those 24 h, the medium was replaced by complete RPMI. Seventy-two hours post-infection, the selection of infected cells was started by adding 5 µg/mL of puromycin (Sigma).

For adherent cells infection, such as HeLa and HAP1, 100 000 to 200 000 cells per well were seeded in a 6-well cell culture plate on the day prior to transduction. Then, the medium was removed and replaced with 1 mL of each lentiviral particles-containing medium with the desired lentiCRISPRv2-sgRNA recombinant plasmid. Cells were incubated with the lentivirus, plus 8 µg/mL of polybrene (Sigma), during 24 h. Twenty-four hours later, the medium of HeLa and HAP1 cells was replaced by complete DMEM or IMDM, respectively. Selection started 72h post-transduction in both HeLa and HAP1 cells. HeLa infected cells were selected through the addition of 3 µg/mL of puromycin. HAP1 infected cells were double-selected by the addition of 2 µg/mL of puromycin plus 5 µg/mL of blasticidin (Sigma) because one sgRNA was cloned into plentiCRISPRv2-Puro and the other into plentiCRISPRv2-Blast.

13| Single-Cell Sorting

Sorting of HeLa and HAP1 single cells by Fluorescence Activated Cell Sorting (FACS) was performed on 96-well cell culture plates with 150 µL of complete media (DMEM or IMDM), with a total of 4 to 5 plates for each cellular condition. Bulk cell populations that underwent sorting were previously splitted in medium without addition of selection markers to recover from selective conditions. Around one million cells were centrifuged for 5 min at 300 g and washed with PBS. Cells were resuspended in FACS buffer (2 % FBS and 0.1 % sodium azide in 1x PBS), filtered with a 100 µm filter and placed into FACS tubes. Filtered cells were subjected to single-cell sorting on a FACS Aria II Cell Sorter (BD Biosciences).

14| Genomic DNA Extraction of Cell Lines

To genotype HeLa, HAP1 and Jurkat E6.1 edited cells, two protocols of genomic DNA extraction were performed. The typical protocol, with phenol:chloroform:isoamyl alcohol was always used for Jurkat cells and bulk populations of HeLa and HAP1 cells.

A 96-well plate protocol was implemented for HeLa and HAP1 clones obtained upon the single-cell sorting.

In the 96-well plate DNA extraction, cells should be previously seeded on a 96-well cell culture plate and allowed to be over-confluent. After removing the culture medium, cells were washed with 150 µL of 1x PBS, and 50 µL of Bradley Lysis Buffer (10 mM Tris-HCl pH 7.5, 10 mM EDTA, 0.5 % Sodium Dodecyl Sulphate (SDS) and 10 mM NaCl) supplemented with 1 mg/mL of Proteinase K solution (20 mg/mL, Ambion) was added to the cell pellet. The plate was sealed with parafilm and incubated in a humidified chamber, overnight at 60ºC. After the incubation period, the plate was allowed to cool to room temperature and 100 µL of an ice-cold solution composed by EtOH/NaCl (75 mM of NaCl in 100 % ethanol) was added to precipitate DNA. The plate was incubated at room temperature during 30 min followed by a centrifugation of 20 min at 3000 rpm. The supernatant was discarded by inverting the plate, 150 µL of cold 70% ethanol was added to rinse the pellet and the plate was centrifuged at 3000 rpm during 10 min. After the centrifugation, the supernatant was removed and the 70% ethanol washing was repeated. The DNA pellet was air dried, resuspended in 30 µL of warm nuclease-free water and stored at -20 ºC.

For the DNA extraction with phenol:chloroform, cells were collected on an Eppendorf and centrifuged 5 min at 300 g. After washing the pellet with 1 mL of 1x PBS, cells were centrifuged again. The cell pellet was resuspended in 500 µL of 1x Sodium chloride-Tris-EDTA (STE) buffer (100 mM NaCl, 10 mM Tris-Cl pH 8.0, 1 mM EDTA) and 25 µL of 20 % SDS plus 10 µL of Proteinase K solution (20 mg/mL). Eppendorfs were sealed with parafilm and incubated overnight at 56ºC with agitation (200 rpm). Thereafter, 20 µL of 5 M NaCl and 550 µL of Phenol: Chloroform: Isoamyl Alcohol (25:24:1 saturated with 10 mM Tris, pH 8.0, 1 mM EDTA; Sigma®) solutions were added. After vortexing and centrifuging samples for 5 min at 14 000 g, the aqueous phase was carefully transferred to a new Eppendorf and 550 µL of chloroform (Millipore) was added. Then, samples were vortexed and centrifuged to recover the aqueous phase, which was again transferred into a fresh Eppendorf. Approximately 2 volumes of absolute ethanol, 1/10 volume of Sodium Acetate 3 M pH 5.2 and 1 µL of glycogen (20 mg/mL; Roche) were added. DNA was allowed to precipitate for, at least, 2 h at -80 ºC. After a centrifugation of 20 min at 4 ºC at maximum speed, the DNA pellet was washed twice with 1 mL of 70 % ethanol, for 10 min at 4 ºC and at maximum speed. After air drying, DNA pellet was resuspended in 50 µL of nuclease-free water and kept at -20 ºC.

15| Nucleic Acids Quantitation

RNA and DNA were quantified by measuring the absorbance at 260 nm in a Nanodrop ND-1000 Spectrophotometer, using 1 µL of sample.

16| Genotyping of CRISPR-edited cells

Screening of ΔPAS1 and ΔPAS2 was performed by PCR in Jurkat, HeLa and HAP1 edited bulk populations, and in clonal populations of HeLa and HAP1 cells obtained by single-cell sorting. The reactions for each sample were prepared with 1 µL of DNA (20-1000 ng) and 19 µL of a mixture with 4 µL of 5x Green GoTaq® Flexi buffer (Promega), 1.5 µL of $MgCl_2$ (25 mM; Promega), 1 µL of dNTPs mixture (10 mM each), 1 µL of each - F and R - primer (10 µM), 0.2 µL of GoTaq® Flexi DNA Polymerase (Promega) and nuclease-free water until the final volume of 20 µL. After an initial denaturation at 95 ºC for 5 min, samples were amplified for 35 cycles at 95 ºC for 45 s, 60 ºC for 60 s and 72 ºC for 90 s, followed by a final extension of 5 min at 72 ºC. Genotyping of ΔPAS1 was performed with the following primers: PAS1-F (5'-GATGGCTTGGAAAAGCAGGC-3') and PAS1-R (5'-ATCTGTAGAGGGAGCAGAAC-3'). Genotyping of ΔPAS2 was made with the following primers: PAS2-F (5'-CTGTCTGCCAAATCCAGTGGA-3') and PAS2-R (5'-GAGGTTTTTGATTTTACTTGGAGGT-3').

17| Total RNA Extraction

For total RNA extraction, cells were harvested on a Nuclease-free Eppendorf by centrifugation at 300 g during 5 min. Hereafter, the supernatant was discarded and 500 µL of TRIzol reagent (Ambion® by Life Technologies™) was added. After an incubation of 5 min at room temperature, 100 µL of chloroform was added and the lysed samples were strongly shaken for 30 s. Thereafter, a 3 min incubation at room temperature was followed by a 15 min centrifugation at 12 000 g at 4ºC. The aqueous phase was cautiously transferred to a new Eppendorf and, from now on, everything was performed on ice. Afterwards, 1 µL of glycogen (20 mg/mL) and the same volume of isopropanol as the aqueous phase were added. The mixture was vortexed and the RNA precipitated at -80 ºC for at least 1 h. After precipitation, samples were thawed at room temperature and centrifuged for 20 min, at 12 000 g at 4 ºC, and the RNA pellet was washed twice with 500 µL of 75 % cold ethanol, centrifuging samples for 10 min at 12 000 g at 4 ºC. Samples were allowed to air dry and the RNA pellet was resuspended in a variable volume of nuclease-free water dependent of the pellet size. The RNA samples were stored at -80 ºC.

18| DNase I Treatment and cDNA Synthesis

DNase I treatment was implemented in 1 µg of RNA to remove the genomic DNA. The procedure was performed in a final volume of 12 µL containing the total RNA, 1 µL of DNase I (Roche), 1.2 µL of DNase I recombinant RNase-free incubation buffer 10x (Roche) and nuclease-free water. After an incubation of 30 min at 37ºC, the enzyme was heat-inactivated by 10 min at 75ºC.

cDNA was synthesized using random hexamers (Sigma) and SuperScript® IV Reverse Transcriptase (SSIV RT; Invitrogen by Thermo Fisher Scientific) enzyme, according to the manufacturer's instructions and minor modifications. Briefly, 1 µL of dNTPs mixture (10 mM each; Thermo Scientific) plus 1 µL of random hexamers (50 µM) were added to the previously prepared mixture for DNase treatment and incubated 5 min at 65 ºC, followed by 5 min at 4 ºC. Then the following components were added for each reaction tube: 4 µL of 5x SSIV Buffer (Invitrogen by Thermo Fisher Scientific), 1 µL of 100 mM Dithiothreitol (DTT; Invitrogen), 0.5 µL of RiboLock RNase Inhibitor (40 U/µL, Thermo Scientific) and 0.5 µL of SSIV RT (200 U/µL). To detect genomic DNA contaminations, a control mixture without SSIV RT was prepared for each sample. The combined reaction mixture was incubated at 23ºC for 10 min, followed by 10 min at 55ºC and finally 10 min at 80ºC to inactivate the enzyme.

19| Quantitative Reverse Transcription PCR (RT-qPCR)

The RT-qPCR reactions were performed using SYBR Select Master Mix (Applied Biosystems), following the manufacturer's guidelines, on a 7500 Fast Real-Time PCR System. All reactions were prepared with 1 µL of cDNA and 9 µL of a mixture containing 5 µL of SYBR, 0.125 µL of each of the forward and reverse primers (0.125 µM) and 3.75 µL of nuclease-free water. Triplicates of each sample were used and the results were assessed using the $\Delta\Delta C_t$ method[88,89] using 18S and GAPDH as normalization controls. The following primers were used to measure *MCL1* pA1 mRNA isoform expression levels: pA1-F (5'-TGCAGACTGGTTGTAGTGGAAC-3') and pA1-R (5'-CATCAACCTCTCAATCCCAGG T-3'). The expression levels of *MCL1* pA2 mRNA isoform were analyzed with the following primers: pA2-F (5'-GCTTGGGGCAGTGAGGGCT-3') and pA2-R (5'-ACCACCTGCCTCCTCCTCC-3').

20| Western Blotting

Whole-cell lysates were obtained through cell lysis on ice using NP40 lysis buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1% NP-40, aprotinin (1 µg/mL), leupeptin (1 µg/mL), pepstatin (1 µg/mL)), supplemented with 1x protease inhibitors (Sigma). To produce the standard curve, eight samples of Bovine Serum Albumin (BSA) with known

concentrations were used. Protein concentrations were quantified by measuring the absorbance at 595 nm, in a plate reader (Synergy 2), using 5 µL of each lysed sample and 250 µL of Bradford reagent (Bio-Rad). A 10 % SDS-PAGE was used to separate equal amounts of the protein lysates, which were then transferred into a nitrocellulose membrane, using the program 3 (dry transfer for 7 min) of iBlot System (Invitrogen). To confirm the transfer to the membrane, Ponceau S staining (NZYtech) was used. Blocking was performed by incubating the membranes, for 60 min at room temperature, with Tris Buffered Saline-0.1% Tween 20 (TBST 0.1 %) containing 5% nonfat dried milk. Mcl-1 was detected through incubation of the membrane with anti-human Mcl-1 mouse monoclonal antibody (Ab22; eBioscience), overnight at 4ºC, at a dilution of 1:500 in 5 % nonfat dried milk with TBST 0.1 % and with goat anti-mouse IgG-HRP secondary antibody (sc-2031, Santa Cruz Biotechnology), for 60 min at room temperature, diluted 1:20 000. As loading control, an anti-α tubulin mouse monoclonal antibody (B-5-1-2; Sigma) was used with a 1:500 000 dilution, followed by a goat anti-mouse IgG-HRP secondary antibody with a 1:20 000 dilution. Detection was achieved using enhanced luminescence (ECL Prime, GE Healthcare) and the immunoblots were visualized on the ChemiDoc™ XRS+ System (Bio-Rad). To quantify the protein band intensities, densitometry analysis on Image lab Software (Bio-Rad) was performed.

21| Flow Cytometry Analysis

Annexin V-Propidium Iodide staining of Jurkat E6.1 and HeLa cells for flow cytometry analysis was performed with the Annexin V – FITC kit EXB0024 (EXBIO) following the recommended procedure, except that the cell density was adjusted to 500 000 cells before adding 1x Annexin V Binding Buffer. Because HeLa are adherent cells and we do not want to lose the suspended population of dead cells, the respective media was centrifuged together with the trypsinized cells. Filtered cells were analyzed on a FACSCANTO II (BD Biosciences) flow cytometer and the results were obtained using the FlowJo v10 software.

22| Staining and Fixation for Microscopy

Since MitoTracker™ Deep Red (Thermo Fisher Scientific) only stains active mitochondria, HeLa living cells were used for the mitochondria staining. Briefly, the medium was removed from the 24-well plate with the cells in coverslips and cells were incubated in 500 µL of complete DMEM with 300 nM of MitoTracker solution. After 30 min of incubation, in a humidified incubator at 37ºC with 5 % CO_2, the medium was removed and the cells were washed three times with 500 µL of 1x PBS, under slow agitation in the dark. From now on, each wash and incubation were performed in the

dark and with slow agitation. Thereafter, PBS was discarded and cells were fixated with 500 µl of 4 % Paraformaldehyde (PFA) solution, during 20 min at room temperature. After the fixation, the PFA solution was removed and coverslips were washed twice with 1x PBS. For nuclei staining, cells were incubated, for 15 min at room temperature, in 300 µL of 1x PBS with 1 µg/mL of Hoechst 33342 dye (Thermo Fisher Scientific). Afterwards, cells in coverslips were washed three times with 1x PBS, mounted in 5 µL of mounting media (Vectashield) and sealed with nail polish.

23| Confocal Microscopy

Imaging of HeLa transfected cells was accomplished on a Leica TCS SP5 confocal microscope. Lasers with different wavelengths were used according to each fluorochrome: 405 nm (Hoechst); 488 nm (pEGFP-fused constructs); 647 nm (MitoTracker™ Deep Red). The obtained images were analyzed and quantified with the ImajeJ software.

24| 3'Rapid Amplification of cDNA Ends (RACE)

To perform 3'RACE of *MCL1* in HAP1 cell line, the SMARTer™ RACE cDNA Amplification kit (Clontech) was used following the manufacturer's protocol. cDNA was synthesized from 1 µg of total RNA using SMARTScribe™ RT. A mixture prepared with 1 µg of RNA and 1 µL of 3'-RACE CDS Primer A (12 µM; 5'-AAGCAGTGGTATCAACGCAGAGTAC$(T)_{30}$VN-3') together with 2.75 µL of deionized water was incubated 3 min at 72 ºC, followed by 2 min at 42 ºC. Afterward, the following reagents were added for each reaction tube: 2 µL of 5x First-Strand Buffer, 1 µL of DTT (20 mM), 1 µL of dNTP mixture (10 mM each), 0.25 µL of RNase inhibitor (40 U/µL) and 1 µL of RT (100 U/µL). Then, the reaction mixture was incubated for 90 min at 42ºC followed by 10 min at 70ºC. Afterwards, cDNA was diluted in 40 µL of Tricine-EDTA Buffer. 3'RACE PCR was done with Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific). For each sample, the reactions were prepared with 4 µL of cDNA and 16 µL of a mixture containing: 4 µL of 5x Phusion HF buffer , 0.4 µL of dNTPs mixture (10 mM each), 1 µL of the *MCL1* gene specific primer exon2-F (10 µM; 5'-CTCGTAAGGACAAAACGGGACTGGCTAG-3'), 2 µL of 10x Universal Primer A Mix (0.4 µM; 5'-CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT-3'), 0.2 µL of Phusion DNA Polymerase (2 U/µL) and nuclease-free water until the final volume of 20 µL. Touchdown PCR were performed as follows: after an initial denaturation step at 98 ºC for 1 min, samples were amplified for 5 cycles at 98 ºC for 10 s and 72 ºC for 3 min, 5 cycles at 98 ºC for 10 s, 70 ºC for 30 s and 72 ºC for 3 min, 27 cycles at 98 ºC for

10 s, 68 ºC for 30 s and 72 ºC for 3 min, followed by a final extension of 10 min at 72 ºC. Fifteen µL of the PCR products were loaded into a 1 % agarose gel.

