UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA VEGETAL

RAC1B alternative splicing regulation in colorectal cancer cells

Telma Filipa Almas Rosado

Mestrado em Biologia Molecular e Genética

Dissertação orientada por:
Doutora Vânia Gonçalves (orientação externa)
Doutor Peter Jordan (orientação interna)

2020

Agradecimentos

Não poderia deixar passar esta oportunidade para fazer alguns agradecimentos a quem me ajudou ao longo desta importante etapa do meu percurso académico. Este foi um caminho com altos e baixos (a ciência é mesmo assim) e bastantes imprevistos, mas tudo se torna mais fácil quando estamos rodeados de pessoas fantásticas.

Em primeiro lugar, quero agradecer aos meus orientadores, Doutora Vânia Gonçalves e Doutor Peter Jordan, por me terem recebido no Laboratório de Oncobiologia e Vias de Sinalização e demonstrarem logo desde o início a sua disponibilidade para me ajudar em tudo o que eu precisasse. Estou muito grata pela confiança que em mim depositaram, pela oportunidade que me deram para crescer tanto profissional como pessoalmente, e também por toda a paciência que tiveram e a ajuda que me deram ao longo de todo este tempo e na correção deste manuscrito.

Quero também deixar um agradecimento aos restantes elementos do grupo, que me acolheram de forma tão simpática e bem-disposta, e me ensinaram o quão mais fácil é resolver os problemas quando estão várias cabeças a pensar ao mesmo tempo. Um grande obrigado à Maria, que me fez uma enorme companhia durante este ano, à Cláudia, que me acompanhou durante os meus primeiros tempos no laboratório, à Márcia, que esteve sempre completamente disponível para me ajudar e aconselhar, principalmente nestes últimos meses, à Patrícia, que teve sempre a preocupação de saber como estava a correr o meu trabalho e dar-me recomendações, e também ao Paulo, à Joana e à Ana por estarem sempre disponíveis para me ajudar no que fosse preciso.

Aproveito para agradecer ainda a todas as pessoas do grupo do Metabolismo do mRNA, que foram sempre muito prestáveis na cedência de tudo o que eu precisasse.

Temos de reconhecer que o trabalho é muito importante, mas o que seria a vida sem os amigos? Agradeço às minhas amigas Andreia, Beatriz, Joana, Mafalda, Rita, Sofia D. e Sofia M. pela companhia que me fizeram desde a licenciatura até agora, por terem acreditado nas minhas capacidades e me terem encorajado a atingir os meus objetivos, sempre com uma dose de humor à mistura. Agradeço também às minhas amigas Gabriela, Inês, Ivana e Mariana pelas boas memórias que me deram ao longo deste mestrado, por estarem sempre presentes, por serem tão simpáticas e divertidas.

Finalmente, não podia deixar de agradecer às pessoas mais importantes da minha vida, a minha família. Um grande obrigado aos meus pais por me terem proporcionado esta oportunidade e me terem apoiado sempre em todas as decisões que tomei, e à minha irmã, a pessoa com quem sei que posso sempre contar e que torna os meus dias mais divertidos.

This work was supported by Fundação para a Ciência e a Tecnologia (FCT), Portugal, through Grant PTDC/BIA-MOL/28386/2017 to Vânia Gonçalves.

Resumo

Em todo o mundo, são diagnosticados anualmente 1.8 milhões de casos de cancro colorretal, uma doença consideravelmente heterogénea e em muitos casos fatal. Este tipo de cancro divide-se em vários subgrupos sendo que um deles se caracteriza pela instabilidade de microssatélites, localização do tumor no cólon proximal, pólipos com morfologia serreada, metilação frequente de ilhas CpG, mutações no gene *BRAF* e sobreexpressão de RAC1B. A expressão de RAC1B resulta de um processo de *splicing* alternativo do gene *RAC1*, uma pequena GTPase com um papel importante no controlo de várias vias de sinalização celular. A existência de vários locais de *splicing* nos pré-mRNAs permite a existência de processos de *splicing* alternativo para além do constitutivo, ou seja, a partir do mesmo mRNA percursor, diferentes exões ou intrões podem ser incluídos ou excluídos do mRNA final, levando, por exemplo, à produção de diferentes isoformas de determinada proteína. Este mecanismo envolve as sequências consenso que são reconhecidas pela maquinaria de *splicing*, mas também a ação de proteínas *trans* que interagem com elementos *cis* presentes no RNA, atuando como repressoras ou ativadoras para os diversos tipos de *splicing* alternativo.

A variante RAC1B resulta da inclusão de um exão alternativo no mRNA maduro do gene *RAC1*. Este denomina-se exão 3b, possui 57 nucleótidos que se traduzem em 19 aminoácidos adicionais na proteína resultante, e confere a esta variante diferentes propriedades relativamente a sinalização celular.

A regulação da proteína RAC1 é feita através de vários fatores que modulam o seu estado de ativação pela ligação a GTP ou GDP e que, consequentemente, afetam a sua localização e atividade. Sabe-se que a regulação de RAC1B é semelhante, mas esta isoforma tem uma menor afinidade para GDP, menor capacidade de hidrolisar GTP e incapacidade para interagir com inibidores de dissociação de GDP, o que resulta numa predominância desta proteína nas células na sua forma ativa.

Apesar de o processo de *splicing* alternativo ser comum a todas as células, a sua desregulação está muitas vezes associada ao desenvolvimento de doenças. A sobreexpressão de RAC1B, para além de associada a este subtipo de cancro colorretal onde promove a sobrevivência das células tumorais e a sua progressão no ciclo celular, foi também identificada em cancro pancreático, da mama, do pulmão e da tiroide. Esta proteína poderá então ser um alvo terapêutico para o tratamento de pacientes com algumas formas de cancro, o que enaltece a importância do estudo da sua regulação nas células humanas.

Tendo em conta que os processos de *splicing* são modulados por diferentes proteínas, é importante identificar e descrever aquelas que influenciam a expressão de RAC1B neste contexto. Os fatores SRSF1 e SRSF3 foram previamente descritos na literatura como moduladores do *splicing* de RAC1B em cancro colorretal, uma vez que SRSF1 promove a inclusão do exão 3b e SRSF3 promove a sua exclusão. Foram também identificadas regiões no pré-mRNA do gene *RAC1* que são necessárias à ligação destes fatores, o que permitiu conhecer melhor os mecanismos de regulação deste processo.

Mais recentemente, várias outras proteínas têm sido descritas como sobreexpressas em diferentes tipos de cancro, bem como capazes de influenciar a expressão de RAC1B em vários tipos de células. Para este trabalho as proteínas RANBP2, PTBP1, ESRP1, DIS3L2 e hnRNP U foram selecionadas como objeto de estudo, sendo o primeiro objetivo deste trabalho a descrição da sua influência na inclusão ou exclusão do exão 3b de RAC1 em células colorretais.

Com esta finalidade, procedeu-se a experiências de sobreexpressão destas proteínas, mas também à depleção da sua expressão endógena em duas linhas celulares colorretais, uma não tumorigénica (NCM460) e uma maligna com sobreexpressão de RAC1B (HT29). Os efeitos das referidas experiências foram verificados através de RT-PCR e *Western blot*, quantificando-se a expressão de RAC1B relativamente à expressão de RAC1 em cada uma das condições.

A depleção da nucleoporina RANBP2 levou a um aumento da expressão de RAC1B relativamente a RAC1 em células NCM460, tanto ao nível do RNA como da proteína, tendo-se verificado também este aumento em células HT29, mas apenas em relação a expressão proteica. Estes

resultados indicam que a RANBP2 provavelmente atua como inibidora da inclusão do exão 3b em RAC1. Contudo, não foi possível confirmar estas observações através das experiências de sobreexpressão, visto que esta não teve efeitos significativos na expressão de RAC1B, excetuando em NCM460 onde levou a um aumento ao nível da proteína, o que contradiz os resultados acima descritos.

A sobreexpressão dos quatro fatores de splicing em estudo não permitiu retirar conclusões acerca do seu efeito na inclusão ou exclusão do exão 3b no mRNA maduro do gene *RAC1*, contudo as experiências de depleção tiveram maior sucesso. Relativamente a PTBP1, a sua depleção provocou uma diminuição da expressão de RAC1B relativamente a RAC1 em ambas as linhas celulares utilizadas, detetável tanto por RT-PCR como por *Western blot*, indicando que esta proteína atua como promotora da inclusão do exão 3b de RAC1. Para o fator de splicing ESRP1, após depleção em células NCM460 houve também diminuição dos níveis de RAC1B relativamente a RAC1 ao nível do transcrito. Tanto ao nível da proteína em NCM460, como nas mesmas experiências em HT29, os resultados obtidos mostraram seguir a tendência de diminuição da inclusão do exão 3b, contudo, não foram considerados estatisticamente significativos. Concluiu-se então que este fator possivelmente também atua como promotor de RAC1B. Em relação a DIS3L2, em NCM460, apesar de os resultados não serem estatisticamente significativos, verifica-se uma tendência para a diminuição da expressão de RAC1B, enquanto que em HT29 há um aumento significativo da expressão desta variante ao nível da proteína, sendo que ao nível do transcrito, apesar de haver uma variação menor, se confirma esta tendência. Neste caso, a DIS3L2 parece promover a inclusão do exão alternativo 3b em células NCM460, mas inibi-lo em células HT29. Finalmente, a depleção de hnRNP U levou a uma diminuição da expressão de RAC1B relativamente a RAC1 em ambas as linhas celulares, não sendo estatisticamente significativo apenas o resultado ao nível da proteína em NCM460, indicando que hnRNP U atua como promotor da expressão de RAC1B.

O segundo objetivo deste trabalho consistia na identificação de locais de ligação dos fatores PTBP1 e ESRP1 ao pré-mRNA do gene *RAC1*. Para isto, procedeu-se à mutagénese de dois possíveis locais de ligação da proteína ESRP1, previamente descritos na literatura, num minigene contendo a região entre os exões 3 e 4 do gene em questão. A mutação dos locais pretendidos no intrão 3b do minigene foi confirmada através de sequenciação do DNA, mas não foi possível proceder à subclonagem dos fragmentos mutados no minigene original, de modo a prevenir a introdução de mutações indesejadas. No futuro deve proceder-se da mesma forma na introdução de mutações em possíveis locais de ligação do fator PTBP1, e poderá sequenciar-se os plasmídeos mutados, como estratégia alternativa à subclonagem, sendo que cada minigene mutado pode depois ser transfetado juntamente com o fator em estudo, ESRP1 ou PTBP1, de modo a investigar se a expressão de RAC1B nas células se altera em relação à transfeção de um minigene não mutado juntamente com o mesmo fator.

Concluindo, com este trabalho foi possível demonstrar que as proteínas RANBP2, PTBP1, ESRP1, DIS3L2 e hnRNP U estão envolvidas na regulação do *splicing* alternativo de RAC1 em células colorretais. Resultados preliminares indicam que RANBP2 inibe a inclusão do exão 3b em RAC1, enquanto que PTBP1, ESRP1 e hnRNP U promovem a sua inclusão. Relativamente ao fator de splicing DIS3L2, parece ter diferentes efeitos dependendo da linha celular utilizada, podendo promover ou inibir a expressão de RAC1B. Estudos futuros deverão ser realizados, de modo a consolidar estes resultados, esclarecer as vias moleculares através das quais a nucleoporina RANBP2 afeta a expressão de RAC1B e definir locais de ligação dos fatores de splicing em estudo, no pré-mRNA do gene *RAC1*. Estes conhecimentos poderão ser úteis para o desenvolvimento de novas terapias para pacientes com este subtipo de cancro colorretal, que inibam especificamente as vias que afetam a inclusão do exão 3b, de modo a restaurar a normalidade da sinalização celular.

Palavras-chave: Cancro colorretal; RAC1B; *Splicing* alternativo; Fatores de *splicing*; Sinalização celular

Abstract

Colorectal cancer (CRC) is a common malignancy with significant mortality, which presents different subtypes due to genetic heterogeneity. The focus of this work was on a subtype that is characterized by microsatellite instability and mutations in the *BRAF* gene, being associated, in most cases, with RAC1B overexpression. RAC1B is a splicing variant of the *RAC1* gene and is characterized by the inclusion of an additional exon through alternative splicing, designated as exon 3b, presenting some regulation and signalling differences when compared to RAC1. RAC1B has been described as a promoter of tumour progression, therefore, it is considered as a potential therapeutic target, as it may allow the design of targeted therapies for patients with this subtype of CRC, but for this, understanding in more detail the mechanisms of RAC1B regulation is extremely important.

Previous research projects in the host lab had already identified SRSF1 and SRSF3 as two splicing factors involved in the regulation of RAC1B splicing in CRC, so the first goal of this thesis work was to determine the effect of other selected proteins on the expression of RAC1B in colorectal cell lines. The proteins of interest are a nucleoporin, RANBP2, and four splicing factors, PTBP1, ESRP1, DIS3L2 and hnRNP U, which were previously described as upregulated in CRC or involved in the RAC1B splicing process in other cell lines. The effect of these proteins on RAC1B alternative splicing was assessed through their overexpression and depletion in two colorectal cell lines, NCM460 and HT29, and the analysis of RAC1 exon 3b inclusion was made at the transcript level by RT-PCR, and at the protein level through SDS-PAGE and Western blot.

The obtained results provided strong evidence that, in fact, these proteins influence RAC1B expression, both at the transcript and protein level. Although the overexpression experiments performed in this work did not lead to many clear conclusions about the effect of the proteins of interest on the RAC1B expression, the siRNA assays were rather conclusive. Taking into account the results obtained through depletion experiments, RANBP2 seems to act as a RAC1B inhibitor in NCM460 cells, while PTBP1, ESRP1, DIS3L2 and hnRNP U are probably enhancers of that RAC1 isoform. In the HT29 cell line, RANBP2 and DIS3L2 were pointed as inhibitors of exon 3b inclusion, while PTBP1, ESRP1 and hnRNP U are most likely enhancers of that alternative splicing event.

The second goal of this work was to identify PTBP1 and ESRP1 RNA-binding motifs in the RAC1 transcript sequence. Here, two predicted ESRP1 binding motifs in the RAC1 transcript were identified and mutated in a *RAC1* minigene, in order to verify if any of these changes could avoid ESRP1 binding and consequently affect the splicing process. The presence of the mutations was confirmed through DNA sequencing, however, it was not possible in the remaining time to reclone the mutant fragments into the original minigene vector, to assure they did not have undesired mutations, allowing to proceed to the planned experiments.

Future studies should be performed in order to confirm the obtained results, as not all of them resulted in statistically significant variations on RAC1B expression levels, and it is also imperative to clarify how the nucleoporin RANBP2 affects RAC1B expression and what binding sites are recognized by the studied splicing factors in the RAC1 pre-mRNA.

Keywords: Colorectal cancer; RAC1B; Alternative splicing; Splicing factors; Cell signaling

Table of contents

List of figures

List of tables

List of abbreviations

ANOVA	Analysis of variance
BRAF	Type-B rapidly accelerated fibrosarcoma
CIMP	CpG island methylator phenotype
CRC	Colorectal cancer
CSD	Cold-shock domain
ddNTP	Dideoxynucleotide triphosphate
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
DIS3L2	DIS3 like 3'-5' exoribonuclease 2
DTT	Dithiothreitol
ECL	Enhanced chemiluminescence
EMT	Epithelial-mesenchymal transition
ESE	Exonic splicing enhancer
ESRP1	Epithelial splicing regulatory protein 1
ESS	Exonic splicing silencer
EtBr	Ethidium bromide
FBS	Fetal bovine serum
FGFR	Fibroblast growth factor receptor
GAP	GTPase-activating protein
GDI	GDP dissociation inhibitor
GDP	Guanosine diphosphate
GEF	Guanine nucleotide exchange factor
GFP	Green fluorescent protein
GTP	Guanosine triphosphate
HCC	Hepatocellular carcinoma
hnRNP	Heterogeneous nuclear ribonucleoprotein
HRP	Horseradish peroxidase
IRES	Internal ribosome entry site
ISE	Intronic splicing enhancer
ISS	Intronic splicing silencer
JNK	c-Jun N-terminal kinase
KRAS	Kirsten rat sarcoma
LB	Lysogeny broth
MAPK	Mitogen-activated protein kinase
MM	Molecular marker
MSI	Microsatellite instability
Myc	Myelocytomatosis oncogene
NADPH	Nicotinamide adenine dinucleotide phosphate
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells

NPC	Nuclear pore complex
p53	Tumour protein 53
PAK	P21 activated kinase
PCR	Polymerase chain reaction
PKM2	Pyruvate kinase M2
PMSF	Phenylmethylsulphonyl fluoride
PTBP1	Polypyrimidine tract binding protein 1
PVDF	Polyvinylidene difluoride
RAC1	Ras-related C3 botulinum toxin substrate 1
RAN	Ras-related nuclear protein
RANBP2	RAN binding protein 2
RANGAP1	RAN GTPase activating protein 1
Ras	Rat sarcoma
Rho	Ras homolog
RNA	Ribonucleic acid
RNB	Ribonuclease II
ROS	Reactive oxygen species
RT-PCR	Reverse transcriptase polymerase chain reaction
SAF-A	Scaffold attachment factor A
SAPK	Stress-activated protein kinase
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
snRNP	Small nuclear ribonucleoprotein
S.O.C.	Super optimal broth with Catabolite repression
SRSF1	Serine/arginine-rich splicing factor 1
SRSF3	Serine/arginine-rich splicing factor 3
STR	Short tandem repeat
SUMO	Small ubiquitin-related modifier
TBE	Tris-Borate-EDTA
TBST	Tris-buffered saline triton
Tm	Melting temperature
TNFα	Tumour necrosis factor α
Topo II	Topoisomerase II
Tub	Tubulin
UV	Ultraviolet radiation
Wnt	Wingless/int-1