25| Nested PCR

Two µL of the 3'RACE PCR products were used to perform Nested PCR of *MCL1* with Phusion High-Fidelity DNA Polymerase, in a final volume of 20 µL. Shortly, 4 µL of 5x Phusion HF buffer, 0.4 µL of dNTPs mixture (10 mM each), 1 µL of *MCL1* exon 3-F primer (10 µM; 5'- AGGTGGCATCAGGAATGTGCTGCTG-3'), 1 µL of Nested Universal Primer A (10 µM; 5'-AAGCAGTGGTATCAACGCAGAGT-3') , 0.2 µL of Phusion DNA Polymerase and nuclease-free water until the final volume were used. After an initial denaturation at 98 ºC for 1 min, samples were amplified for 30 cycles at 98 ºC for 10 s, 68 ºC for 30 s and 72 ºC for 90 s, followed by a final extension of 10 min at 72 ºC. The whole volume of the PCR products were loaded into a 1 % agarose gel.

26| Statistical analysis

In the graphics, data are presented as mean ± SD. Statistical differences were tested with two-tailed unpaired *t* test with Welch's correction or one-way ANOVA with Holm-Sidak's correction. Only p-values < 0.05 were considered statistically significant, where ** means p< 0.01, *** means p<0.001, **** means p< 0.0001 and ns means not significant. All graphics were made using GraphPad Prism version 6.01 for Windows, GraphPad Software, La Jolla California USA, www.graphpad.com.

Results and Discussion

1| Physiological function of *MCL1* APA-derived mRNA isoforms

1.1| Genome editing of *MCL1* PASs influences Mcl-1 protein levels in Jurkat cells

Mcl-1 is an anti-apoptotic member of the Bcl-2 protein family that plays a major role in cell survival since it promotes cell viability[56,63]. By Northern Blot, 3' RACE and sequencing, previous studies from our laboratory have discovered two *MCL1* mRNA isoforms generated by 3'UTR-APA in human T cells. Both isoforms, named pA1 (the shortest) and pA2 (the longest), are generated by usage of the PAS1 or PAS2, respectively, and only differ in the 3'UTR length, encoding the same protein (Figure 8).

In order to understand the functional role of *MCL1* APA-derived mRNA isoforms and since Mcl-1 has an anti-apoptotic function, we went to study the physiological impact of each isoform on cell survival. First, we applied the CRISPR/Cas9 technology[84] to delete either the proximal or distal *MCL1* PASs in Jurkat E6.1 cells (human T cells), allowing to study the function of pA1 and pA2 *MCL1* isoforms, individually. The genome regions that after transcription give rise to the 3'UTRs are known to exhibit a high content of AT-rich sequences, making it more difficult to apply CRISPR/Cas9 technology due to the lower occurrence of PAM sequences in these regions. Nevertheless, at the moment, CRISPR/Cas9 is the most efficient technology to perform genome editing and generate stable edited cells. To delete *MCL1* PASs, we used a two-sgRNA genome editing strategy that consisted of using two sgRNAs flanking each PAS. In fact, previous studies have applied this strategy to successfully delete the 3'UTR. For instance, a study conducted in rabbits used the two-sgRNA genome editing strategy to delete the tyrosinase 3'UTR[90]. For our strategy to succeed, both sgRNAs must cleave the target DNA at the same time, to delete the region flanked by them that includes the PAS. To delete the proximal PAS (ΔPAS1) we used PAS1_sgRNA1 and PAS1_sgRNA2, while PAS2_sgRNA1 and PAS2_sgRNA2 were used to delete the distal PAS (ΔPAS2) (Figure 10). Thus, we generated lentiviral particles containing these sgRNAs in HEK 293T cells, and then infected Jurkat cells with the lentiviral particles.

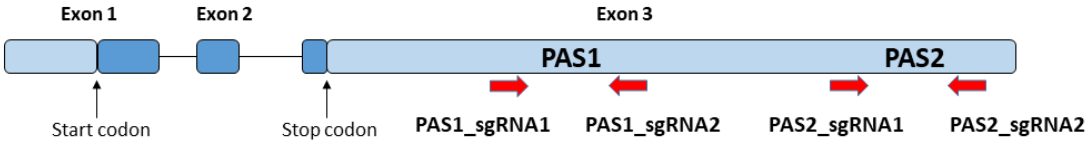


Figure 10: CRISPR/Cas9 strategy | Schematic representation of the strategy used to perform genome editing of the proximal (PAS1) or distal (PAS2) *MCL1* PASs. Light blue: UTRs; dark blue: CDS; red arrows: sgRNAs; lines: introns.

In the ΔPAS1 edited cells we will have the deletion of the PAS responsible for the generation of the shortest isoform, which means that in this condition only the *MCL1* pA2 isoform will be produced. Likewise, ΔPAS2 corresponds to the deletion of the distal PAS responsible for the generation of the longest isoform, so we can study the function of the *MCL1* pA1 isoform.

Genotyping techniques, such as PCR and sequencing, were implemented to evaluate whether the genome editing occurred successfully. Regarding ΔPAS1, the expected size of the band, after gel electrophoresis of the PCR products, corresponding to the wild type (wt) allele was 254 bp, while the allele with the deletion, from now on referred to as deleted allele, would be expected to have a band size of ~184 bp. By gel electrophoresis analysis (Figure 11A), we concluded that deletion of the region flanked by the two sgRNAs, which includes the proximal PAS, was not successful. Sequencing of the PCR product after DNA extraction of the electrophoretic band (Figure 11C) showed that despite of not causing the deletion, both sgRNAs worked by introducing an indel, therefore changing the sequence of the region flanked by them. This means that it would not be possible to perform a second round of infection with those sgRNAs, since they will no longer recognize their target sequence. Concerning ΔPAS2, the electrophoresis (Figure 11B) showed that the deletion of the region that includes the distal PAS occurred with success in the bulk population, since we can clearly distinguish the wild type allele (band size of 790 bp) from the deleted one (with an expected band size of ~500 bp). Sequencing of the DNA corresponding to the deleted allele (Figure 11D) confirmed the deletion of the region that includes the PAS2. Unlike ΔPAS1, on ΔPAS2 edited cells both sgRNAs worked at the same time, meaning that the genome editing occurred successfully. The interpretation of the results obtained from DNA sequencing shows some of the complications of the two-sgRNA genome editing strategy. One of the difficulties to be aware of is that, when the selection strategy is based only in one antibiotic resistance, we cannot control the selection of cells that were not infected with both sgRNAs. This means that the selected cells population will consist of a mixture of cells infected with both sgRNAs and cells infected with only one sgRNA. Furthermore, in

order to delete untranslated regions both sgRNAs must work simultaneously. When the two sgRNAs do not work at the same time, indels could be generated without performing the deletion of the sgRNAs flanked region. Nevertheless, nowadays this remains the most efficient strategy to perform genome editing of PASs.

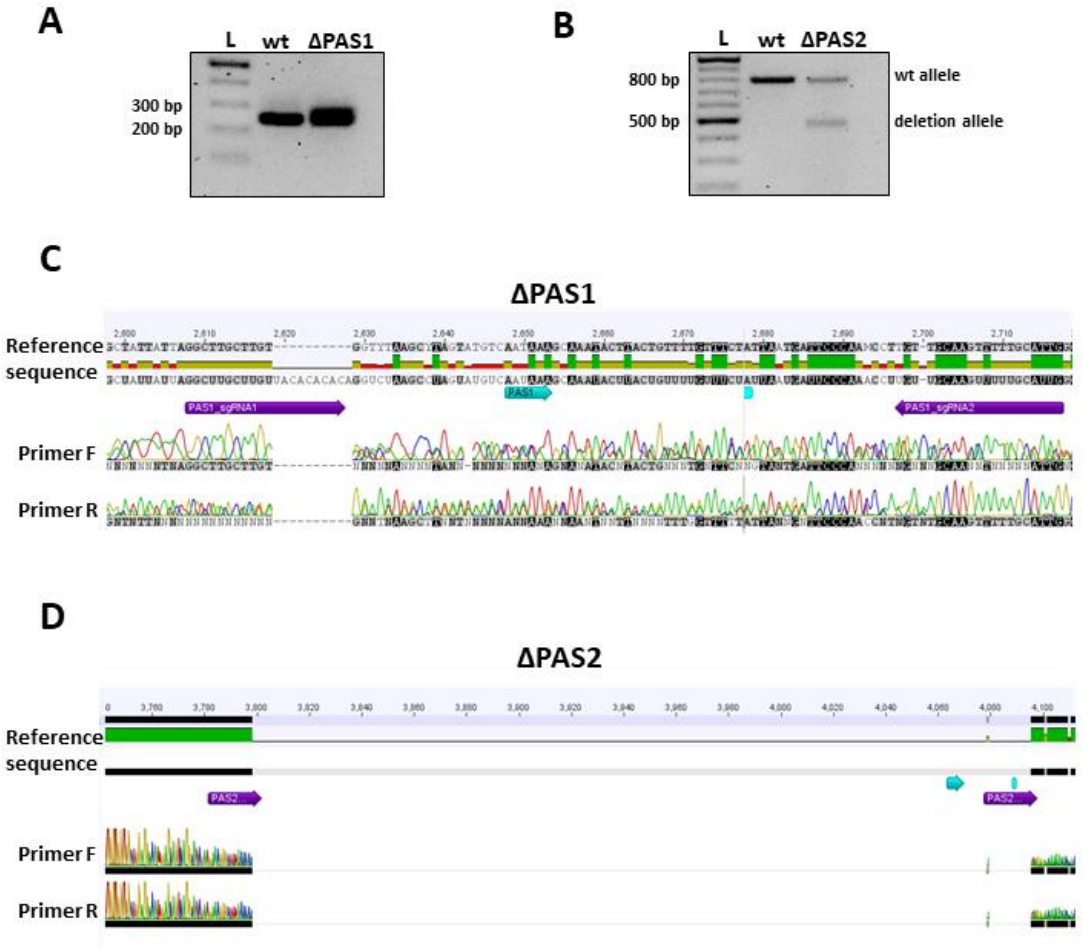


Figure 11: Genotyping of Jurkat E6.1 cells subject to CRISPR/Cas9 genome editing to delete either *MCL1* PAS1 (ΔPAS1) or PAS2 (ΔPAS2) | (A, B) Electrophoretic profile of ΔPAS1 and ΔPAS2 conditions in comparison with non-edited cells (wt). On each panel, the first lane corresponds to the molecular weight; DNA ladder (L); **(C, D)** Genomic alignment of *MCL1* 3'UTR sequence with the sequencing results obtained by sephadex DNA purification of the ΔPAS1 band (~254 bp) (**C**) and the ΔPAS2 band corresponding to the deletion allele (~500 bp) (**D**). In the reference sequences are indicated the PASs (blue arrows) and their polyA sites (blue boxes), as well the sgRNAs (purple arrows) used for the genome editing.

Having obtained the CRISPR edited Jurkat cells, we then ask if ΔPAS1 and ΔPAS2 affected *MCL1* mRNA expression and Mcl-1 protein levels. Although having not deleted the PAS1, we decided to measure *MCL1* pA1 mRNA expression and Mcl-1 protein levels also in this condition, since the sequence of the PAS, AATAAA, was altered in this cell population, at least in one of the alleles. By RT-qPCR, it seems that the genome editing does not affect mRNA levels (Figure 12A), since ΔPAS1 and ΔPAS2

cause only a slight decrease in the expression levels of *MCL1* pA1 and pA2 mRNAs, respectively. This minor reduction on the mRNA levels is most likely because the genome editing did not occur in homozygosity, therefore we are quantifying the mRNA levels obtained from the wt allele. However, Mcl-1 protein levels are strongly decreased in both ΔPAS1 and ΔPAS2 conditions (Figure 12B).

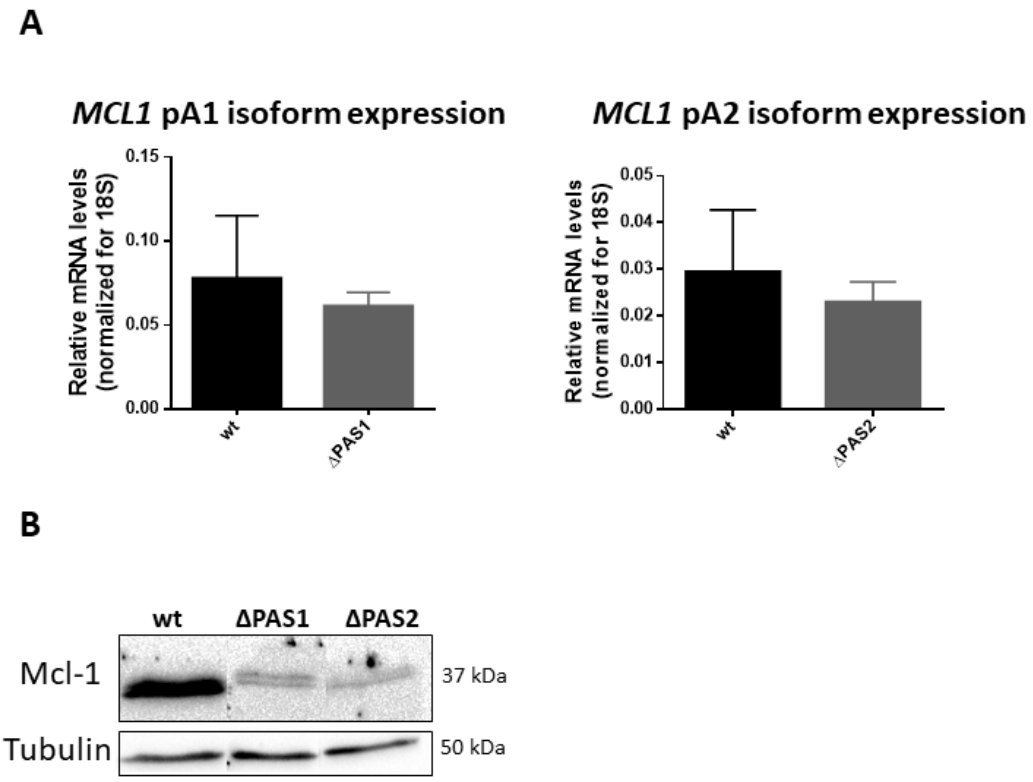


Figure 12: Deletion of PAS1 and PAS2 decreases Mcl-1 protein levels in Jurkat E6.1 cells | (A) Relative *MCL1* pA1 (left) and pA2 (right) mRNA levels of wt, ΔPAS1 and ΔPAS2 cells were accessed by RT-qPCR and normalized for 18S. Data are shown as mean + SD of three independent experiments. Statistical significance was determined by an unpaired *t* test with Welch's correction. **(B)** Western blot of Mcl-1 in ΔPAS1, ΔPAS2 and wt Jurkat E6.1 cells, using a monoclonal anti-human Mcl-1 antibody. A monoclonal anti-tubulin antibody was used to detect α-tubulin as a loading control.

This suggests that a slight decrease in the expression of *MCL1* pA1 and pA2 mRNAs causes a significant impact on Mcl-1 protein levels. Therefore, we can proceed with the functional assays. Taken together, these results enabled to demonstrate that the genome editing was successful as it affects Mcl-1 protein levels. Importantly, our results show that genome editing of untranslated regions, such as deletions of PASs, affects protein expression.

1.2| *MCL1* APA-derived mRNA isoforms contribute to Mcl-1 anti-apoptotic function

To evaluate the physiological effect of both *MCL1* APA-derived mRNA isoforms on Mcl-1 cellular function, we measured the contribution of each isoform for Jurkat E6.1 cell viability by Annexin V-Propidium Iodide assays. The results (Figure 13) clearly show an increase in the amount of apoptotic cells in both CRISPR conditions (ΔPAS1 and ΔPAS2).