1. Introduction

1.1. Sporadic colorectal cancer

With 1.8 million newly diagnosed cases and 900 000 deaths annually, colorectal cancer (CRC) is currently the third most common and second deadliest cancer worldwide, being also a highly heterogeneous malignancy concerning anatomical location, pathological criteria and molecular changes.[1–4] CRC is usually initiated by some genetic alteration that predisposes cells to tumoural development, and this may be followed by three other steps: promotion, when there is a high cell proliferation rate; progression, when other malignant genetic changes are acquired; and, finally, metastasis, when cancer cells are disseminated from the primary tumour to other parts of the organism, generating secondary tumours.[1]

CRC is mostly sporadic (60-65%), depending on environmental and nutritional factors rather than hereditary mutations.[1–3] Around 25% of CRC cases are associated with a family history of the disease, but only 5% of the cases constitute hereditary cancer syndromes.[1] Several sporadic CRC subgroups have been identified and one displays the microsatellite instability (MSI) phenotype, which affects almost 15% of the patients and is characterized by variations in the length of short tandem repeats (STR).[3,5,6] This characteristic is due to a combination of a tendency in repetitive sequences for DNA polymerase slippage and inactivation of the DNA mismatch repair system.[3,5,6] This type of tumours usually occur in the proximal colon and cause polyps with a serrated morphology.[3,5] MSI-high CRC contains alteration in at least 3 of 5 STR markers and is often associated with a CpG island methylator phenotype (CIMP) and mutations in the *BRAF* gene, which encodes a serine/threonine protein kinase involved in activation of the MAPK pathway.[1,3,5–7] It should be noted that KRAS activating mutations also activate the MAPK pathway and occur in about 30% of colorectal tumours; however, BRAF activating mutations are an alternative oncogenic event, thus, these mutations allow the identification of molecularly distinct tumours (Figure 1.1).[3,6] The most frequent BRAF mutation, $BRAF^{V600E}$, consists in the replacement of a valine by a glutamic acid in exon 15 mimetizing an amino acid phosphorylation that increases BRAF activity, and was identified as responsible for the overstimulation of the MAPK pathway.[1,5–9]

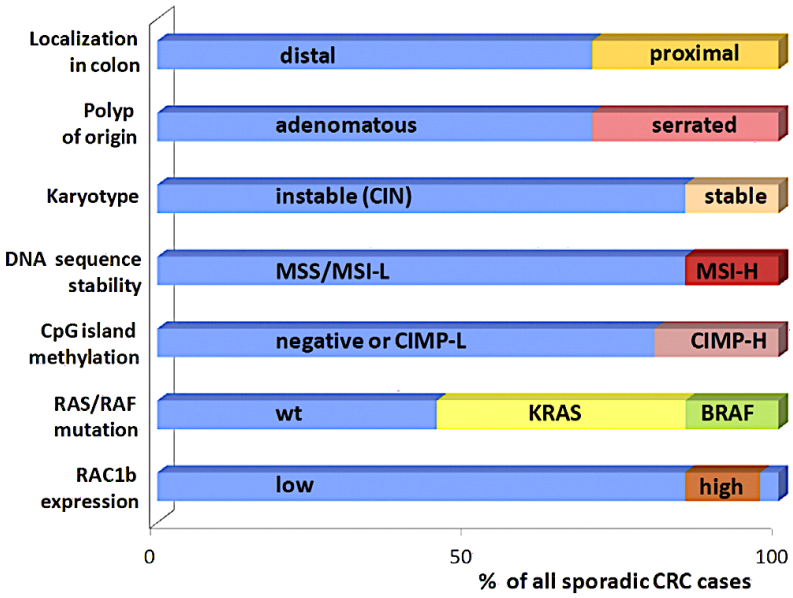


Figure 1.1 - Main molecular features characterizing the serrated pathway in sporadic CRC. Blue bars represent all sporadic CRCs (100%) and show in colour the percentage of CRCs presenting the indicated molecular or anatomical features which are shared with the serrated pathway.[3] (From Matos *et al.* 2016)

About 80% of the BRAF^{V600E} CRC cases also show RAC1B overexpression, which is an alternative splicing variant of the small GTPase RAC1 that will be further discussed in this work.[5,6,10]

1.2. Constitutive and alternative splicing mechanisms

The process of splicing may be defined as an essential step for gene expression by which introns, noncoding regions, are removed from pre-mRNA transcripts and exons, coding regions, are joined in order to form a mature mRNA product.[11–14]

Splicing requires a macromolecular ribonucleoprotein complex, the spliceosome, that is responsible for the recognition of consensus sequences on the pre-mRNA, namely the 5' splice site, 3' splice site, branch-point site and polypyrimidine tract, in order to catalyse RNA excision and join the desired fragments in the same order as they appear in the gene.[11–13,15,16]

The spliceosome is composed of five small nuclear ribonucleoprotein particles (snRNPs), named U1, U2, U4, U5 and U6, besides many other associated proteins.[11,15–18] These snRNPs are assembled on the pre-mRNA in a defined order, forming a series of complexes designated as H, E, A, B and C (Figure 1.2).[11–13]

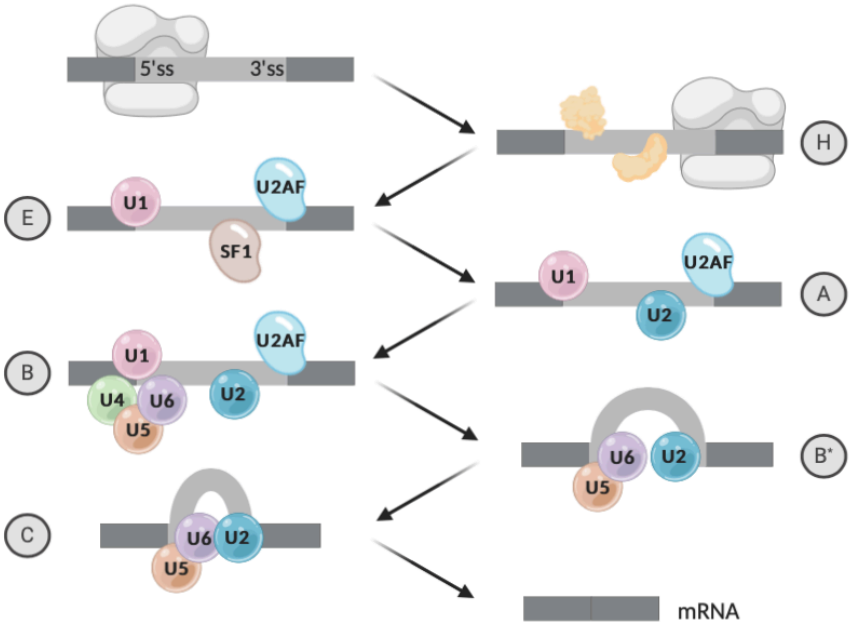


Figure 1.2 – Schematic representation of spliceosome assembly. The spliceosome begins to be assembled co-transcriptionally, passing through defined sequential steps, designated as complexes H, E, A, B and C, as indicated in the figure. Exons are represented in dark grey while introns present a lighter colour, DNA polymerase is represented in white and hnRNPs are represented in yellow. The activated B complex is represented as B*. The result of the splicing process is the mature messenger RNA (illustrated at the bottom). Created with BioRender.com. (Adapted from Lee and Rio 2015)[12]

The H complex consists in the binding of different heterogeneous nuclear ribonucleoproteins (hnRNPs) on the pre-mRNA, in a co-transcriptional manner, influencing the spliceosomal assembly on different splice sites.[11,19] Subsequently, the E complex is formed, as the 5' splice site is recognized by the snRNP U1, SF1 binds to the branch point and U2AF interacts with the polypyrimidine tract and the 3' splice site.[11,13,16] This complex indicates a commitment to splicing, bridging the exons or introns (depending on the considered model, of intron or exon-definition) and approximating the splice sites.[11,16] Then, U2 replaces SF1 at the branch point, leading to the formation of complex A, which defines a commitment of the splice site pairings.[11,16] As a tri-snRNP complex composed by U4, U5 and U6 interacts with the 5' splice site, the pre-catalytic B complex is formed, being activated after a few rearrangements.[12,16] During catalytic activation of the B complex, U1 and U4 are destabilized and released from the RNA, additionally, as U6 loses its base-pairing with U4, it forms an internal stem loop

and becomes paired to U2, creating the active site of the spliceosome.[12,16] The catalysis of the first splicing step is made by the activated B complex, and only after a few RNA conformational changes, the catalytically competent C complex is formed.[11,15,16] This allows the finalization of the splicing reaction, when there are two trans-esterification reactions which lead to lariat formation, intron removal and, finally, exon ligation.[11,15,16] For its catalytic activity and RNA composition, the spliceosome is classified as a ribozyme.[12] The spliceosomal assembly and disassembly process is rather dynamic, and several of its steps are reversible.[12,13]

Besides constitutive splicing, there are also alternative splicing events on nearly all human genes (>95-100%), that are possible due to the existence of multiple splicing sites in the precursor mRNAs.[12,18] Alternative splicing is a mechanism by which particular exons or introns may be differentially included or excluded from the final produced mRNA, as different splice sites can be joined.[11] Splicing sites may be differentially used based on a variety of mechanisms: (1) variations in the sequence degeneration level on the 5' and 3' splice sites, (2) auxiliary cis-acting splicing elements that recruit different regulatory factors, (3) phosphorylation state and subcellular distribution of regulatory splicing factors, (4) the length of the flanking introns, (5) existence of pre-mRNA secondary structure, or (6) nucleosome positioning, chromatin structure and histone modifications.[11,16]

Alternative splicing is an essential post-transcriptional process that has several regulatory functions. It can regulate protein expression and allow the production of numerous protein isoforms from the same pre-mRNA molecule, generating a more diverse proteome from a given gene number within the genome.[12,15,17] It can also lead to the introduction of premature stop codons on the transcripts causing their subsequent degradation.[12,15,17] Even in untranslated regions, the occurrence of alternative splicing processes may influence mRNA localization, mRNA stability or translation efficiency.[20,21]

Although essential, splice site consensus sequences are insufficient for exon and intron definition, and trans-acting proteins, which may be repressors or activators, help splice site selection through interaction with cis-acting elements.[11–13,16,20,21] These elements are short sequences present on the pre-mRNA, usually with 4 to 18 nucleotides, and may be exonic splicing enhancers (ESE), intronic splicing enhancers (ISE), exonic splicing silencers (ESS) or intronic splicing silencers (ISS), depending on their function and position.[11–13,16,20,21] Splicing enhancers (ESEs and ISEs) bind positive trans-acting factors, such as serine and arginine-rich (SR) proteins that generally promote spliceosome assembly on weak splice sites via their RS domain, preventing exon skipping and promoting protein interactions within the spliceosome throughout exons and introns or between enhancers and nearby splice sites.[13,15,16,18] On the other hand, splicing silencers (ESSs and ISSs) bind negative trans-acting factors, such as hnRNPs that usually inhibit spliceosome assembly at those sites by sterical hindrance or formation of inhibiting RNA secondary structures.[12,13,15,16] Both hnRNPs and SR proteins are key factors in exon-intron definition, and their balance defines the location of spliceosome assembly, being suggested that enhancing elements have a greater importance in constitutive splicing, while silencers are more relevant in alternative splicing.[12,13]

These regulatory proteins, which interact with cis-acting elements, are able to recognize single-stranded RNA, thus, the formation of secondary structures prevents their binding and the recruitment of specific trans-acting partners.[16] Also, variations in local concentrations of these factors, in their ratios or their activity, allow the regulation of splice-site recognition, both in constitutive and alternative splicing.[11,13,16,17] Additionally, there are tissue-specific splicing patterns, determined by specific combinations of regulators that are present in particular cell types, and also alternative splicing networks related to different physiologic states of cells, such as stress response or disease.[11,15]

There are different types of alternative splicing events as illustrated in Figure 1.3. Major types of alternative splicing are: exon skipping or cassette exons, where one or more exons and respective flanking introns are included or excluded from the final mRNA; alternative 5' or 3' splicing sites, when there are competing splice site options to define the exon boundaries; differential intron

retention, where an intron is not removed from the mature mRNA as it would be expected; mutually exclusive exons, a rare event in which two consecutive exons cannot be simultaneously included in the final mRNA; alternative transcription start sites, when there is an alternative first exon which may be included in the mature mRNA transcript; and multiple polyadenylation sites, in which, similarly to the previous example, there is an alternative last exon that may be used.[13,15,17,20] Cassette exon is the most frequent type of alternative splicing in higher eukaryotes, possibly being related to the generation of greater phenotypic complexity.[13,20] It was also described in the literature that alternative cassette exons and respective flanking introns have higher sequence conservation levels when compared to constitutive exons and their flanking introns, although they are shorter and have weaker splice sites.[20] It is thought that these characteristic features are due to the need of a proper regulation of alternative splicing, namely through the exonic splicing regulatory sequences, and a correct placement of the spliceosome at the exon-intron junctions.[20]

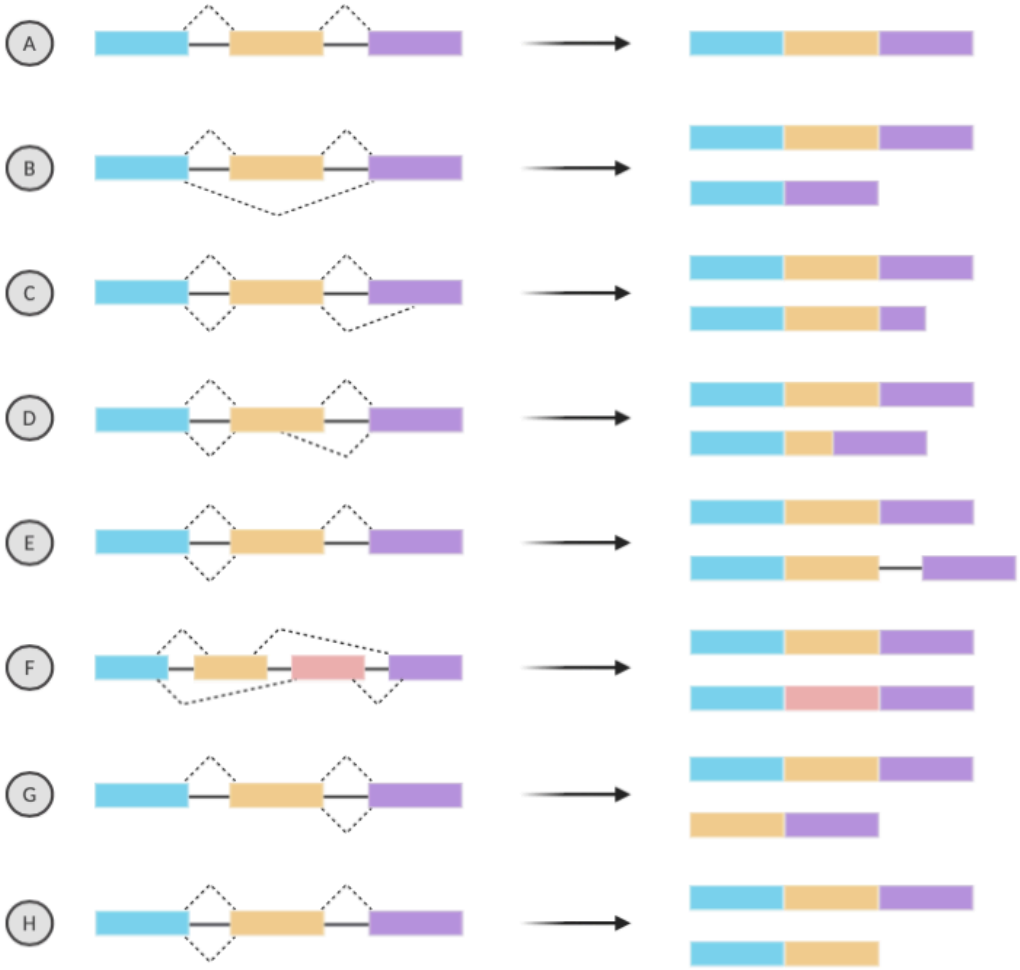


Figure 1.3 – Different types of alternative splicing events. Schematic representation of splicing processes with exons represented in colour and interspersed by introns. (A) Constitutive splicing; (B) Exon skipping or cassette exon; (C) Alternative 3' splice site; (D) Alternative 5' splice site; (E) Intron retention; (F) Mutually exclusive exons; (G) Alternative transcription start site; (H) Alternative terminal exons with polyadenylation sites. Created with BioRender.com. (Adapted from Di *et al.* 2019)[17]