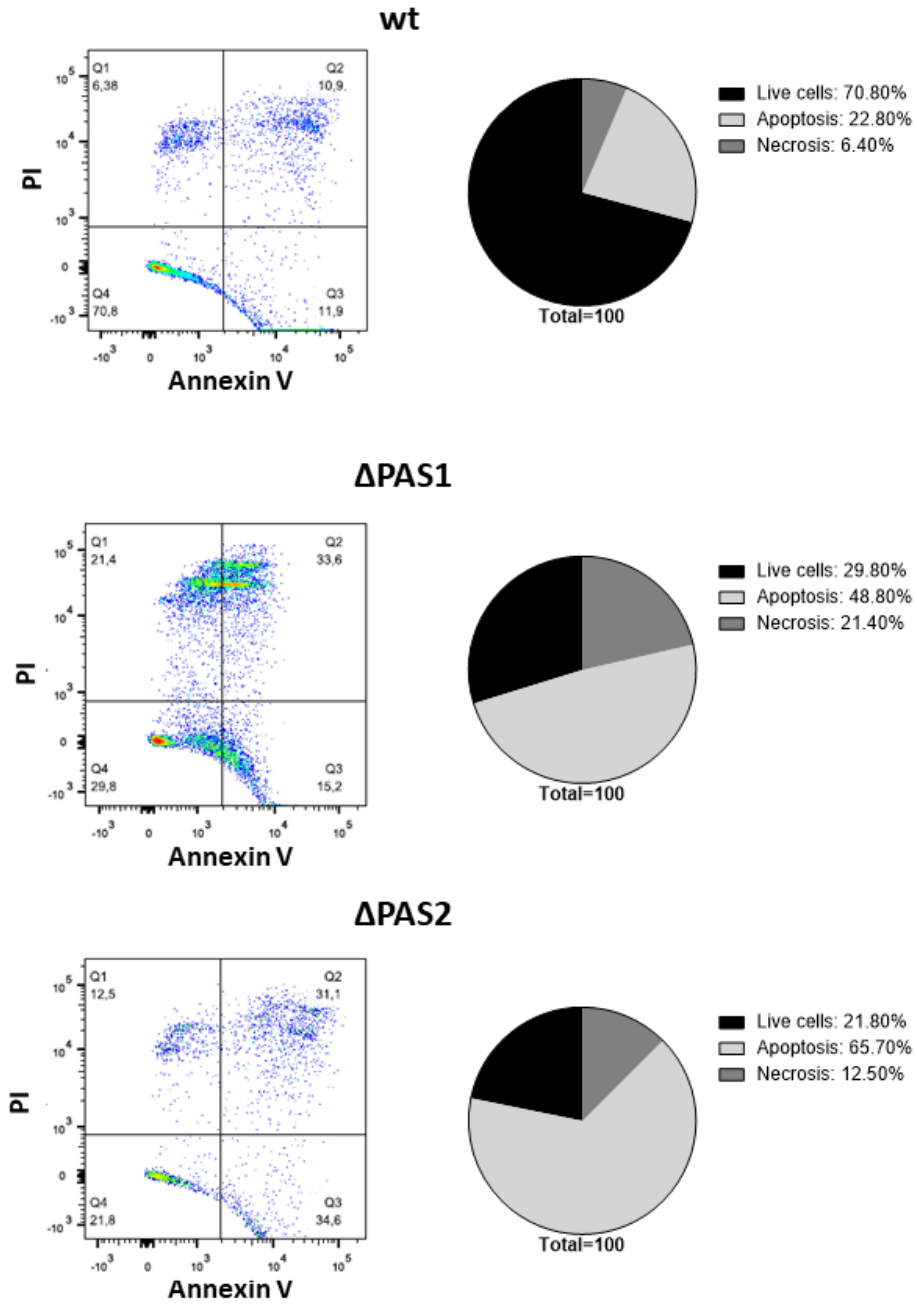


Figure 13: Both pA1 and pA2 mRNA isoforms contribute to Mcl-1 anti-apoptotic function in Jurkat E6.1 cells | Annexin V-PI staining followed by Flow Cytometry analysis performed in wt, ΔPAS1 and ΔPAS2 cells. Q1 (necrosis); Q2 (late apoptosis); Q3 (early apoptosis); Q4 (live cells). Pie charts represent the results obtained for each condition; early and late apoptosis were combined and designated as apoptosis.

The increase of apoptosis in both CRISPR conditions, where Mcl-1 protein levels are reduced, is in agreement with the anti-apoptotic function of Mcl-1. For instance, a recent study showed that apoptosis is induced due to decreased protein levels of Mcl-1[91]. In fact, several studies have shown that Mcl-1 inhibition results in apoptosis induction. Since in both ΔPAS1 and ΔPAS2 CRISPR cells there is the decrease of Mcl-1 protein levels (Figure 12B), our results show that the Mcl-1 cellular function is given by both pA1 and pA2 mRNA isoforms.

1.3| Genome editing of *MCL1* PASs influences Mcl-1 protein levels in HeLa cells

To increase the robustness of our results we needed to obtain homozygous cell populations by single sorting of the CRISPR edited cells. It is well known that Jurkat E6.1 cells are difficult to grown in single cell conditions, and our group previous attempts to perform Jurkat single cell sorting had failed. Therefore, we decided to implement the CRISPR/Cas9 technology in HeLa cells, which are easier to manipulate and suitable to perform single cell sorting. Previously, we used one pair of sgRNAs to delete each PAS in Jurkat E6.1 cells. However, in order to get two CRISPR conditions for each PAS, we decided to use two pairs of sgRNAs, independently, to delete each signal in HeLa cells. This was just a precautionary measure recommended in the literature[84] as occasionally, and for reasons still unknown, some sgRNAs are not efficient. Regarding the distal PAS, besides the pair of sgRNAs (PAS2_sgRNA1 and PAS2_sgRNA2) already used in Jurkat, HeLa cells were also infected with another pair of sgRNAs (PAS2_sgRNA1 and PAS2_sgRNA3) flanking this signal. Concerning the PAS1, it was not possible to design another pair of sgRNAs because: a) the deleted region could not be very large, b) could not interfere with the use of the distal PAS, and c) there was no other PAM sequence near the region flanked by the pair of sgRNAs already used on Jurkat cells. Therefore, in the case of PAS1, we only had one sgRNA pair. Through the genotyping techniques of the ΔPAS1 edited bulk population (Figure 14A and 14D), we concluded that, both sgRNAs worked since we observed alteration in the genomic sequence, but not simultaneously and, therefore, did not generate the deletion, since the deleted allele (with an expected band size of ~184 bp) was not amplified. In the ΔPAS2_1/2 condition, where we used PAS2_sgRNA1 and PAS2_sgRNA2, and in the ΔPAS2_1/3 condition, with PAS2_sgRNA1 and PAS2_sgRNA3, the electrophoresis (Figure 14B and 14C) showed that the deletion of the sgRNAs flanked region occurred successfully. Through sequencing (Figure 14E and 14F) of the DNA corresponding to the deleted allele (with

an expected band size of ~500 bp in ΔPAS2_1/2 and ~410 bp in ΔPAS2_1/3) we confirmed that the PAS2 was deleted, therefore the genome editing was well succeed in the bulk population .

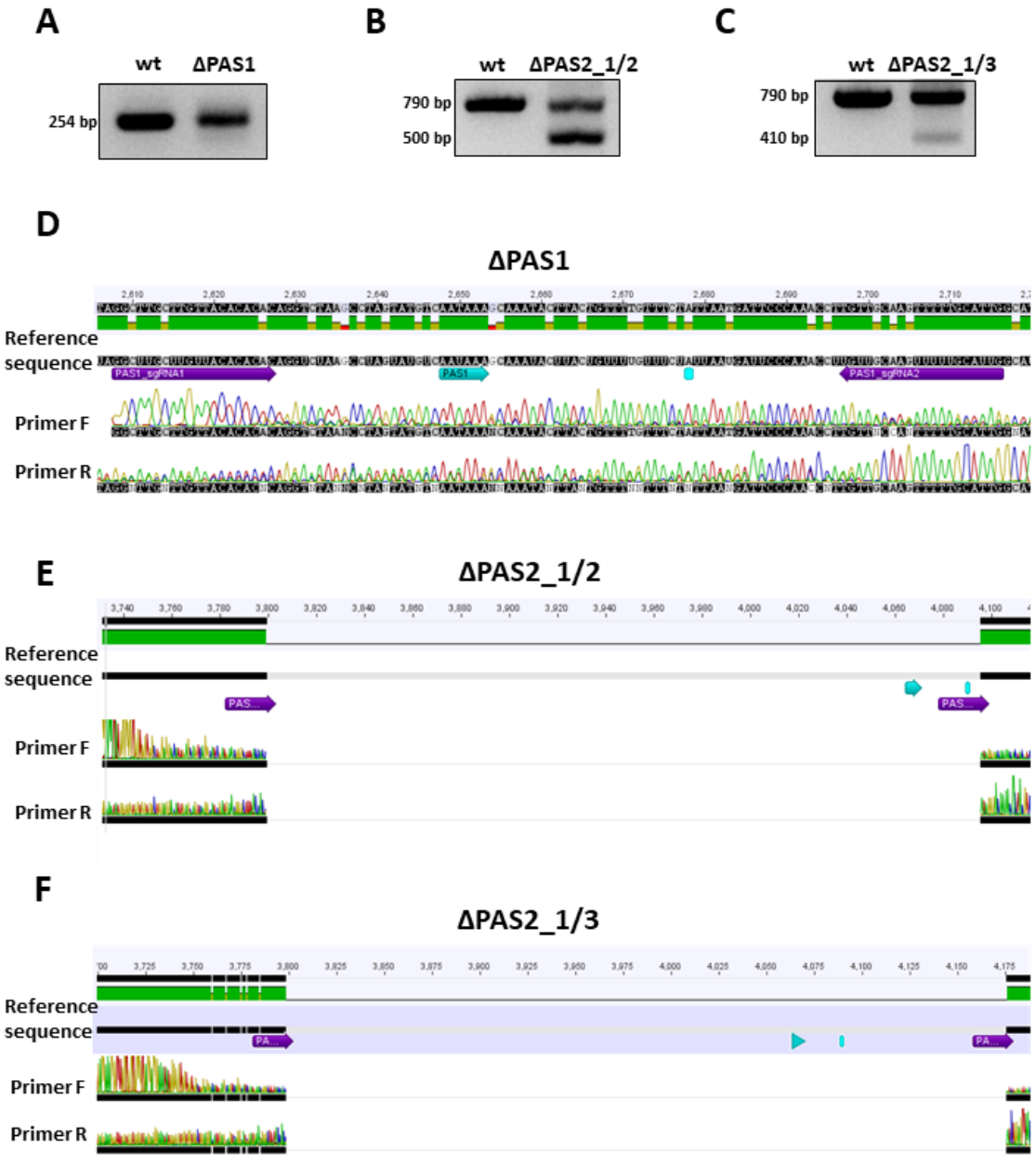


Figure 14: Genotyping of HeLa bulk populations subject to CRISPR/Cas9 technology to delete either *MCL1* PAS1 (ΔPAS1) or PAS2 (ΔPAS2) | (A-C) Electrophoretic profile of the DNA amplification of ΔPAS1 , ΔPAS2_1/2 and ΔPAS2_1/3 conditions in comparison with non-edited cells (wt). On each panel, the correspondent molecular weight is indicated in the left; **(D-F)** Genomic alignment of *MCL1* 3'UTR sequence with the sequencing results obtained by sephadex DNA purification of the ΔPAS1 band (~254 bp) (**D**), the ΔPAS2_1/2 band corresponding to the deletion allele (~500 bp) (**E**) and the ΔPAS2_1/3 band corresponding to the deletion allele (~410 bp) (**F**). In the reference sequences are indicated the PASs (blue arrows) and their polyA sites (blue boxes), as well the sgRNAs (purple arrows) used for the genome editing.

After confirming that the deletion alleles were present in the bulk population, we proceeded to single cell sorting of the HeLa CRISPR edited cells by FACS in three 96-well cell culture plates for each condition (288 clones), to obtain homozygous cell populations with PAS deletion. Although PAS1 was not deleted and its sequence remains unchanged, we decided to proceed with sorting of this condition, since cell bulk populations are heterogeneous and, therefore, ΔPAS1 cells could not have been detected by sequencing if they are in low levels in the total cell population. In each well, a clonal cell line, from now on referred to as clones, generated from a single cell was obtained. Upon an expansion period of ~ 2-3 weeks, there were 27 clones from the ΔPAS1 condition, 22 from ΔPAS2_1/2 and 39 from ΔPAS2_1/3. These numbers of clones were low because the vast majority of the cells died, suggesting that *MCL1* APA-derived mRNA isoforms are essential for cell survival.

The analysis of the electrophoretic profile of the PCR genotyping of the clones obtained initially suggested that we had some homozygous clonal edited cells, since only the deleted allele was amplified. Figure 15 is a representative illustration that serves as an example of this first electrophoresis, showing some clones from ΔPAS2_1/3 condition. The arrows indicate the putative homozygous clonal edited cells.

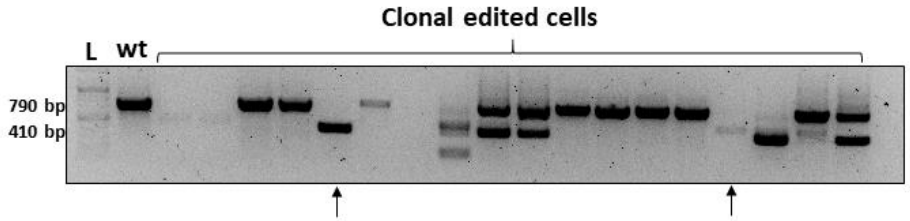


Figure 15: Representative electrophoresis of HeLa clonal edited cells | Electrophoretic profile of the DNA amplification of clonal cells from ΔPAS2_1/3 condition obtained from single cell sorting of CRISPR edited cells, in comparison with non-edited cells (wt). DNA ladder (L). Arrow indicates a putative homozygous cell population.

To verify that we had indeed homozygous clones we repeated the genotyping using primers targeting the deleted region (Figure 16A). However, after genotyping we observed amplification of the PCR products (Figure 16B), which means we did not obtain any viable homozygous clone in the single cell sorting.

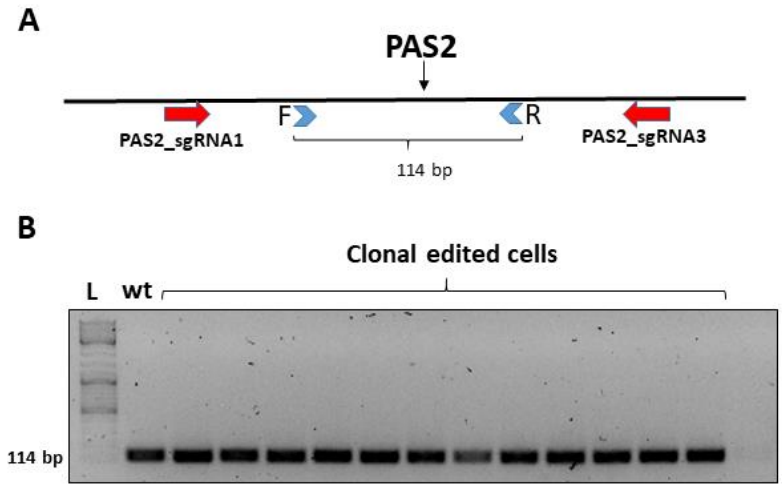


Figure 16: Representative electrophoresis of HeLa clonal edited cells with primers targeting the supposed deleted region | (A) Scheme with the localization of the primer pair used (blue arrowheads). The two sgRNAs used to delete the PAS2 are also depicted as well as the localization of the PAS2. **(B)** Electrophoretic profile of the putative homozygous clonal edited cells, in comparison with non-edited cells (wt); This figure only shows some clones from ΔPAS2_1/3 condition, representing the results acquired with all the obtained clones from each condition; DNA ladder (L).