Although alternative splicing is a normal and frequent process in cells, its deregulation is associated with several diseases, such as cancer. Splicing was identified as being involved in almost all steps of cancer development, therefore, some alternative splicing products and splicing factors have become biomarkers or therapeutic targets.[13,16,18,22] It was previously documented that some splicing variants may contribute to cancer genesis and progression, acting in proliferation, metastasis or

apoptosis and consequently redirecting the fate of cancer cells.[17,23] In addition, genomic mutations occurring in canonical splice sites can lead to the activation of cryptic sites usually not recognized by the spliceosome, leading to aberrant splicing products and production of neoantigens that can be targeted for immunotherapy.[15]

1.3. The small GTPase RAC1 and its splicing variant RAC1B

RAC1 is a small GTPase that belongs to the Rho family within the Ras superfamily of proteins and is known to have a role in the control of signalling pathways involved in cytoskeletal dynamics and gene transcription.[10,24,25] It has lipidic modifications which target it to cell membranes and can be found in an active or inactive state when bound to GTP or GDP respectively.[10,24,25] The transitions between both states are regulated in the cell by guanine nucleotide exchange factors (GEFs), GTPase-activating proteins (GAPs) and GDP dissociation inhibitors (GDIs) (Figure 1.4).[10,24,25] GEFs promote RAC1 activation by exchanging GDP for GTP, leading to a conformational change, GAPs promote RAC1 inactivation by accelerating its intrinsic GTPase activity and GDIs sustain RAC1 inactivation by sequestering it in the cytoplasm.[10,24,25] These changes on the activation state are initiated by growth factor receptor activation, integrin activity or other effector molecules that modulate the activity or localization of the referred RAC1 regulators, leading to cell behaviour changes.[25]

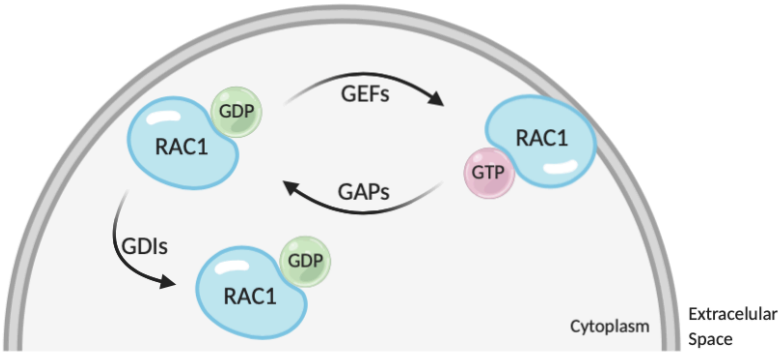


Figure 1.4 – Schematic representation of RAC1 activity regulation. RAC1 cycles between active and inactive states, characterized by the binding to GTP or GDP respectively, through the action of GEFs, GAPs and GDIs. Activated RAC1 is located at the cell membrane.[10,24,25] Created with BioRender.com.

Upon GTP binding, RAC1 undergoes a conformational switch that allows interaction with and consequent activation of many effector molecules, which can then trigger signalling cascades, stimulating for instance actin polymerization, laminin deposition, lamellipodia formation, DNA synthesis, the JNK/SAPK protein kinase cascade, or the activity of NF-κB that regulates cell cycle progression.[10,24–26] It was also described that if RAC1 activity is not properly regulated it may have a tumorigenic influence, namely being involved in epithelial-mesenchymal transition, cell invasion or survival.[24,25,27]

RAC1B was identified as an alternative splicing variant of the *RAC1* gene, having 19 more amino acids in-frame between codons 75 and 76, corresponding to an extra exon with 57 nucleotides, named exon 3b (Figure 1.5).[10,24,28]

This isoform shows some differences in terms of regulation and signalling when compared to RAC1. Although RAC1B relies on GEFs and GAPs for its activation and downregulation just as RAC1, an increase in its intrinsic exchange rate of GDP to GTP, due to reduced GDP affinity, and decreased GTP hydrolysis were also documented.[24,28] In cells, a main property described for RAC1B was an inability to interact with GDIs, due to the 19 additional amino acids, remaining bound to the

plasma membrane instead of being sequestered in the cytoplasm, therefore, there is a predominance of this isoform in an active state in the cells, as it can continuously contact membrane localized GEFs.[24,28,29]

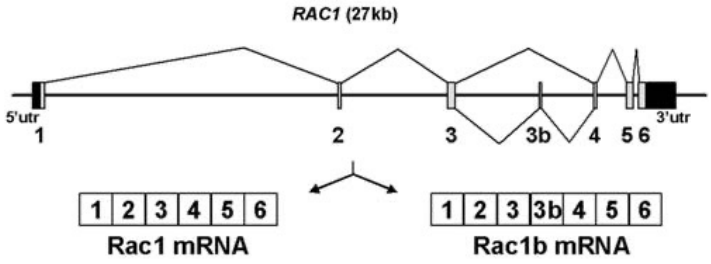


Figure 1.5 – Schematic representation of the human *RAC1* gene. The *RAC1* gene has seven coding exons, including the alternative exon 3b that may be excluded (left side) or included (right side) in the final mRNA.[14] (From Gonçalves *et al.* 2009)

In terms of downstream signalling, RAC1B cannot induce the cytoskeletal changes described for RAC1, due to its inability to activate PAK and stimulate the JNK pathway.[6,24,29] However, RAC1B is able to stimulate the NF-κB pathway and to promote the production of reactive oxygen species (ROS), as it forms a complex with NADPH oxidase at the membrane.[6,24,28] In mammary epithelial cells, RAC1B was found to activate epithelial-mesenchymal transition (EMT) and cause DNA damage via ROS production, with a consequent increase in genomic instability.[6,24,28,30] RAC1B is also able to activate the Wnt pathway, creating a negative feedback loop that leads to its own degradation through ubiquitination, and decreases cell adhesion, being this an oncogenic pathway in many colorectal tumours.[14,28]

RAC1B overexpression has been associated with colorectal, breast, lung, thyroid and pancreatic cancers.[10,31–34] In colon tumour cell lines, RAC1B is found in lower protein levels than RAC1, nevertheless, when analysing the activation state of both proteins it is noticeable that the levels of active RAC1B are greatly superior to the levels of active RAC1, due to the reduced intrinsic GTPase activity and its inability to be inhibited by GDIs.[10,24,28] Given the observed association of $BRAF^{V600E}$ CRC with RAC1B, several studies allowed the conclusion that RAC1B synergizes with the activated oncogene *BRAF* in order to promote cell survival and cell cycle progression.[6,28,35,36]

1.4. Splicing-related proteins and their influence on RAC1B expression

Alternative splicing events may be modulated by splicing factors, which in turn are regulated through different mechanisms, namely expression level variations and/or post-translational modifications.[14]

As previously referred, RNA-binding proteins may be divided in SR proteins and hnRNPs, and the expression of some was found to be deregulated in tumours.[35] It was also described that SR proteins and hnRNPs usually have opposing effects on splicing.[16,37] They recognize multiple splicing regulatory sequence elements and these elements are also recognized by several different splicing factors.[35]

SRSF1 (previously known as ASF/SF2) and SRSF3 (previously known as SRp20) constitute examples of antagonistic SR proteins that are known to regulate RAC1B splicing.[14,18] SRSF1 was found to be overexpressed in colon tumours among others, being involved in multiple alternative splicing mechanisms; however, it was identified as an intervenient in cancer progression rather than tumour initiation.[35] The tumorigenic role of SRSF3 is also well documented, for example as a direct regulator of the tumour suppressor protein p53.[18] In the context of a CRC model studied in the host lab, the overexpression of SRSF1 leads to an increased expression of RAC1B, being classified as an enhancer for the alternative exon 3b inclusion, whereas SRSF3 depletion has the same effect, being identified as a silencer.[14,35] Through determination of exon 3b inclusion levels following mutation of conserved

sequence motifs, the exon 3b positions +7 and +11 were recognised as necessary for SRSF1 binding, while the positions +18 and +19 are required for SRSF3 binding.[14]

More recently, other proteins were identified as influencing RAC1B expression, being some of them described below and further studied in this work.