Despite of not having homozygous clonal cells, we selected six heterozygous clones from each condition (ΔPAS1, ΔPAS2_1/2 and ΔPAS2_1/3) and measured the expression of *MCL1* mRNA isoforms in those cells. The reason for selecting only a few clones from each condition was due to the difficulty in maintaining so many clonal cell populations at the same time. By RT-qPCR (Figure 17), we measured the expression levels of pA1 mRNA on the six selected clones from ΔPAS1 condition, and the clone with less pA1 mRNA expression was chosen. The same rationale was applied for the two ΔPAS2 conditions.

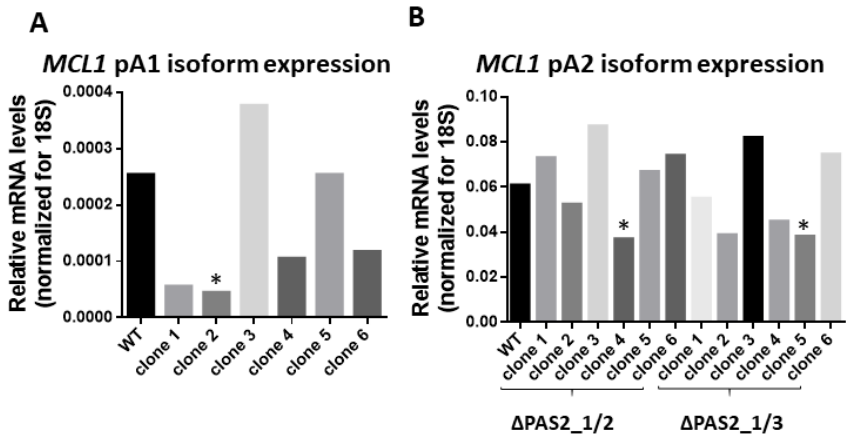


Figure 17: Measurement of *MCL1* mRNA levels from six heterozygous clones from each condition | (A) Relative *MCL1* pA1 mRNA levels of wt and heterozygous clones from ΔPAS1 condition. **(B)** Relative *MCL1* pA2 mRNA levels of wt and heterozygous clones from ΔPAS2_1/2 and ΔPAS2_1/3 conditions. Clones selected are represented by asterisks.

In the ΔPAS1 condition, we selected a clonal cell population that, according to the PCR and sequencing (Figure 18A and 18D), possesses three possible alleles, which is in agreement with the fact of HeLa cells having three chromosomes 1. Thus, this clone has the deleted allele (with an expected band size of ~184 bp) where the genome editing occurred successfully deleting the PAS1. Besides the wt allele (band size of 254 bp), where only one sgRNA worked, we also obtained another allele where both sgRNAs worked but did not generate the deletion.

Concerning PAS2 (Figure 18B, 18C, 18E and 18F), we selected a clonal cell population from each CRISPR condition with PAS2 deleted. Moreover, sequencing of the DNA corresponding to the upper band allowed us to conclude that in ΔPAS2_1/2 condition both sgRNAs worked in that allele but not simultaneously, and in PAS2_1/3 condition only one sgRNA worked. Despite of not having homozygous cell populations, we decided to proceed with the work on these heterozygous clones because we had previously shown by western blot (Figure 12B) that Mcl-1 protein levels were reduced in the bulk population of CRISPR edited Jurkat cells.

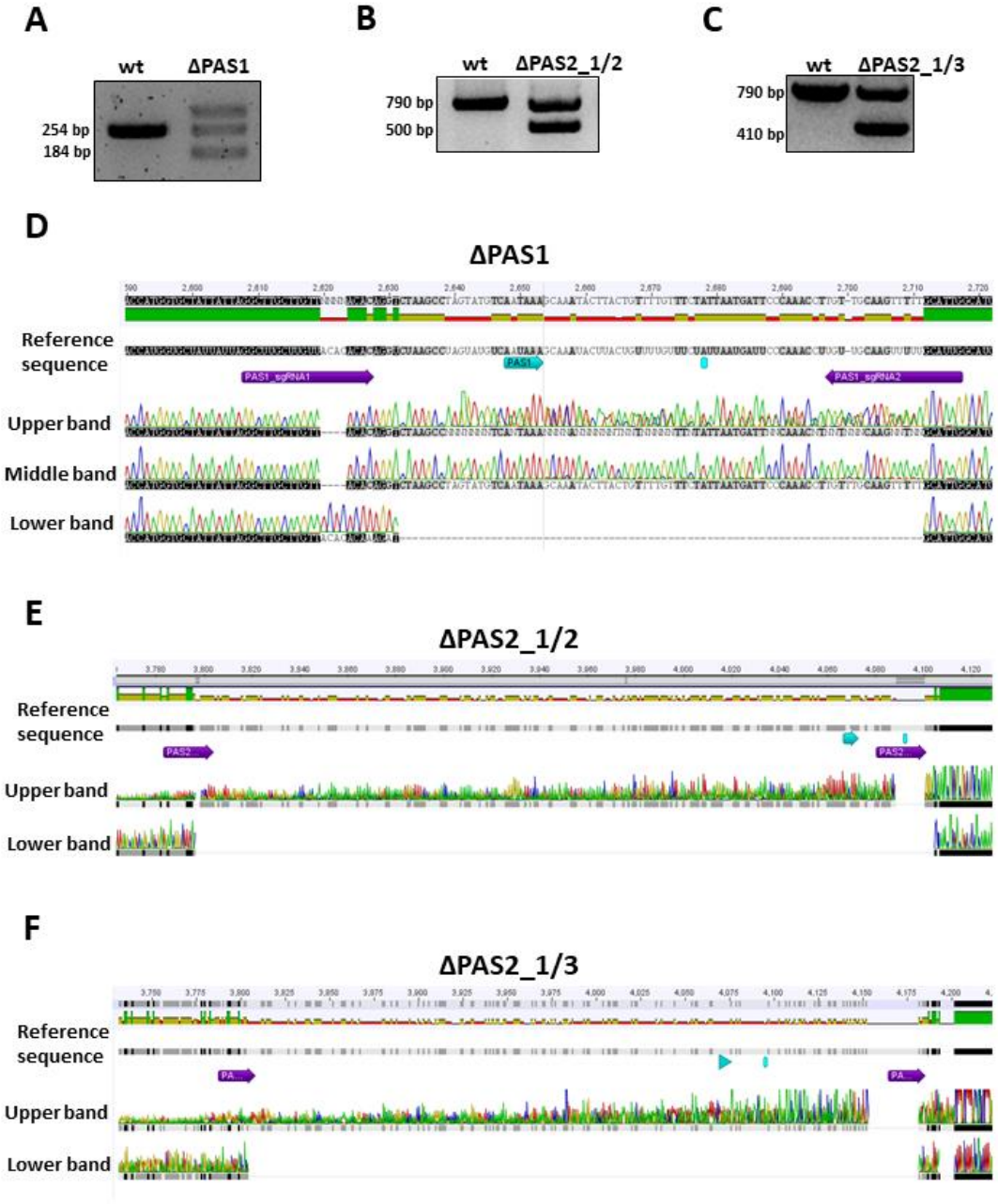


Figure 18: Genotyping of HeLa clonal cell lines obtained by sorting of HeLa CRISRP/Cas9 bulk population | (A-C) Electrophoretic profile of ΔPAS1, ΔPAS2_1/2 and ΔPAS2_1/3 conditions in comparison with non-edited cells (wt). On each panel, the correspondent molecular weight is indicated in the left; **(D-F)** Genomic alignment of *MCL1* 3'UTR sequence with the sequencing results obtained by sephadex DNA purification of all bands obtained for each condition. In the reference sequences are indicated the PASs (blue arrows) and their polyA sites (blue boxes), as well the sgRNAs (purple arrows) used for the genome editing.

After confirming that our genome editing strategy was successful in heterozygosity, its impact on *MCL1* pA1 and pA2 mRNA isoforms and on Mcl-1 protein levels was investigated. The results achieved through RT-qPCR (Figure 19A), even though not significant (possibly due to the heterozygosity of the clones), suggest that, in the ΔPAS1 there is a reduction on *MCL1* pA1 isoform expression, and in both ΔPAS2 conditions, the expression of *MCL1* pA2 isoform decreases. By Western Blot (Figure

19B) we confirmed that the decrease in the mRNA expression levels of pA1 and pA2 isoforms translates into a decrease on Mcl-1 protein levels. Taken together, these results show that ΔPAS1 and ΔPAS2 heterozygous clones present a decrease in *MCL1* mRNA and protein levels and further demonstrate that genome editing of untranslated regions affects protein levels.

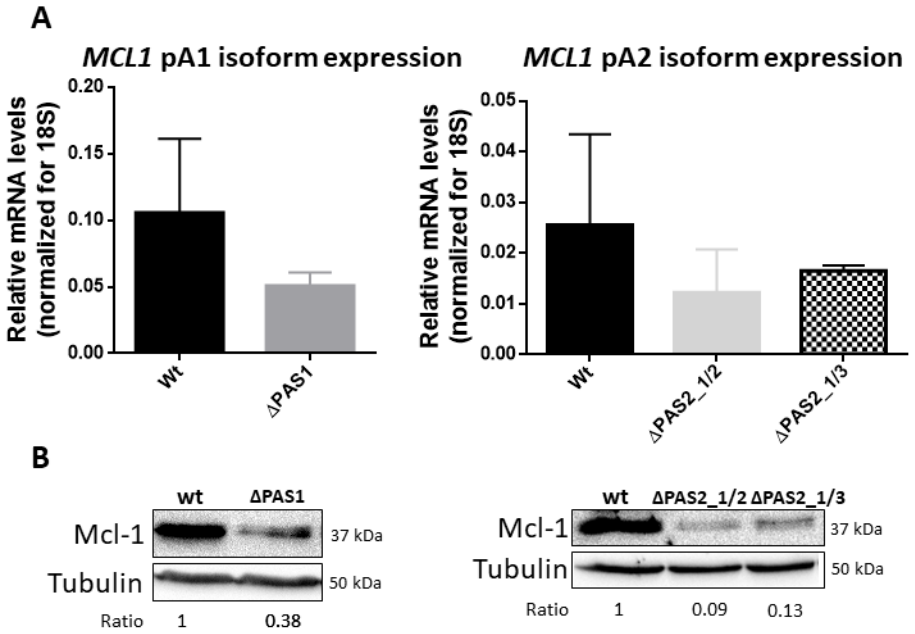


Figure 19: Deletion of PAS1 and PAS2 decreases Mcl-1 protein levels in HeLa clonal cell lines | (A) Relative *MCL1* pA1 (left) and pA2 (right) mRNA levels of wt, ΔPAS1, ΔPAS2_1/2 and ΔPAS2_1/3 cells accessed by RT-qPCR and normalized for 18S. Data are shown as mean + SD of three independent experiments. Statistical significance was determined by an unpaired *t* test with Welch's correction, for *MCL1* pA1 mRNA levels, and by a one-way ANOVA with Tukey's correction, for *MCL1* pA2 mRNA levels. **(B)** Western blot of Mcl-1 in wt, ΔPAS1, ΔPAS2_1/2 and ΔPAS2_1/3 cells, using a monoclonal anti-human Mcl-1 antibody. A monoclonal anti-tubulin antibody was used to detect α-tubulin as a loading control. Mcl-1 protein quantification using densitometry of band signal intensity and normalized to tubulin protein signal intensity is shown below the WB.

1.4| The functional role of *MCL1* APA-derived mRNA isoforms is not cell type-specific

After establishing CRISPR cells, we then investigated the physiological effect of each *MCL1* APA-derived mRNA isoform on Mcl-1 cellular function. Measurement of the apoptosis levels in HeLa clones (Figure 20) showed that in all CRISPR conditions (ΔPAS1, ΔPAS2_1/2 and ΔPAS2_1/3) there is an increase in the amount of apoptotic cells, comparing to the wt. This is in agreement with the results obtained with Jurkat cells and allowed us to reinforce that both pA1 and pA2 play a role on the Mcl-1 anti-apoptotic function. Moreover, these results also demonstrate that the functional role of *MCL1* APA-derived mRNA isoforms is not cell type-specific.

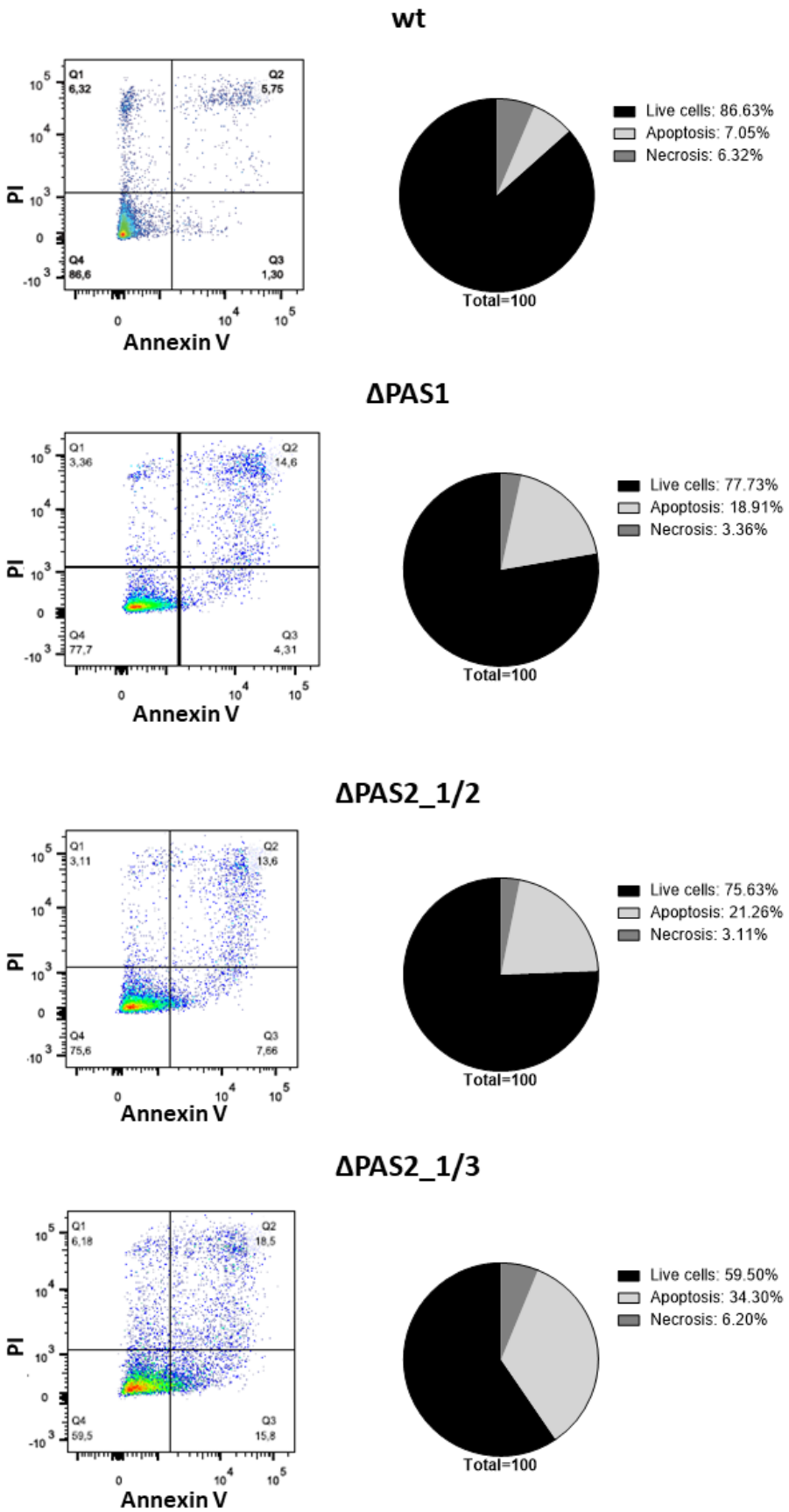


Figure 20: *MCL1* APA-derived mRNA isoforms contribute to Mcl-1 anti-apoptotic function in Hela clonal cell lines | Annexin V-PI staining followed by Flow Cytometry analysis performed in wt, ΔPAS1, ΔPAS2_1/2 and ΔPAS2_1/3 cells. Q1 (necrosis); Q2 (late apoptosis); Q3 (early apoptosis); Q4 (live cells). Pie charts represent the results obtained for each condition. Q3 and Q3 were combined and designated as apoptosis.

1.5| Genome editing of *MCL1* PASs in an haploid cell line

Taking into account that obtaining the homozygous cell population failed both in Jurkat and in Hela cells, we decided to perform CRISPR/Cas9 genome editing of the PASs in a haploid cell line. We used HAP1 cells, an adherent cell line derived from Chronic Myelogenous Leukemia, to delete either the proximal and distal *MCL1* PAS. Firstly, we confirmed by 3' RACE (Figure 21) that *MCL1* generates two mRNA isoforms (pA1 and pA2) by 3'UTR-APA in this haploid cell line.

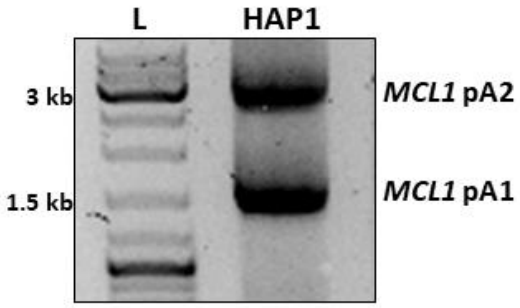


Figure 21: *MCL1* 3'RACE in HAP1 cell line | Two bands corresponding to the *MCL1* APA-derived mRNA isoforms (pA1 and pA2) are represented in the second lane. The first lane corresponds to the molecular weight DNA ladder (L).

To delete the proximal PAS we used PAS1_sgRNA1 and PAS1_sgRNA2, the same pair already used for Jurkat and HeLa cells, while PAS2_sgRNA1 and PAS2_sgRNA3 were used to delete the distal PAS. In order to avoid selecting cells that were not infected with both sgRNAs, we decided to implement a double-selection strategy. For this, HAP1 cells were infected with lentiviral particles containing sgRNAs cloned in all-in-one lentiviral vectors with different antibiotic resistance. Thus, PAS1_sgRNA1 was cloned in plentiCRISPRv2-Puro, PAS1_sgRNA2 in plentiCRISPRv2-Blast, PAS2_sgRNA1 in plentiCRISPRv2-Puro and PAS2_sgRNA3 in plentiCRISPRv2-Blast. However, this double-selection strategy with puromycin and blasticidin selection does not guarantee that all selected cells will have the respective PAS deleted, since, as previously mentioned, for this to happen both sgRNAs have to work simultaneously.

Through the genotyping of HAP1 cells bulk population (Figure 22A and 22C), we concluded that the deletion of the region including the PAS1 was not successful, since the PAS sequence remains unchanged. Regarding the distal PAS (Figure 22B and 22D), the results showed that the genome editing occurred with success in some cells of the bulk population, as we can still detect the wt allele.

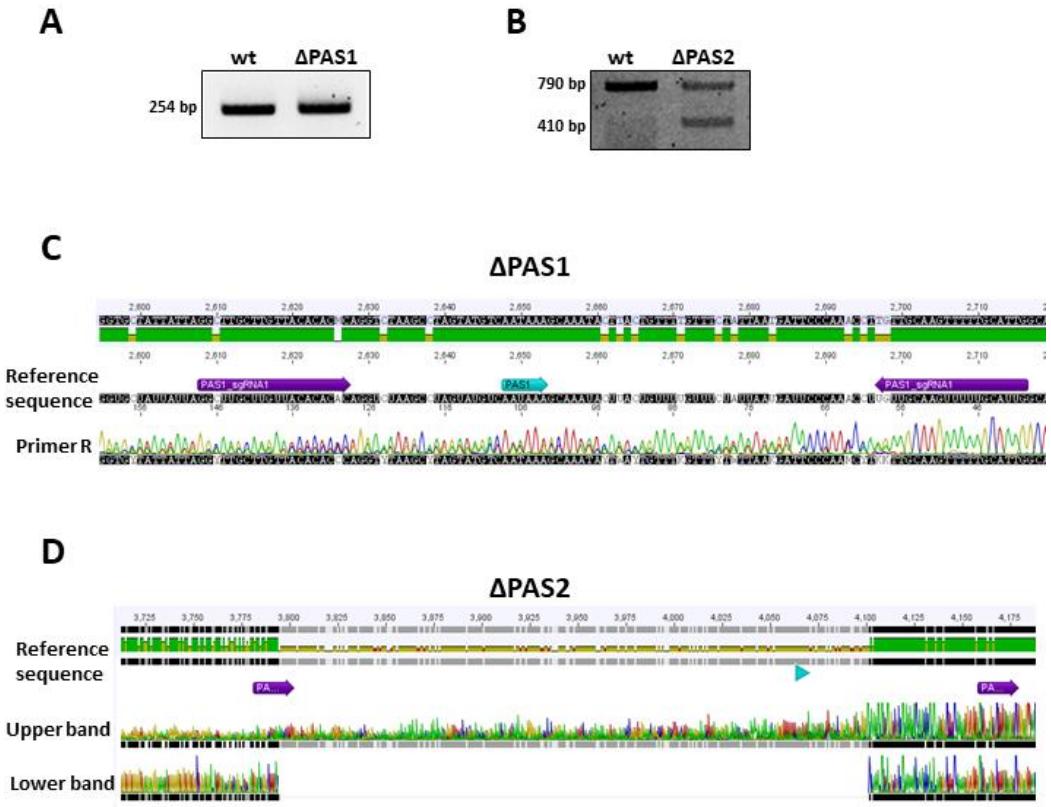


Figure 22: Genotyping of HAP1 cells subject to CRISPR/Cas9 technology to delete either *MCL1* PAS1 (ΔPAS1) or PAS2 (ΔPAS2) | (A,B) Electrophoretic profile of ΔPAS1 and ΔPAS2 conditions in comparison with non-edited cells (wt). On each panel, the correspondent molecular weight is indicated in the left; **(C,D)** Genomic alignment of *MCL1* 3'UTR sequence with the sequencing results obtained by sephadex DNA purification off all bands obtained for each condition. In the reference sequences are indicated the PASs (blue arrows) and their polyA sites (blue boxes), as well the sgRNAs (purple arrows) used for the genome editing.

We then performed cell sorting to obtain homozygous cell populations derived from single cells. Since the genome editing of the ΔPAS1 condition was not successful, we only performed sorting of ΔPAS2 condition. After ~ 2 weeks in culture, we observed growing of the single cell in some of the wells as depicted in Figure 23.

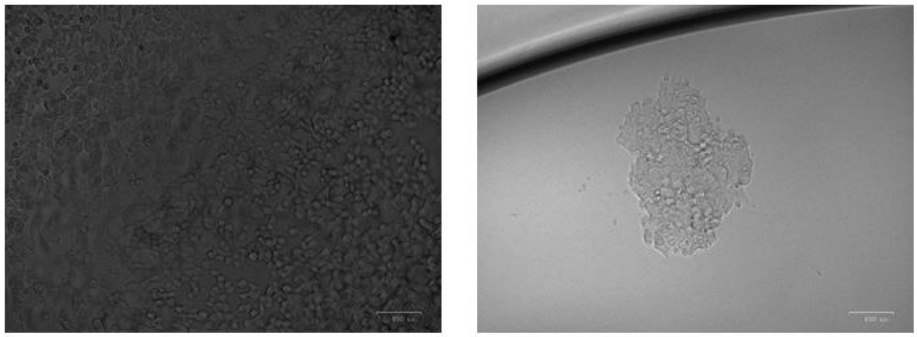


Figure 23: HAP1 clonal cells obtained by sorting of HAP1 CRISPR/Cas9 bulk population | Morphological aspect of a clonal cell line generated from a single cell (right) in comparison with a clonal cell line generated from 200 cells (left) after 2 weeks in culture in a 96-well plate.

Upon single-cell sorting of HAP1 ΔPAS2 bulk population, the survival rate of the clones was ~42% since from a total of 480 single-cell sorted clones only 200 survived and originated clonal cells in the 96-well plate. However, after PCR genotyping, only 10 of these 200 clones were possible homozygous (5%) but it was not possible to perform the functional assays, as cells did not survive after some time in culture. Figure 24 is a representative illustration showing only a few genotyped clones, with arrows indicating the homozygous clonal edited cells obtained. The figure also shows the presence of cells with the wt allele and a clonal cell line with two alleles (asterisk), suggesting that this clone is heterozygous. However, HAP1 is a haploid cell line, indicating that this cell population was probably generated from more than one single cell. Possibly a technical mistake occurred during sorting, resulting in the presence of two cells in the well that generated this clonal cell population.

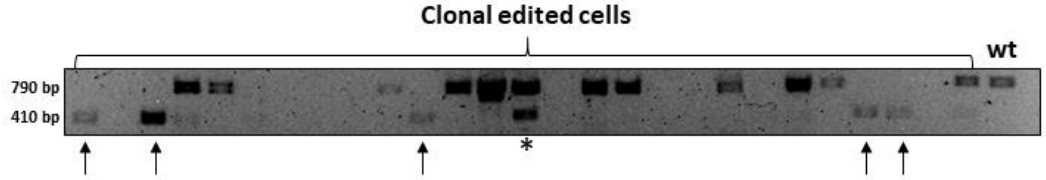


Figure 24: Genotyping of HAP1 clonal cell lines obtained by sorting of HeLa CRISRP/Cas9 bulk population | Electrophoretic profile of the DNA amplification of clonal cells from ΔPAS2 condition obtained from single cell sorting of CRISPR edited cells, in comparison with non-edited cells (wt). Arrows indicate homozygous clones and asterisk indicates a clonal cell line with two alleles.

The fact that homozygous clones did not survive could explain the low survival rate and, therefore, the difficulty in getting homozygous clones on HeLa cells. Together with the flow cytometry, these results strongly suggest that *MCL1* pA1 and pA2 contribute to Mcl-1 cellular function. Taking into account that nothing is known about the functional role of the two *MCL1* APA-derived mRNA isoforms in Mcl-1 cellular function, this is a very important result as it indicates that both pA1 and pA2 mRNAs are essential for cell viability.

2| Impact of the *MCL1* alternative 3'UTRs on Mcl-1 protein localization

2.1| *MCL1* alternative 3'UTRs influence Mcl-1 subcellular localization and protein levels

Although some studies have reported the presence of Mcl-1 in the nucleus, which is in agreement with its anti-proliferative function[57], it is well known that Mcl-1 localization is predominantly mitochondrial, where it performs its anti-apoptotic function[60]. It is also known that, the N-terminal region drives Mcl-1 to the mitochondria, and the C-terminal TM domain allows anchoring in the outer mitochondrial membrane[57,60].

Recent evidence indicate that alternative 3'UTRs play a role in protein localization[30]. This prompted us to evaluate the influence of *MCL1* alternative 3'UTRs on Mcl-1 protein subcellular localization. For this, our laboratory had previously overexpressed EGFP-*MCL1* constructs in HeLa cells and analyzed their expression and localization by confocal microscopy. In Figure 25A, cells were transfected with either pEGFP-*MCL1*CDS-pA1, harboring *MCL1* CDS followed by the pA1 3'UTR sequence, or pEGFP-*MCL1*CDS-pA2, with the proximal PAS mutated and the pA2 3'UTR sequence following *MCL1* CDS. To help in viewing the localization of the different constructs at the confocal microscope, mitochondria were stained with MitoTracker DeepRed dye and, the nucleus was stained with the Hoechst dye. Measurement of Mcl-1 protein levels through the quantification of EGFP expression, together with the analysis of the images obtained by confocal microscopy, allowed us to conclude that, despite varying only in their 3'UTR length, the protein production levels and subcellular distribution of Mcl-1 encoded by pA1 mRNA isoform is distinct, when comparing with Mcl-1 derived from the pA2 isoform. As shown in Figure 25B, Mcl-1 encoded by pA1 mRNA isoform has a well-defined mitochondrial accumulation, the standard Mcl-1 localization according to its anti-apoptotic function, whereas Mcl-1 encoded by pA2 mRNA isoform displays a more disperse distribution throughout the cell. This suggest that the subcellular distribution of Mcl-1 protein is regulated by the alternative 3'UTR of pA1 and pA2 mRNA isoforms. Furthermore, Mcl-1 pA2 encoded protein is less expressed than Mcl-1 encoded by pA1 mRNA (Figure 25C), which is in agreement with luciferase assays already done in the laboratory. The fact that Mcl-1 pA2 encoded protein is less expressed than Mcl-1 derived from pA1 mRNA, and displays different subcellular localization could indicate that Mcl-1 protein distribution depends on the regulation of protein expression.

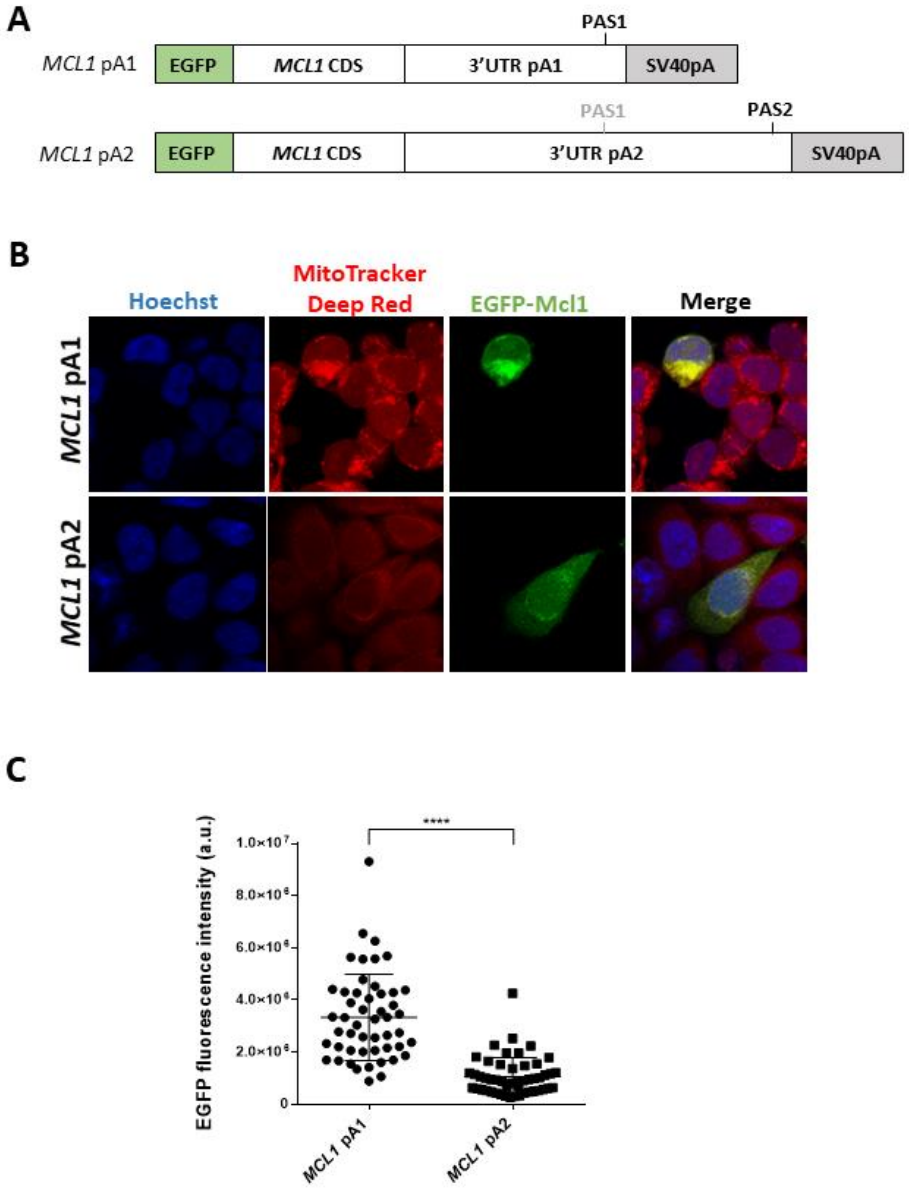


Figure 25: *MCL1* alternative 3'UTRs influence Mcl-1 subcellular localization and protein levels | (A) Scheme of EGFP-*MCL1* fused constructs. *MCL1* pA1 corresponds to the construct harboring *MCL1* CDS followed by the pA1 3'UTR sequence. *MCL1* pA2 corresponds to the construct with the PAS1 mutated and the pA2 3'UTR sequence following *MCL1* CDS. All constructs contain the SV40 polyA signal downstream of the 3'UTR sequence; **(B)** Fluorescence confocal microscopy of HeLa cells transfected with either *MCL1* pA1 or *MCL1* pA2 constructs (green). Cells were subjected to nucleus labelling with Hoechst (blue) and mitochondria labelling with MitoTracker Deep Red (red). Differences in protein expression levels are not detected in this figure since the images were acquired with different settings. Image acquisition was performed using Leica SP5 confocal microscope and ImageJ software. **(C)** Quantification of the EGFP fluorescence intensity obtained by the different constructs was performed in 50 cells using ImageJ and maintaining the same settings in all conditions. Statistical significance was determined by an unpaired *t* test with Welch's correction, where **** p<0.0001.