1.4.1. RANBP2

RAN binding protein 2 (RANBP2) is a large pleiotropic protein, with a structure that includes four RAN-binding domains, eight tandem zinc fingers, a SUMO E3 ligase domain and repetitions of phenylalanine and glycine for the binding of transport receptors.[2,38–40] As the name suggests, RANBP2 binds to RAN, a small GTP binding protein from the RAS superfamily, with an asymmetrical distribution in the cells, existing higher concentrations of RAN-GTP in the nucleus and RAN-GDP in the cytoplasm.[38,39,41] RAN-GTP facilitates the release of cargo molecules imported to the nucleus but also binds to exportins, being transported to the cytoplasm, where it interacts with RANGAP1, leading to RAN-GTP conversion to the GDP-bound form and consequent release of the transported macromolecules.[41] RANGAP1 is located on the nuclear envelope, in the cytoplasmic side of the nuclear pore complex (NPC) through interaction with RANBP2, which facilitates its function.[41] The NPCs are large protein channels present in the nuclear membrane with the function of macromolecular traffic control between the cytoplasm and the nucleus, and RANBP2 is one of its constituent nucleoporins, being a cytosolic component of the filaments attached to the NPC cytoplasmic ring.[2,39] Because of its relationship with RAN and its structural characteristics, RANBP2 was firstly described in the literature as a nucleocytoplasmic transport regulator, however, many other of its functions have now been documented, associating this protein with mitosis and cancer.[38,39]

Concerning nucleocytoplasmic transport, RANBP2 is important in the formation of nuclear speckles, which are clusters of proteins, mainly belonging to the SR family and showing different levels of phosphorylation, that are adjacent to some chromosomes and have functions related to gene expression, as pre-mRNA splicing or mRNA export.[41] This suggests that RANBP2 has an active role in the distribution of phosphorylated SR proteins in the nucleus, thus affecting alternative splicing events.[41]

Regarding mitosis, it was shown that RANBP2 is necessary for the assembly and proper function of kinetochores, mitotic spindles and centrosomes, ensuring proper chromosome segregation and genome integrity.[2] The reason why RANBP2 is so important in this process, is related to its enzymatic function, a SUMO E3-type ligase activity.[40,42] The E3 ligase domain binds to the E2 enzyme Ubc9 and also interacts with SUMO-modified forms of RANGAP1, creating a complex that makes part of the kinetochore during mitosis.[40,42] Moreover, this domain is involved in topoisomerase (Topo) II sumolyation, which is a necessary process for the correct localization of Topo II in the centromeres, allowing its function in catalysing the separation of catenated chromosomes after their replication.[40] Hence, it can be assumed that RANBP2 is a tumour suppressor, as it prevents the formation of multipolar spindles, aneuploidies or other chromosomal abnormalities.[2,38,40] Furthermore, mice with low levels of this protein showed a tendency to develop tumours.[2]

Of particular interest for this thesis are some studies that attribute an oncogenic role to RANBP2 in $BRAF^{V600E}$ mutated colon cancer cells.[2,43] It was described that this type of cells present a reduction in microtubule outgrowth during spindle formation, but also an upregulation of RANBP2, which is thought to balance the defects caused by the BRAF mutation and allow mitotic progression and survival of the cancer cells, because upon RANBP2 silencing these cells develop defects in kinetochore structure and composition, problems in mitotic progression and apoptosis.[43] Given that RAC1B is also overexpressed in BRAF mutant colon cells, RANBP2 was an interesting candidate protein included in this work.

1.4.2. PTBP1

PTBP1 (also known as hnRNP I) is expressed in almost all human cell types, although in variable levels, and can translocate between the nucleus and the cytoplasm.[44–47] It has four RNA recognition motifs, being usually bound to polypyrimidine-rich RNA sections in order to regulate alternative splicing, acting as an enhancer or a repressor depending on the binding region.[44–47] It may also bind DNA, acting as a transcription factor.[44–47] PTBP1 has many other functions, namely in mRNA location, stability, transport and metabolism, in transcription, polyadenylation, IRES mediated translation, protein stability and localization, regulation of genes involved in apoptosis, cellular growth, proliferation, differentiation, migration and invasion.[44,46,47] Additionally, PTBP1 promotes glycolysis in cancer cells, rather than oxidative phosphorylation, through SRSF3 overexpression and promotes the alternatively spliced glycolytic enzyme variant PKM2.[44] It also stimulates neuronal cell growth and differentiation, and promotes T cell activation.[44]

There is an alternative exon 11 in PTBP1 itself, which has to be included in order to produce the functional protein.[44] This allows an autoregulation process, where PTBP1 causes exon skipping and leads to nonsense-mediated decay of its own mRNA, but it is also regulated by miRNAs, lncRNAs, H_2O_2 and other RNA-binding proteins, such as ESRP1.[44]

This protein is known to be overexpressed in many tumours, often representing a poor overall survival and disease-free survival rate, indicating a correlation between alternative splicing and carcinogenesis.[45–47] As a splicing factor, PTBP1 was identified by a bioinformatics approach as a stimulator of the alternative splicing of isoform RAC1B, which may promote colon cancer formation.[44,45] It also leads to the inclusion of exon 11 in cortactin, which enhances cell migration and invasion in CRC, as it is involved in cytoskeletal dynamics and cell motility.[47] On the other hand, PTBP1 knockdown in HCT116 and SW480 colorectal cells, decreased cell proliferation rate and induced cell cycle arrest and apoptosis.[46]

Interestingly, PTBP1 silencing in HT29 cells has different effects depending on the level of downregulation.[48] With more than a 90% decrease of PTBP1, there was a reduction in cell-cell adhesion, proliferation and colony formation, but an increase in invasiveness; however, a PTBP1 reduction of approximately 60% led to a decrease in invasiveness, being concluded that PTBP1 expression is related to reduced invasiveness.[48]

1.4.3. ESRP1

The ESRP1 belongs to the hnRNP family and influences the expression of several genes, mainly involved in EMT, cell motility and cytoskeletal dynamics.[29,49] It is upregulated during carcinogenesis but downregulated during invasion, namely in EMT.[49] EMT is observed in normal organ development, but also in wound healing with tissue fibrosis and metastatic processes, involving several transcriptional changes.[29,49,50] ESRP1 negatively regulates EMT in colon, breast and pancreatic cancer, suppressing malignant phenotypes.[29,49–51] There are splicing changes during this complex process which impact gene regulation, and it has been determined that ESRP1 downregulation during EMT has a correlation with splicing changes of several exons.[50] For example, in oral squamous cancer cells, ESRP1 knockdown increased RAC1B transcript and protein levels.[49]

ESRP1 has cell- and location-specific functions, for instance, when located in the nucleus it may modulate alternative splicing events, but when in the cytoplasm it may control translation processes.[29,49,51] It presents three motifs for RNA recognition but it is also involved in protein-protein interactions, exerting some influence on intracellular signalling pathways.[51] In the context of CRC, ESRP1 overexpression leads to the activation of the fibroblast growth factor signalling pathway.[51] When the fibroblast growth factor receptors (FGFR) are stimulated, tumour progression and epithelial cell

growth are induced in an anchorage-independent manner.[51] It was also described in CRC cells that ESRP1 interacts with other proteins which promote mitosis and proliferation, acquiring a pro-tumorigenic function.[51]

1.4.4. DIS3L2 and hnRNP U

DIS3L2 is a conserved cytoplasmic 3'-5' exoribonuclease, composed of two cold-shock domains (CSD), an S1 domain which allows its binding to RNA, and a ribonuclease II domain (RNB).[23,52] It is able to identify decay signals in RNA transcripts, in the form of 3' polyuridylation, leading to the degradation of mRNAs, miRNAs or ncRNAs.[52,53] Nevertheless, it was reported that DIS3L2 may degrade RNA in a polyuridylation-independent manner.[52] Furthermore, DIS3L2 is involved in mRNA metabolism, cell proliferation, differentiation and apoptosis, and is associated with several diseases, such as Perlman syndrome, an overgrowth syndrome where this protein has usually loss of function mutations, Wilms' tumour, a kidney cancer with partial or total DIS3L2 deletion, or Marfan-like syndrome, characterized by skeletal overgrowth and a truncated DIS3L2 locus.[52–54]

The role of DIS3L2 in cancer remains controversial, as it seems to be variable, depending on the considered type of tumour.[23,52] On one hand, it may be viewed as an oncogenic factor, as it cooperates with the RNA-binding protein LIN28 to suppress the maturation of LET-7, a tumour suppressing miRNA.[23,52,55] But on the other hand, it may be viewed as a tumour suppressor, as its knockdown was shown to cause an increase in cell proliferation and growth, decreased levels of apoptosis, occurrence of mitotic errors and abnormal expression of several proteins involved in cell cycle regulation, while its overexpression suppressed cell cycle progression and cellular growth.[23,52,53] In CRC tissue samples, a lncRNA AC105461.1 was identified as significantly downregulated and positively correlated with DIS3L2 expression, furthermore, its overexpression in CRC cell lines slowed down cell growth, while its knockdown caused higher levels of cell growth and invasion.[55]