With these aforementioned results, we needed to understand whether the scattered localization of Mcl-1 pA2 encoded protein was due to the fact of being less expressed or by features contained in the longer 3'UTR. For that, we co-transfected cells with pEGFP-*MCL1*CDS-pA1 plus a *MCL1*-specific LNA, to decrease Mcl-1 protein expression. A scramble-LNA was used as a negative control. The expression and localization of the construct was analyzed by confocal microscopy (Figure 26).

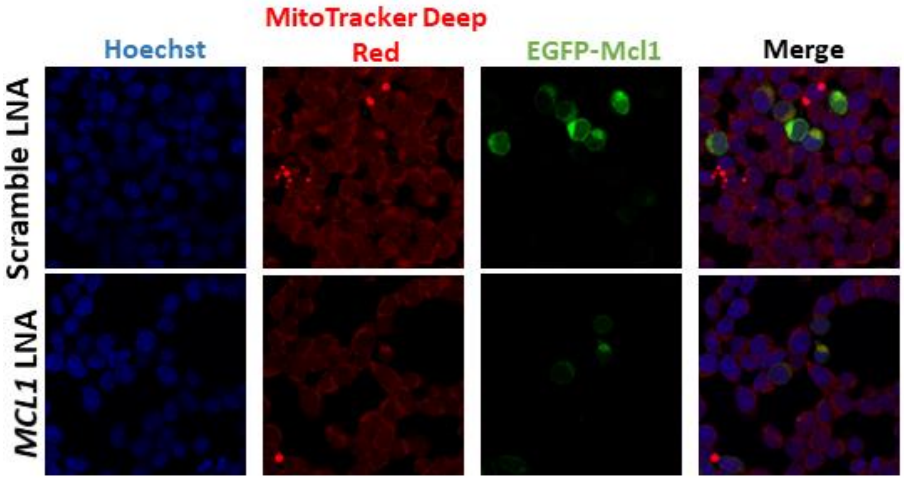


Figure 26: Mcl-1 localization throughout the cell is not due to a question of lower protein amount | Fluorescence confocal microscopy of HeLa cells co-transfected with *MCL1* pA1 (green) and either a MCL1-specific LNA or a scramble LNA. Cells were subjected to nucleus labelling with Hoechst (blue) and mitochondria labelling with MitoTracker Deep Red (red). Image acquisition was performed using Leica SP5 confocal microscope and ImageJ software.

If Mcl-1 protein distribution depended on the regulation of protein expression, decreasing Mcl-1 pA1 encoded protein levels would be expected to alter the Mcl-1 subcellular localization. Notably, we observed that, despite Mcl-1 expression levels decreased due to the LNA, Mcl-1 protein derived from pA1 mRNA isoform still localizes in the mitochondria. This result indicates that Mcl-1 diffused localization throughout the cell is not due to a question of a lower protein amount but rather due to a specific sequence or feature that is only present in the 3'UTR of pA2 mRNA isoform. Importantly, our result suggest that *MCL1* alternative 3'UTRs regulate Mcl-1 protein localization, as it was described by Mayr and Berkovits[30] for CD47 protein. Since pA2 3'UTR has 1410 more nucleotides (nt) than pA1 3'UTR, we decided to investigate the region of pA2 3'UTR involved in the regulation of Mcl-1 pA2 encoded protein localization.

2.2| A 3'UTR region of about 450 nt, between PAS1 and PAS2, is responsible for the subcellular localization of Mcl-1 encoded by pA2 mRNA

In order to define a specific sequence that is only present at the 3'UTR of pA2 isoform, responsible for modulating Mcl-1 subcellular localization, we created and analyzed by confocal microscopy EGFP-*MCL1* fused constructs with different 3'UTRs lengths (1920 and 2371 nt), comprised between PAS1 and PAS2 (Figure 27A). To that purpose, HeLa cells were transfected with one of the following constructs: pEGFP-*MCL1*CDS-pA1; pEGFP-*MCL1*CDS-1920, with the proximal PAS mutated and a 3'UTR sequence with 1920 nt, following *MCL1* CDS; pEGFP-*MCL1*CDS-2371, harboring the proximal PAS mutated and a 3'UTR sequence with 2371 nt, following *MCL1* CDS; and pEGFP-*MCL1*CDS-pA2. The results clearly show that, while pEGFP-*MCL1*CDS-1920 protein display the same mitochondrial localization as the Mcl-1 protein encoded by pA1 mRNA, pEGFP-*MCL1*CDS-2371 protein shows the scattered distribution as Mcl-1 encoded by pA2 (Figure 27B). These results lead to the identification of a region, of approximately 450 nt, located between the two PASs (from nucleotides 1920 to 2371 of the 3'UTR), that is responsible for the dispersed localization of Mcl-1, and therefore, responsible for regulating Mcl-1 subcellular localization. The molecular mechanism underlying this regulation of protein distribution is likely to involve binding of RBPs to the 450 bp region and the establishment of protein interactions.

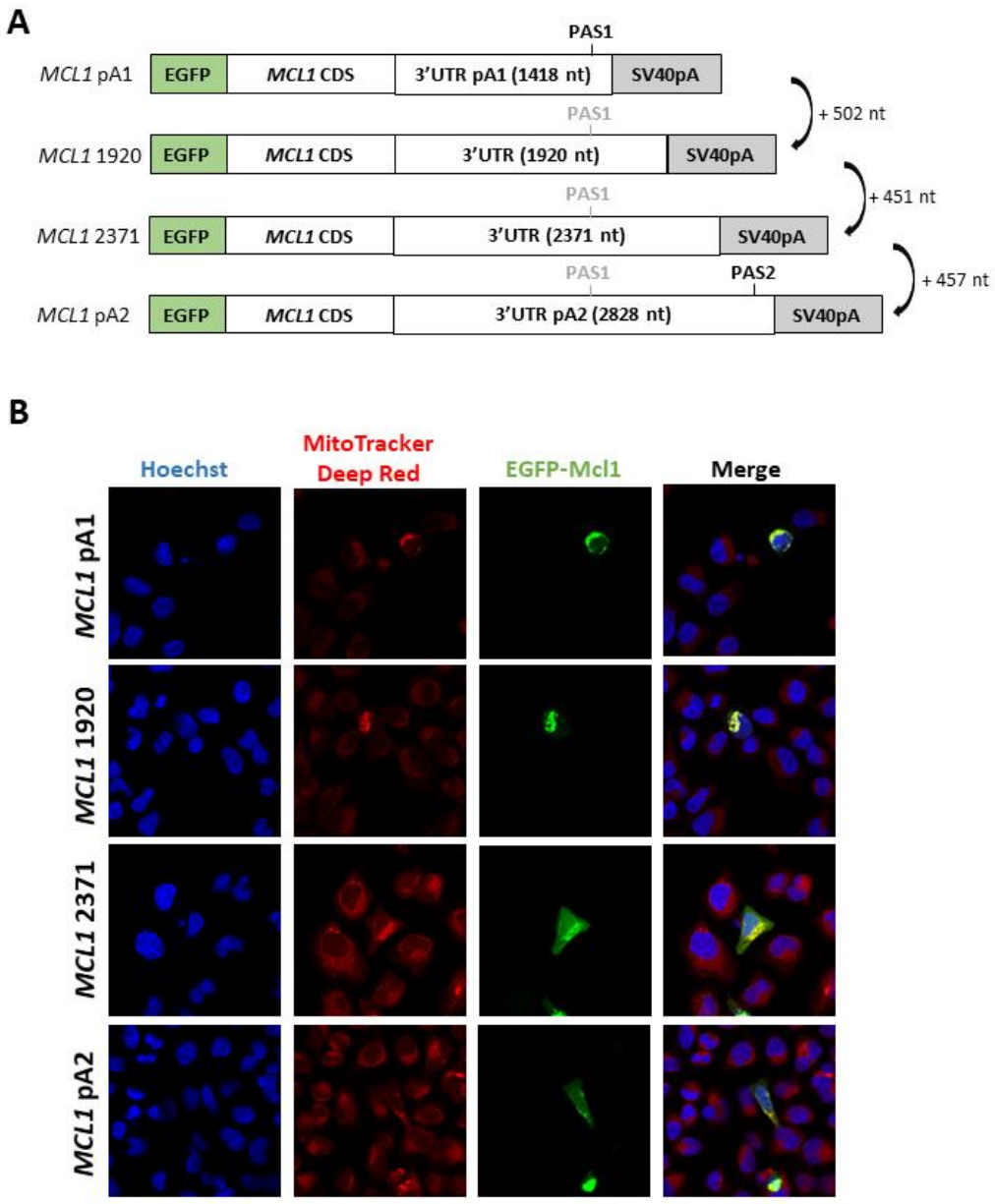


Figure 27: A 3'UTR sequence localized between 1920 and 2371 nucleotides regulates Mcl-1 localization | (A) Scheme of EGFP-*MCL1* fused constructs. *MCL1* pA1 corresponds to the construct harboring *MCL1* CDS followed by the pA1 3'UTR sequence. *MCL1* 1920 and *MCL1* 2371 represent the constructs harboring *MCL1* CDS followed by the 3'UTR sequence with 1920 and 2371 nt, respectively and with the PAS1 mutated. *MCL1* pA2 corresponds to the construct with the PAS1 mutated and the pA2 3'UTR sequence downstream of *MCL1* CDS. All constructs contain the SV40 polyA signal downstream of the 3'UTR sequence; **(B)** Fluorescence confocal microscopy of HeLa cells transfected with either *MCL1* pA1, *MCL1* 1920, *MCL1* 2371 or *MCL1* pA2 constructs (green). Cells were subjected to nucleus labelling with Hoechst (blue) and mitochondria labelling with MitoTracker Deep Red (red). Image acquisition was performed using Leica SP5 confocal microscope and ImageJ software.

Recently, Mayr and Berkovits[30] showed that *CD47* alternative 3'UTRs are responsible for defining the subcellular localization of the CD47 protein. The authors observed that CD47 protein encoded by the longest 3'UTR mRNA isoform localizes predominantly in the cell surface, while CD47 protein derived from the shorter 3'UTR mRNA isoform mainly localizes in the endoplasmic reticulum. Based on several

experiences, they propose a model of 3'UTR dependent protein localization (3'UDPL). According to the model, the HuR RBP binds to the long, but not the short, 3'UTR of *CD47* and recruits SET to the translation site. After *CD47* mRNA translation, SET interacts with the C terminus of the long encoded protein driving CD47 protein to the cell surface through active RAC1[30].

Inspired by Mayr's work[30] on 3′UTR-dependent protein interactors that control protein subcellular localization, we performed an *in silico* analysis to search for specific RBPs that could bind exclusively to the 3'UTR region between 1920 and 2371 nt of pA2 isoform. For that, we used RBPDB and RBPmap[92] databases, which revealed the existence of two RBPs that putatively could bind only to the 450 bp region (Figure 28). The putative binding location of the two candidate RBPs, Zinc finger protein 36 (ZFP36, also known as TTP) and Zinc finger Ran-binding domain-containing protein 2 (ZRANB2), is represented in figure 28.

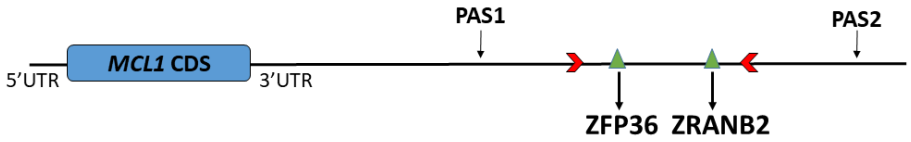


Figure 28: Putative binding sites of specific RBPs on *MCL1* 3'UTR | ZFP36 and ZRANB2 binding sites (green) are depicted in the illustration. The red arrows delimit the specifc 450 nt region. PAS1 and PAS2 are represented.

Binding of ZFP36 to the *Mcl1* mRNA was already observed in a recent study performed in mice[93]. Moreover, ZFP36 binds to AREs present in the 3'UTR of several mRNAs, being capable of destabilizing these transcripts since it promotes deadenylation or removal of their polyA tail[94]. In human cells, ZRANB2 is known to be present in the nucleus and to interact with important splicing factors such as U170K and U2AF[95]. The National Center for Biotechnology Information (NCBI) web page allowed us to conclude that the expression of ZFP36 and ZRANB2 is ubiquitous across different tissues. According to Mayr's model, HuR would be expected to be one of the candidate RBPs. However, this does not means that the molecular mechanism underlying the regulation of Mcl-1 protein distribution is not based on the 3'UDPL model. In fact, other RBPs capable of mediating protein-protein interactions, may also be able to recruit proteins to the translation site. Therefore, we suggest that ZFP36 or ZRANB2 may play a similar role to HuR.

Afterwards, we should perform RNA immunoprecipitation to prove if ZFP36 and ZRANB2 bind to the 450 bp region between 1920 and 2371 nt. For this, cells should be transfected with pEGFP-*MCL1*CDS-2371 construct followed by RNA

immunoprecipitation using antibodies for the RBPs and measurement of EGFP mRNA levels by RT-qPCR. The pEGFP-*MCL1*CDS-1920 should be used as a negative control, since it not harbors the RBP-binding sites.

2.3| *MCL1* alternative 3'UTRs regulate Mcl-1 protein levels

The previous observation that Mcl-1 pA2 encoded protein is less expressed than Mcl-1 protein derived from pA1 mRNA suggests that the pA2 3'UTR acts as an inhibitor of Mcl-1 protein production (Figure 25C). In fact, it is well stablished in the literature that longer 3'UTRs are more prone to negative regulation as they comprise more binding sites for miRNAs and RBPs[5,7]. Therefore we were interested in studying the function of *MCL1* alternative 3'UTRs in modulating Mcl-1 protein levels. For this, we measured Mcl-1 protein levels, by quantification of EGFP expression of pEGFP-*MCL1*CDS only, pEGFP-*MCL1*CDS-pA1 and pEGFP-*MCL1*CDS-pA2 encoded proteins. Comparing with pEGFP-*MCL1*CDS, the results showed that the presence of pA1 and pA2 3'UTR causes a large decrease in Mcl-1 protein expression (Figure 29A), meaning that the 3'UTR indirectly regulates Mcl-1 protein levels probably due to interactions with miRNAs and/or RBPs. This is in agreement with previous studies, performed in different cell lines, showing that Mcl-1 protein levels are inhibited by binding of miRNAs or RBPs to the 3'UTR. For instance, miR-29[96] and CUGBP2 RBP[97] overexpression was demonstrated to inhibit Mcl-1 expression in human H69 cholangiocytes and HCT-116 colon cancer cells, respectively. Comparing pEGFP-*MCL1*CDS-pA1 and pEGFP-*MCL1*CDS-pA2, we concluded that, as the 3'UTR length increases, the amount of protein produced decreases. According to this, our laboratory has previously discovered that the translation of *MCL1* pA2 mRNA is inhibited by miR-17 (unpublished results), which may explain the low expression levels of the Mcl-1 pA2 encoded protein. Several studies have established an association between miRNA-mediated regulation and the low protein levels produced by mRNAs with longer 3'UTRs. For instance, the presence of three binding sites for miR-15/16 in the long 3'UTR of *Cyclin D2* has been shown to be responsible for the low protein levels produced by this isoform[36].

Measurement of Mcl-1 protein levels in EGFP-*MCL1* fused constructs with different 3'UTRs lengths (pEGFP-*MCL1*CDS-1920 and pEGFP-*MCL1*CDS-2371) allowed us to reinforce the negative correlation between the 3'UTR length and Mcl-1 expression levels (Figure 29B). Given that miR-17 binding site is within the 450 nt region downstream of 1920 nt, it is likely to explain the decrease on Mcl-1 protein levels in pEGFP-*MCL1*CDS-2371 comparing with pEGFP-*MCL1*CDS-1920. This reinforces our

previous observation that miR-17 mediated inhibition regulates Mcl-1 protein levels. Inhibition of Mcl-1 protein levels can also be mediated by RBPs, as the above-mentioned CUGBP2. Taking into account that, ZFP36 and ZRANB2 putatively bind to the 450 nt region downstream of 1920 nt, as miR-17, perhaps they are involved in the regulation of Mcl-1 protein levels. In fact, the lack of statistical significant difference between Mcl-1 protein levels in pEGFP-*MCL1*CDS-pA1 and pEGFP-*MCL1*CDS-1920, reinforces our speculations concerning the role of miR-17, ZFP36 and ZRANB2 on the regulation of Mcl-1 protein levels. Therefore, inhibition of Mcl-1 expression may depend on a synergistic effect resulting from the combined participation of miRNAs, like miR-17, together with the two RBPs. Moreover, the difference in Mcl-1 protein levels between pEGFP-*MCL1*CDS-2371 and pEGFP-*MCL1*CDS-pA2 suggests that the regulation of Mcl-1 protein levels also relies on the concerted action of several miRNAs and RBPs whose binding sites are not present between the 1920 and 2371 nt. Accordingly, we can not exclude the possibility that the regulation of Mcl-1 protein levels could also rely on RBPs and miRNA that bind across the 3'UTR of *MCL1* and be dependent on their stoichiometric quantity[30].

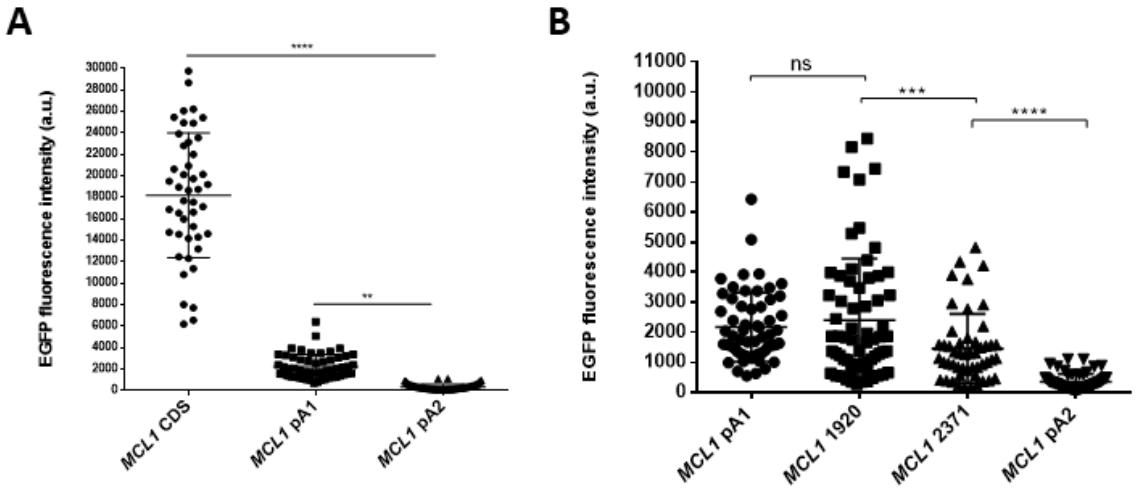


Figure 29: *MCL1* 3'UTR length influences Mcl-1 protein expression | (A) Quantification of the EGFP fluorescence intensity obtained by *MCL1* CDS, *MCL1* pA1 and *MCL1* pA2 constructs; **(B)** Quantification of the EGFP fluorescence intensity obtained by *MCL1* pA1, *MCL1* 1920, *MCL1* 2371 and *MCL1* pA2 constructs. (A) and (B) quantifications were performed with 50 cells using ImageJ and maintaining the same settings in all conditions. Statistical significance was determined by one-way ANOVA with Holm-Sidak's correction, where **=p<0.01, ***=p<0.001 ****=p< 0.0001, ns = not significant.

2.4| Working model for the role of *MCL1* 3'UTRs on Mcl-1 protein localization

We hypothesize a working model concerning the role of the 3'UTR on Mcl-1 protein localization (Figure 30). In this model, we suggest that one of the two candidate RBPs (protein A), capable of binding to the 450 nt region of the 3'UTR of pA2 mRNA isoform, recruits another protein (protein B) that will interact with Mcl-1 pA2 encoded protein. Upon *MCL1* pA2 mRNA translation, we suggest that the interaction of protein B with Mcl-1 protein is responsible for the more dispersed subcellular localization. The molecular mechanisms underlying this process remain unknown but may be based on the possibility that protein B interaction with Mcl-1 pA2 encoded protein decreases the Noxa's efficiency in recruiting Mcl-1 for mitochondria. Noxa is a BH3-only protein, with mitochondrial localization, known for interacting with Mcl-1 and recruiting it to mitochondria[98]. Taking into account that, the subcellular localization of Mcl-1 pA2 encoded protein is more dispersed but mitochondria accumulation is still observed, we suggest that protein B interaction with Mcl-1 pA2 encoded protein does not inhibit Mcl-1 interaction with Noxa but makes it more difficult. This explains the fact that Mcl-1 pA1 encoded protein shows a well-defined mitochondrial accumulation, while Mcl-1 encoded by pA2 mRNA isoform, displays a more disperse distribution. According to this, a study showed that lack of Noxa results in diffuse Mcl-1 distribution, without evident mitochondrial accumulation[98]. Since the N-terminal region and the C-terminal TM domain are involved in mitochondrial localization of Mcl-1 and its anchoring in the outer mitochondrial membrane, it is also possible that protein B interaction with Mcl-1 hides the N-terminal region and/or the C-terminal TM domain, regardless of Mcl-1 interaction with Noxa, making the mitochondrial enrichment more difficult.

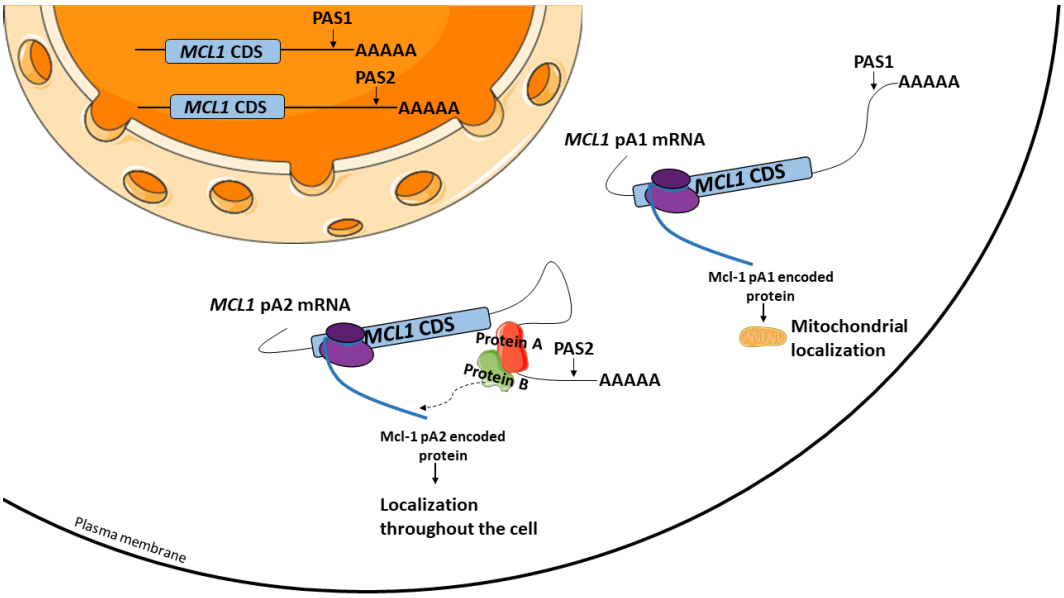


Figure 30: Working model for the role of *MCL1* 3'UTRs on Mcl-1 protein localization | We hypothesize that protein A (red RBP) binds to the 3'UTR of *MCL1* pA2 mRNA isoform and recruits protein B (green), which interacts with Mcl-1 pA2 encoded protein. Upon *MCL1* pA2 mRNA translation, we suggest that the interaction of protein B with Mcl-1 protein is responsible for the scattered subcellular localization.

Conclusions

The main purpose of this work was to go deeper into the knowledge concerning the regulation of *MCL1* APA isoforms. Although many studies focused on *MCL1* regulation by alternative splicing, there were no studies performed regarding the regulation of *MCL1* by APA. The aims of this work are highly relevant for the gene regulation field because it is well known that APA plays a major role in gene expression. In fact, more than 70% of mammalian genes can be regulated by APA since they possess more than one PAS in its sequence[8]. About 50% of human genes can generate mRNA isoforms with different 3'UTRs by APA. It is well established that, alternative 3'UTRs influence mRNA stability, localization and translation[5,25]. Moreover, it was recently demonstrated that alternative 3'UTRs are capable of regulating protein function[29] and localization[28,30]. Since nothing was known concerning the functional role of the two *MCL1* APA-derived mRNA isoforms, this study focused on understanding their physiological effect on Mcl-1 function, mainly in cell survival, and their impact on Mcl-1 protein production and subcellular distribution.

Regarding the CRISPR technology, we concluded that there are some complications associated to the two-sgRNA approach used. First, when the two sgRNAs are resistant to the same antibiotic, it is not possible to select only the cells infected with both sgRNAs. Second, to delete the sgRNAs flanked region, is not enough to select cells infected with the two sgRNAs, it is also required that both work simultaneously. Despite these difficulties, we concluded that genome editing of untranslated regions impacts protein levels. Importantly, we showed that a small reduction in pA1 and pA2 mRNA expression leads to a significant impact on Mcl-1 protein levels in Jurkat E6.1 and HeLa cells. Flow Cytometry analysis in the CRISPR edited cells showed that ΔPAS1 and ΔPAS2 conditions increase the apoptosis levels allowing concluding that both pA1 and pA2 mRNAs contribute to Mcl-1 cellular function. The results obtained were consistent in both Jurkat E6.1 and HeLa cells, indicating that the functional role of the two *MCL1* APA-derived mRNA isoforms is not cell type-specific. The fact that functional assays could not be performed in the HAP1 homozygous clonal population because cells did not survive, strongly suggest that the two *MCL1*-APA derived mRNA isoforms are vital for cell survival.

Regarding Mcl-1 subcellular localization, we showed that Mcl-1 dispersed distribution is not due to the lower amounts of protein produced by pA2 mRNA, but rather due to a specific RNA structure or sequence that is only present at the 3'UTR of pA2 mRNA. This allowed us to conclude that Mcl-1 protein localization is regulated by alternative 3'UTRs. Furthermore, we concluded that this specific feature lies in a region of approximately 450 nt, located between PAS1 and PAS2. In addition, we found two

RBPs (ZFP36 and ZRANB2) that putatively bind to the 450 bp region only, and proposed them as candidates involved in the regulation of Mcl-1 subcellular localization. With these results, we hypothesize a model based on the existence of a RBP (protein A) that binds to the 450 nt region present at the 3'UTR of pA2 mRNA and recruits another protein (protein B) who interacts with Mcl-1 protein. After translation of *MCL1* pA2 mRNA, protein B is likely to interact with the newly translated Mcl-1 protein causing the more dispersed localization. Moreover, we speculate that protein B interaction with Mcl-1 pA2 encoded protein decreases the Noxa's efficiency in recruiting Mcl-1 for the mitochondria. On the other hand, we can also suggest that the interaction of protein B with Mcl-1 masks the N-terminal region and/or the C-terminal TM domain, resulting in less mitochondria accumulation.

Concerning the protein expression, the presence of *MCL1* alternative 3'UTRs reduces Mcl-1 protein levels, allowing us to conclude that the 3'UTR act as a negative regulator of Mcl-1 protein production. In fact, we established a negative correlation between the 3'UTR length and Mcl-1 protein expression levels. Taking into account that miR-17 binding site lies within de 450 bp region, located between the 1920 and 2371 nt, we speculate that ZFP36 and ZRANB2, which bind in this region too, may also be involved in the regulation of Mcl-1 expression levels. In addition, we propose that other miRNAs and/or RBPs, whose binding sites are not encountered between the 1920 and 2371 nt are likely to be involved in the regulation of Mcl-1 protein levels.

Taken together, our results allowed understanding the role of the two *MCL1*-APA derived isoforms in Mcl-1 protein production, function and localization. We demonstrated that both pA1 and pA2 *MCL1* APA-derived mRNA isoforms contribute to Mcl-1 anti-apoptotic function. In addition, we also concluded that Mcl-1 protein expression and subcellular localization is regulated by *MCL1* alternative 3'UTRs. In general, this research allowed reinforcing the knowledge concerning the function of 3'UTR-APA in gene expression. In particular, this work enlightened the regulation and function of *MCL1* APA isoforms in human cells.

Future perspectives

In order to unravel the physiological impact of each isoform on Mcl-1 function, and since the homozygous clones did not survive, we intend to perform further functional assays in CRISPR heterozygous cells. Regarding cell survival, we propose assessing cell viability through other assays, such as the MTT assay. We expect to have results complementary to those obtained with the Annexin V-PI assay and, if this is the case, it will allow us to validate and increase the robustness of our results. Since Mcl-1 anti-apoptotic function is accomplished in mitochondria, we propose to evaluate mitochondrial activity by measuring, for instance, cytochrome c release or the mitochondrial transmembrane potential. Furthermore, we propose performing immunofluorescence using a Mcl-1 antibody on the CRISPR cells with pA1 or pA2 deletion, in order to evaluate the endogenous expression and localization of pA1 or pA2 encoded proteins without the need of using constructs overexpression.

Concerning the regulation of Mcl-1 expression and subcellular localization, we seek to dissect the molecular mechanisms behind the differences in Mcl-1 protein levels and localization. To prove the putative binding of ZFP36 and ZRANB2 to the 3'UTR region between 1920 and 2371 nt, we intend to transfect cells with the pEGFP-*MCL1*CDS-2371 construct and perform RNA immunoprecipitation using antibodies against ZFP36 and ZRANB2 proteins. As a negative control, we will use the pEGFP-*MCL1*CDS-1920 construct, as it not harbors binding sites for those RBPs. The obtained results will determine the candidate RBPs that regulate Mcl-1 subcellular localization. Given that the RBPs can bind to *MCL1* mRNA without altering the subcellular localization of Mcl-1, it would be interesting to individually knockdown the candidate RBPs and analyze the localization of Mcl-1 protein encoded by the pA2 mRNA isoform. This will allow the identification of the specific RBP that modulates the subcellular distribution of Mcl-1 pA2 encoded protein. Afterwards, we intend to unravel the specific protein B that, according to our model, interacts with protein A and the Mcl-1 protein encoded by pA2 mRNA. For that, we will search for protein interactions using online databases like BIOGRID and Interactome. To validate the interaction of the *in silico* identified protein B with Mcl-1 pA2 encoded protein, we intend to immunoprecipitate pEGFP-*MCL1*CDS-pA2 encoded protein, using a GFP antibody, and test if in fact it interacts with protein B. We should use pA1 encoded protein as a control, since it is not expected to interact with protein B. To confirm protein B interaction with protein A, we plan to pull-down the specific RBP by immunoprecipitation and test if it interacts with protein B by western blot. These experiments will allow identifying and validating the proteins of our proposed model.

References

1. Alberts B, Wilson JH, Hunt T. *Molecular biology of the cell*, 5th edn. Garland Science (2008).

2. Carlos Azevedo CES. Biologia Celular e Molecular. *Lidel*, (2012).

3. Lodish H. *Molecular Cell Biology*. W. H. Freeman (2008).

4. Millevoi S, Vagner S. Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. *Nucleic acids research* **38**, 2757-2774 (2010).

5. Lutz CS, Moreira A. Alternative mRNA polyadenylation in eukaryotes: an effective regulator of gene expression. *Wiley Interdiscip Rev RNA* **2**, 22-31 (2011).