The hnRNP U protein (also known as SAF-A) is the largest of the heterogeneous nuclear ribonucleoproteins and is mainly found in the cell nucleus as part of the nuclear matrix.[37,56–58] It is composed by an N-terminus with mainly acidic residues that allows DNA-binding, a glutamine stretch, a nuclear localization sequence and a C-terminus rich in glycine, including an RGG box which mediates RNA-binding.[56–58] hnRNP U is involved in alternative splicing regulation, most frequently with cassette exon inclusion, as in the case of caspase-9 pre-mRNA, where hnRNP U prevents an exon cassette exclusion, consequently promoting the expression of caspase-9a, a pro-apoptotic isoform.[37,57] It also has a role in RNA stability control, as verified for TNFα which mRNA lifetime was doubled through stabilization by hnRNP U.[58,59] Moreover, it is involved in 3D genome organization during interphase and throughout mitosis, as its knockdown may lead to chromosomal instability.[58,59] Mutations or microdeletions in the gene encoding hnRNP U can lead to brain disorders or cancer development.[58,59] Furthermore, hnRNP U is involved in transcription regulation, RNA transport, cell apoptosis, DNA damage response, telomere length control and the epigenetic inactivation of the X chromosome.[23,56,57,60]

It was recently described in the literature that DIS3L2 expression was upregulated in hepatocellular carcinoma (HCC), gradually increasing along the cancer stages and being associated with survival time.[23] In liver cell lines, DIS3L2 directly interacts with hnRNP U in the nucleus through its CSD, being further demonstrated that DIS3L2 or hnRNP U knockdown led to a decrease in RAC1B expression without affecting overall RAC1 expression levels and, as expected, their overexpression had the opposite outcome.[23] For the regulation of this alternative splicing event DIS3L2 was described to bind to the exon 3b region through its S1 domain, allowing the subsequent hnRNP U binding to the RAC1 pre-mRNA by its RGG domain, which results in exon inclusion (Figure 1.6).[23]

It was further verified that the effect of DIS3L2 and hnRNP U overexpression or knockdown on cell proliferation is reversed by RAC1B knockdown or overexpression respectively, which highlights

the importance of RAC1B in the oncogenic capacity of DIS3L2 and hnRNP U to regulate cell proliferation and tumour growth in HCC.[23]

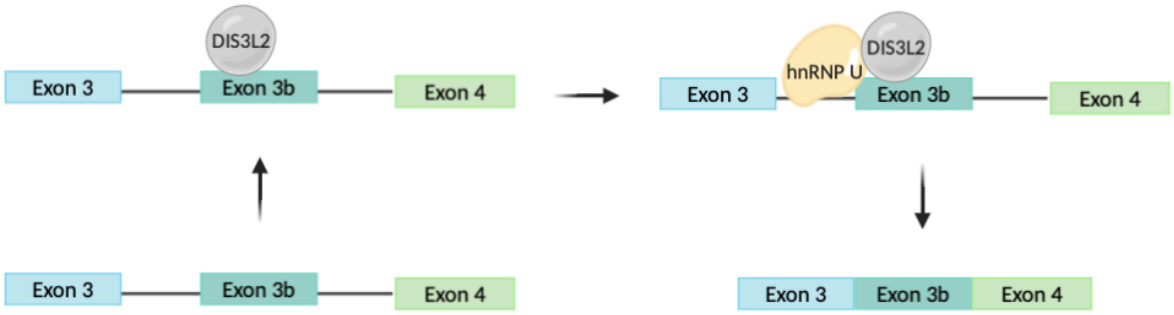


Figure 1.6 – Schematic representation of RAC1B splicing through DIS3L2 and hnRNP U in HCC. Sequential binding of DIS3L2 and hnRNP U to exon 3b leads to exon inclusion into the mature mRNA. Exons are represented in colour and interspersed by introns. Created with BioRender.com. (Adapted from Xing *et al.* 2019)[23]

2. Objectives

A subgroup of CRC cases has been described in the literature as presenting RAC1B overexpression, which influences cell viability and proliferation, leading to the identification of this isoform as a potential therapeutic target.[5] Thus, understanding the mechanisms of RAC1B regulation is extremely important to allow the design of targeted therapies for these patients.

It is also known that alternative splicing may be modulated by different proteins, and some have already been determined for RAC1B modulation: SRSF1 and SRSF3.[14] Additionally, it was described that RANBP2, PTBP1 and ESRP1 are overexpressed in colon tumours, possibly being involved in RAC1B alternative splicing.[43–45,51] In previous pilot experiments, the host lab found that PTBP1 and ESRP1 overexpression led to a decrease in exon 3b inclusion.[61] Contradictorily, ESRP1 depletion also led to reduced levels of RAC1B transcript and protein in several cell lines, demanding for further investigation of this splicing factor.[61] For RANBP2 overexpression, a decrease in exon 3b inclusion was shown, while its depletion had the opposite outcome at the transcript and protein levels, as expected.[61] Very recently, the overexpression of DIS3L2 and hnRNP U in HCC was reported to promote the expression of the RAC1B isoform, therefore, we may consider the hypothesis of a similar mechanism in CRC.[23]

The goals for this work are described below and mainly focused on the study of the above referred factors and their influence on the alternative splicing of the *RAC1* gene, including the elucidation of some of the corresponding molecular mechanisms.

1) Determine the effect of RANBP2, PTBP1, ESRP1, DIS3L2 and hnRNP U overexpression and depletion on RAC1B transcript and protein levels in colorectal cell lines;
2) Identify PTBP1 and ESRP1 RNA-binding motifs in the RAC1 transcript sequence.

3. Materials and Methods

3.1. Cell culture

The cell lines used in this experimental work were NCM460 and HT29, both maintained in RPMI supplemented with 10% (v/v) fetal bovine serum (FBS) (Invitrogen) and frequently checked for mycoplasm contamination. The cells were grown in a humified atmosphere at 37ºC containing 5% CO_2 and were always manipulated in a laminar flow hood with sterile materials to prevent culture contamination. As both cell lines are adherent, remaining attached to a solid substrate, a trypsin solution (Gibco) was used for subculturing. Cells were always used in this study until a maximum of 22 passages.

The NCM460 cell line was originally obtained from the normal colon mucosa of a 68 year old Hispanic male with a gastric tumour, and selected through subculturing techniques, presenting some stem cell properties.[62] These epithelial cells are not exactly like the original tissues, but are very similar, constituting an important model for *in vitro* studies.[61,62]

HT29 is a human colon adenocarcinoma cell line, isolated from a primary tumour of a 44 year old Caucasian female.[63] These cells are comparable to human fetal colonic cells, but are malignant, having a high rate of glucose consumption and carrying a $BRAF^{V600E}$ mutation.[63]

The required passage confluence for NCM460 cells is approximately 80% while for HT29 is around 100%.[61]

3.2. Cell transfection and lysis

In this work, a chemical method was used for cell transfection, where positively charged lipids bind the negatively charged nucleic acids to be transfected creating positively charged complexes, which in turn bind the cell membranes and are subsequently taken by the cells, resulting in variations of gene expression.[64,65]

The plasmid transfections were performed in 6-well plates, with a 60-80% confluence for NCM460 and 40-60% confluence for HT29 cells. Per well, 10 µl of LipofectAMINE 2000 (Invitrogen) and 2.5 µg of the desired DNA were used (except for RANBP2 where 3 µg were used) according to the manufacturer's instructions. The transfected plasmids coded for the cDNAs of RANBP2, PTBP1, ESRP1, DIS3L2 or hnRNP U, and a plasmid coding for GFP (pcDNA3_GFP) was used as control for these experiments.

The RANBP2 expression vector (pkTol2Chy_RANBP2_1-3224) was kindly provided by Dr. Masakazu Hamada from the Mayo Clinic in Rochester, Minnesota, USA, and the DIS3L2 expression vector (pFLAG_CMV2_DIS3L2) was kindly provided by Dr. Richard I. Gregory from the Stem Cell Program, Boston Children's Hospital, Boston, USA. Besides the hnRNP U expression vector (pFLAG_CMV2_hnRNPU) which will be described in *Chapters 3.5* and *4.1*, the plasmids coding for PTBP1 (pcDNA_myc_PTBP1) and ESRP1 (pcDNA_myc_ESRP1) were previously cloned in the host lab on a pCR™2.1-TOPO® vector from the TOPO® TA Cloning® kit (Invitrogen), and then subcloned into a pcDNA3_myc expression vector (engineered by the host lab).[61]

The siRNA transfections were performed in 24-well plates, with a desired confluence of 30-40% for NCM460 and 20-30% for HT29 cells. Per well, 2 µl of LipofectAMINE 2000 and 50 ρmol of the desired siRNAs were used (except for RANBP2 where 60 ρmol were used) according to the manufacturer's instructions. The transfected siRNAs targeted the endogenous proteins corresponding to the above DNA plasmids, and an siRNA against luciferase was used as transfection control for these experiments.