6. Di Giammartino DC, Nishida K, Manley JL. Mechanisms and consequences of alternative polyadenylation. *Mol Cell* **43**, 853-866 (2011).

7. Curinha A, Oliveira Braz S, Pereira-Castro I, Cruz A, Moreira A. Implications of polyadenylation in health and disease. *Nucleus* **5**, 508-519 (2014).

8. Tian B, Manley JL. Alternative cleavage and polyadenylation: the long and short of it. *Trends Biochem Sci* **38**, 312-320 (2013).

9. Tian B, Manley JL. Alternative polyadenylation of mRNA precursors. *Nat Rev Mol Cell Biol*, (2016).

10. Xiao MS, Zhang B, Li YS, Gao Q, Sun W, Chen W. Global analysis of regulatory divergence in the evolution of mouse alternative polyadenylation. *Mol Syst Biol* **12**, 890 (2016).

11. Yeh HS, Yong J. Alternative Polyadenylation of mRNAs: 3'-Untranslated Region Matters in Gene Expression. *Mol Cells* **39**, 281-285 (2016).

12. Danckwardt S*, et al.* Splicing factors stimulate polyadenylation via USEs at non-canonical 3' end formation signals. *EMBO J* **26**, 2658-2669 (2007).

13. Tian B, Hu J, Zhang H, Lutz CS. A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic acids research* **33**, 201-212 (2005).

14. Erson-Bensan AE. Alternative polyadenylation and RNA-binding proteins. *J Mol Endocrinol* **57**, F29-34 (2016).

15. Shi Y*, et al.* Molecular architecture of the human pre-mRNA 3' processing complex. *Mol Cell* **33**, 365-376 (2009).

16. Elkon R, Ugalde AP, Agami R. Alternative cleavage and polyadenylation: extent, regulation and function. *Nat Rev Genet* **14**, 496-506 (2013).

17. Mandel CR, Bai Y, Tong L. Protein factors in pre-mRNA 3'-end processing. *Cell Mol Life Sci* **65**, 1099-1122 (2008).

18. Chan SL*, et al.* CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3' processing. *Genes Dev* **28**, 2370-2380 (2014).

19. Schonemann L*, et al.* Reconstitution of CPSF active in polyadenylation: recognition of the polyadenylation signal by WDR33. *Genes Dev* **28**, 2381-2393 (2014).

20. Sun Y*, et al.* Molecular basis for the recognition of the human AAUAAA polyadenylation signal. *Proc Natl Acad Sci U S A* **115**, E1419-E1428 (2018).

21. Brown KM, Gilmartin GM. A mechanism for the regulation of pre-mRNA 3' processing by human cleavage factor Im. *Mol Cell* **12**, 1467-1476 (2003).

22. Venkataraman K, Brown KM, Gilmartin GM. Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition. *Genes Dev* **19**, 1315-1327 (2005).

23. Schafer P*, et al.* Reconstitution of mammalian cleavage factor II involved in 3' processing of mRNA precursors. *RNA* **24**, 1721-1737 (2018).

24. Zheng D, Tian B. RNA-binding proteins in regulation of alternative cleavage and polyadenylation. *Adv Exp Med Biol* **825**, 97-127 (2014).

25. Braz SO*, et al.* Expression of Rac1 alternative 3' UTRs is a cell specific mechanism with a function in dendrite outgrowth in cortical neurons. *Biochim Biophys Acta Gene Regul Mech* **1860**, 685-694 (2017).

26. Derti A*, et al.* A quantitative atlas of polyadenylation in five mammals. *Genome Res* **22**, 1173-1183 (2012).

27. Hoque M*, et al.* Analysis of alternative cleavage and polyadenylation by 3' region extraction and deep sequencing. *Nat Methods* **10**, 133-139 (2013).

28. Mayr C. Evolution and Biological Roles of Alternative 3'UTRs. *Trends Cell Biol* **26**, 227-237 (2016).

29. Lee SH, Mayr C. Gain of Additional BIRC3 Protein Functions through 3'-UTR-Mediated Protein Complex Formation. *Mol Cell* **74**, 701-712 e709 (2019).

30. Berkovits BD, Mayr C. Alternative 3' UTRs act as scaffolds to regulate membrane protein localization. *Nature* **522**, 363-367 (2015).

31. Grassi E, Mariella E, Lembo A, Molineris I, Provero P. Roar: detecting alternative polyadenylation with standard mRNA sequencing libraries. *BMC Bioinformatics* **17**, 423 (2016).

32. Legendre M, Ritchie W, Lopez F, Gautheret D. Differential repression of alternative transcripts: a screen for miRNA targets. *PLoS Comput Biol* **2**, e43 (2006).

33. Sandberg R, Neilson JR, Sarma A, Sharp PA, Burge CB. Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science* **320**, 1643-1647 (2008).

34. Lianoglou S, Garg V, Yang JL, Leslie CS, Mayr C. Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. *Genes Dev* **27**, 2380-2396 (2013).

35. Liu D*, et al.* Systematic variation in mRNA 3'-processing signals during mouse spermatogenesis. *Nucleic acids research* **35**, 234-246 (2007).

36. Mayr C, Bartel DP. Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* **138**, 673-684 (2009).

37. Diaz-Munoz MD, Turner M. Uncovering the Role of RNA-Binding Proteins in Gene Expression in the Immune System. *Front Immunol* **9**, 1094 (2018).

38. Ji Z, Lee JY, Pan Z, Jiang B, Tian B. Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. *Proc Natl Acad Sci U S A* **106**, 7028-7033 (2009).

39. Zhang H, Lee JY, Tian B. Biased alternative polyadenylation in human tissues. *Genome Biol* **6**, R100 (2005).

40. Evsikov AV*, et al.* Cracking the egg: molecular dynamics and evolutionary aspects of the transition from the fully grown oocyte to embryo. *Genes Dev* **20**, 2713-2727 (2006).

41. Braz SO*, et al.* Expression of Rac1 alternative 3' UTRs is a cell specific mechanism with a function in dendrite outgrowth in cortical neurons. *Biochim Biophys Acta*, (2017).

42. Alt FW*, et al.* Synthesis of secreted and membrane-bound immunoglobulin mu heavy chains is directed by mRNAs that differ at their 3' ends. *Cell* **20**, 293-301 (1980).

43. Takagaki Y, Seipelt RL, Peterson ML, Manley JL. The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B cell differentiation. *Cell* **87**, 941-952 (1996).

44. Xia Z*, et al.* Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3'-UTR landscape across seven tumour types. *Nat Commun* **5**, 5274 (2014).

45. Ji Z, Tian B. Reprogramming of 3' untranslated regions of mRNAs by alternative polyadenylation in generation of pluripotent stem cells from different cell types. *PLoS One* **4**, e8419 (2009).

46. Neve J*, et al.* Subcellular RNA profiling links splicing and nuclear DICER1 to alternative cleavage and polyadenylation. *Genome Res* **26**, 24-35 (2016).

47. Li W, Park JY, Zheng D, Hoque M, Yehia G, Tian B. Alternative cleavage and polyadenylation in spermatogenesis connects chromatin regulation with post-transcriptional control. *BMC Biol* **14**, 6 (2016).

48. Pinto PA*, et al.* RNA polymerase II kinetics in polo polyadenylation signal selection. *EMBO J* **30**, 2431-2444 (2011).

49. Yonaha M, Proudfoot NJ. Specific transcriptional pausing activates polyadenylation in a coupled in vitro system. *Mol Cell* **3**, 593-600 (1999).

50. Lutz CS, Murthy KG, Schek N, O'Connor JP, Manley JL, Alwine JC. Interaction between the U1 snRNP-A protein and the 160-kD subunit of cleavage-polyadenylation specificity factor increases polyadenylation efficiency in vitro. *Genes Dev* **10**, 325-337 (1996).

51. Gunderson SI, Polycarpou-Schwarz M, Mattaj IW. U1 snRNP inhibits pre-mRNA polyadenylation through a direct interaction between U1 70K and poly(A) polymerase. *Mol Cell* **1**, 255-264 (1998).

52. Kaida D*, et al.* U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* **468**, 664-668 (2010).

53. Berg MG*, et al.* U1 snRNP determines mRNA length and regulates isoform expression. *Cell* **150**, 53-64 (2012).

54. Castelo-Branco P, Furger A, Wollerton M, Smith C, Moreira A, Proudfoot N. Polypyrimidine tract binding protein modulates efficiency of polyadenylation. *Mol Cell Biol* **24**, 4174-4183 (2004).

55. Licatalosi DD*, et al.* HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**, 464-469 (2008).

56. Thomas LW, Lam C, Edwards SW. Mcl-1; the molecular regulation of protein function. *FEBS Lett* **584**, 2981-2989 (2010).

57. Germain M, Duronio V. The N terminus of the anti-apoptotic BCL-2 homologue MCL-1 regulates its localization and function. *J Biol Chem* **282**, 32233-32242 (2007).

58. Ashkenazi A, Fairbrother WJ, Leverson JD, Souers AJ. From basic apoptosis discoveries to advanced selective BCL-2 family inhibitors. *Nat Rev Drug Discov* **16**, 273-284 (2017).

59. Michels J, Johnson PW, Packham G. Mcl-1. *Int J Biochem Cell Biol* **37**, 267-271 (2005).

60. Akgul C. Mcl-1 is a potential therapeutic target in multiple types of cancer. *Cell Mol Life Sci* **66**, 1326-1336 (2009).

61. Chipuk JE, Moldoveanu T, Llambi F, Parsons MJ, Green DR. The BCL-2 family reunion. *Mol Cell* **37**, 299-310 (2010).

62. Belmar J, Fesik SW. Small molecule Mcl-1 inhibitors for the treatment of cancer. *Pharmacol Ther* **145**, 76-84 (2015).

63. Craig RW. MCL1 provides a window on the role of the BCL2 family in cell proliferation, differentiation and tumorigenesis. *Leukemia* **16**, 444-454 (2002).

64. Anstee NS, Vandenberg CJ, Campbell KJ, Hughes PD, O'Reilly LA, Cory S. Overexpression of Mcl-1 exacerbates lymphocyte accumulation and autoimmune kidney disease in lpr mice. *Cell Death Differ* **24**, 397-408 (2017).

65. Carrington EM*, et al.* Anti-apoptotic proteins BCL-2, MCL-1 and A1 summate collectively to maintain survival of immune cell populations both in vitro and in vivo. *Cell Death Differ* **24**, 878-888 (2017).

66. Xiang W, Yang CY, Bai L. MCL-1 inhibition in cancer treatment. *Onco Targets Ther* **11**, 7301-7314 (2018).

67. Rinkenberger JL, Horning S, Klocke B, Roth K, Korsmeyer SJ. Mcl-1 deficiency results in peri-implantation embryonic lethality. *Genes Dev* **14**, 23-27 (2000).

68. Opferman JT, Letai A, Beard C, Sorcinelli MD, Ong CC, Korsmeyer SJ. Development and maintenance of B and T lymphocytes requires antiapoptotic MCL-1. *Nature* **426**, 671-676 (2003).

69. Arbour N*, et al.* Mcl-1 is a key regulator of apoptosis during CNS development and after DNA damage. *J Neurosci* **28**, 6068-6078 (2008).

70. Steimer DA, Boyd K, Takeuchi O, Fisher JK, Zambetti GP, Opferman JT. Selective roles for antiapoptotic MCL-1 during granulocyte development and macrophage effector function. *Blood* **113**, 2805-2815 (2009).

71. Edwards SW, Derouet M, Howse M, Moots RJ. Regulation of neutrophil apoptosis by Mcl-1. *Biochem Soc Trans* **32**, 489-492 (2004).

72. Jamil S, Sobouti R, Hojabrpour P, Raj M, Kast J, Duronio V. A proteolytic fragment of Mcl-1 exhibits nuclear localization and regulates cell growth by interaction with Cdk1. *Biochem J* **387**, 659-667 (2005).

73. Fujise K, Zhang D, Liu J, Yeh ET. Regulation of apoptosis and cell cycle progression by MCL1. Differential role of proliferating cell nuclear antigen. *J Biol Chem* **275**, 39458-39465 (2000).

74. Jamil S, Mojtabavi S, Hojabrpour P, Cheah S, Duronio V. An essential role for MCL-1 in ATR-mediated CHK1 phosphorylation. *Mol Biol Cell* **19**, 3212-3220 (2008).

75. Quinn BA*, et al.* Targeting Mcl-1 for the therapy of cancer. *Expert Opin Investig Drugs* **20**, 1397-1411 (2011).

76. Bae J, Leo CP, Hsu SY, Hsueh AJ. MCL-1S, a splicing variant of the antiapoptotic BCL-2 family member MCL-1, encodes a proapoptotic protein possessing only the BH3 domain. *J Biol Chem* **275**, 25255-25261 (2000).

77. Kim JH, Sim SH, Ha HJ, Ko JJ, Lee K, Bae J. MCL-1ES, a novel variant of MCL-1, associates with MCL-1L and induces mitochondrial cell death. *FEBS Lett* **583**, 2758-2764 (2009).

78. Dzhagalov I, Dunkle A, He YW. The anti-apoptotic Bcl-2 family member Mcl-1 promotes T lymphocyte survival at multiple stages. *J Immunol* **181**, 521-528 (2008).

79. Opferman JT*, et al.* Obligate role of anti-apoptotic MCL-1 in the survival of hematopoietic stem cells. *Science* **307**, 1101-1104 (2005).

80. Pierson W*, et al.* Antiapoptotic Mcl-1 is critical for the survival and niche-filling capacity of Foxp3(+) regulatory T cells. *Nat Immunol* **14**, 959-965 (2013).

81. Barrangou R, Horvath P. A decade of discovery: CRISPR functions and applications. *Nat Microbiol* **2**, 17092 (2017).

82. van der Oost J, Westra ER, Jackson RN, Wiedenheft B. Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nat Rev Microbiol* **12**, 479-492 (2014).

83. Mali P*, et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823-826 (2013).

84. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8**, 2281-2308 (2013).

85. Komor AC, Badran AH, Liu DR. CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. *Cell* **169**, 559 (2017).

86. Nishimasu H*, et al.* Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science* **361**, 1259-1262 (2018).

87. Sanjana NE, Shalem O, Zhang F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat Methods* **11**, 783-784 (2014).

88. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402-408 (2001).

89. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic acids research* **29**, e45 (2001).

90. Song Y*, et al.* CRISPR/Cas9-mediated mutation of tyrosinase (Tyr) 3' UTR induce graying in rabbit. *Sci Rep* **7**, 1569 (2017).

91. Chen LS, Keating MJ, Wierda WG, Gandhi V. Induction of apoptosis by MCL-1 inhibitors in chronic lymphocytic leukemia cells. *Leuk Lymphoma*, 1-4 (2019).

92. Paz I, Kosti I, Ares M, Jr., Cline M, Mandel-Gutfreund Y. RBPmap: a web server for mapping binding sites of RNA-binding proteins. *Nucleic acids research* **42**, W361-367 (2014).

93. Ebner F*, et al.* The RNA-binding protein tristetraprolin schedules apoptosis of pathogen-engaged neutrophils during bacterial infection. *J Clin Invest* **127**, 2051-2065 (2017).

94. Hausburg MA*, et al.* Post-transcriptional regulation of satellite cell quiescence by TTP-mediated mRNA decay. *Elife* **4**, e03390 (2015).

95. Mangs AH, Morris BJ. ZRANB2: structural and functional insights into a novel splicing protein. *Int J Biochem Cell Biol* **40**, 2353-2357 (2008).

96. Mott JL, Kobayashi S, Bronk SF, Gores GJ. mir-29 regulates Mcl-1 protein expression and apoptosis. *Oncogene* **26**, 6133-6140 (2007).

97. Subramaniam D*, et al.* Translation inhibition during cell cycle arrest and apoptosis: Mcl-1 is a novel target for RNA binding protein CUGBP2. *Am J Physiol Gastrointest Liver Physiol* **294**, G1025-1032 (2008).

98. Nakajima W, Hicks MA, Tanaka N, Krystal GW, Harada H. Noxa determines localization and stability of MCL-1 and consequently ABT-737 sensitivity in small cell lung cancer. *Cell Death Dis* **5**, e1052 (2014).