Cell lysis was done 24 h after transfection with DNA plasmids or 48 h after siRNA transfection, by scrapping the cells in NP40 lysis buffer (50 mM Tris-HCl; pH 7.5; 2 mM $MgCl_2$; 100 mM NaCl; 10% Glycerol; 1% NP40) with protease inhibitors (1.5 μM aprotinin; 23 μM leupeptin; 10 μM E64; 1 μM EGTA; 15 μM pepstatin A; 1 μM PMSF; 1 μM 1,10-phenanthroline). The resulting lysate was then divided into two aliquots. From the first, RNA extraction was performed, using the RNA extraction kit NucleoSpin® RNA (Macherey Nagel), according to the manufacturer's instructions. To the second aliquot, sample buffer 5x (250 mM Tris-HCl, pH 6.8; 25% Glycerol; 10% SDS; 325 mM DTT; 1.25 mg Bromophenol blue) with Benzonase (0.2% v/v) and $MgCl_2$ (1% v/v) was added to obtain the total protein extract.

3.3. Semi-quantitative RT-PCR and agarose gel electrophoresis

In order to analyse gene expression at the transcript level, 2 μg of RNA from the cell lysates were reverse transcribed into cDNA using the qScript™ XLT cDNA SuperMix (Quanta Bio), according to the manufacturer's instructions. As in some experiments the obtained RNA concentration was too low to allow the use of the desired 2 μg, the maximum possible quantities of RNA were used for the synthesis of cDNA and the same amount of RNA was always used between samples of the same assay.

The cDNA was analysed through polymerase chain reaction (PCR), a fast method used to amplify specific DNA fragments through a series of temperature cycles, with each cycle consisting of the DNA denaturation, annealing of primers (small terminal oligonucleotides) and synthesis of complementary DNA through the incorporation of dNTPs by a DNA polymerase.[66] Serial dilutions of control cDNA were made to construct an equation of linear regression and demonstrate the linearity of the amplification reaction, enabling the semi-quantitative comparison of the obtained PCR products.

The semi-quantitative RT-PCR reactions were performed in a C1000 Touch™ thermal cycler (BIO-RAD), using buffer B (10 mM Tris-HCl pH 9 1.5 M; 50 mM $MgCl_2$; 1.5 mM KCl; 0.1% Bacta Gelatin (DifcoLab); 0.2 mM of each dNTP), GoTaq® G2 Flexi DNA polymerase (Promega) and the primers described in Table 3.1.

Table 3.1 - List of forward (F) and reverse (R) primers used for the semi-quantitative RT-PCR reactions. Respective nucleotide sequence and melting temperature (Tm) in degrees Celsius are indicated. Tm was calculated using the ApE software (v 2.0.60).

Target	Primer sequence (5'→ 3')	Tm (ºC)
RAC1 / RAC1B	**F:** CATGATCGACTACGACGTTCCTGATTATGCGG	65
RAC1	**R:** ACAAGCAAATTGAGAACACATCTGTT	58
RAC1B	**R:** ATATCCTTACCGTACGTTTCTCCAA	57
RANBP2	**F:** CCATGAGGCAGAGAGGAACA	57
	R: GGTCACAGGCCATCATTTCC	57
PTBP1	**F:** CCAGCCCATCTACATCCAGT	57
	R: ACATCCAGGGTCACAGGGTA	59
ESRP1	**F:** CCCCTACAAATGTTAGAGACTGT	55
	R: GCGGCCGCTAAATACAAACCCATTCTTTGGG	67
DIS3L2	**F:** CACAGCATTGCAAAGCAACT	56
	R: AGACTGGACTGCCTCAAGGA	59
hnRNP U	**F:** AGAGGTGGAATGCCCAACAG	58
	R: TCTGACCCCAGAATTGACCC	58
Pol2	**F:** CGCAATGAGCAGAACGGCGC	63
	R: TCTGCATGGCACGGGGCAAG	64

The optimized PCR programmes which were used in this thesis work are described in Supplementary table I.

A Green GoTaq® Flexi Buffer (Promega) was added to the final amplified products, which were separated by size on a 2.5% agarose gel with Tris-Borate-EDTA buffer 1x (TBE 1x) and 0.5 μg/ml of ethidium bromide (EtBr), using an electric current of 70-90V. The agarose gel electrophoresis technique allows the separation of DNA fragments by molecular weight as the loaded DNA samples go through the pores of an agarose gel when an electric current is applied, being the migration distance inversely proportional to the size of the DNA molecules.[67] Given that EtBr, which becomes intercalated between the DNA bases, emits fluorescence upon UV light exposure, the gel was then visualized in a transilluminator (UVITEC Cambridge) and photographed for further analysis.[67]

3.4. SDS-PAGE and Western blot

Protein extracts obtained as described in *Chapter 3.2* were denatured for 10 min at 95ºC, in order to be analysed by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE). This technique allows the separation of proteins by molecular weight, as an SDS buffer masks their different charges, creating a uniform negative charge and consequently allowing their migration towards the positively charged electrode.[68] The samples were loaded on an 8-12% (v/v) polyacrylamide gel with a 4% (v/v) stacking layer, using a Precision Plus Protein™ Standards protein ladder (BIO-RAD) as reference. The SDS-PAGE was then performed in a Mini-PROTEAN Vertical Electrophoresis Cell (BIO-RAD) chamber at 20 mA per gel for about 1 h in SDS buffer (8.7 mM SDS, 63 mM Tris, 480 mM Glycine).

Subsequently, a Western blot was performed, being used for specific protein detection and as a semi-quantitative process to evaluate protein expression variations.[68] Proteins were transferred to a Immun-blot® polyvinylidene difluoride (PVDF) membrane (BIO-RAD) using a Mini Trans-Blot® Cell (BIO-RAD), a wet electroblotting system, at 100 V for 1 h in blot buffer (48 mM Tris, 38.6 mM Glycine, 1.41 mM SDS, 20% (v/v) Methanol), and then stained and fixed with a Coomassie solution (0.25% (m/v) Brilliant Blue G (Sigma-Aldrich), 45% (v/v) Methanol, 10% (v/v) Acetic acid).

After removing the excess stain with a destaining solution (45% (v/v) Methanol, 10% (v/v) Acetic acid) and washing the membranes 3 times with TBST (20 M Tris, 6.7 M NaCl, 3% (v/v) HCl 37%, 0.05% (v/v) Triton X-100 pH 7.6) under agitation, they were incubated for an hour with a blocking solution of 5% (m/v) non-fat dry milk in TBST, to avoid unspecific binding of antibodies. The primary antibody incubation was made overnight, at 4ºC with agitation, using the antibodies and dilutions described in Table 3.2. After washing the membranes 3 times for 5 min with TBST, they were incubated for 1 h under agitation with the corresponding horseradish peroxidase-conjugated (HRP) secondary antibody (Goat Anti-Rabbit IgG or Goat Anti-Mouse IgG, BIO-RAD) with a dilution of 1:3000 on 5% (m/v) non-fat dry milk in TBST.

After three final washings for 10 min with TBST, the membranes were submitted to an enhanced chemiluminescence (ECL) reaction in a dark chamber with a mix of two solutions, being the first one composed of 100 mM Tris-HCL (pH 8.8), 37.5 ρM luminol (Roche) and 0.4 mM p-coumaric acid (Roche), and the second one by 100 mM Tris-HCL (pH 8.8) and 49 mM hydrogen peroxide (Sigma-Aldrich). The HRP on the secondary antibodies catalyses the reaction between the luminol, a hydrogen donor, and the hydrogen peroxide, a hydrogen acceptor, creating a luminescent signal, which was captured through exposure to X-ray films (Fujifilm).[69] The films were automatically developed in a Medical X-ray Processor (KODAK) and digitalized for further analysis.

Table 3.2 - List of the primary antibodies used for Western blot experiments. Respective commercial suppliers and used dilutions on 5% (m/v) non-fat dry milk in TBST are indicated.

Primary antibody	Supplier	Dilution
Mouse Anti-RAC1	Merck - Millipore	1:1000
Rabbit Anti-RAC1B	Merck - Millipore	1:2000
Mouse Anti-RANBP2	Santa Cruz Biotechnology	1:500
Mouse Anti-PTBP1	Santa Cruz Biotechnology	1:500
Rabbit Anti-ESRP1	Invitrogen - Thermo Fisher Scientific	1:1000
Rabbit Anti-DIS3L2	Novus Biologicals	1:250
Mouse Anti-hnRNP U	Santa Cruz Biotechnology	1:2000
Mouse Anti-Tubulin	Sigma-Aldrich	1:600000
Mouse Anti-PCNA	Merck - Millipore	1:3000
Rabbit Anti-GFP	Abcam	1:500 / 1:2000
Mouse Anti-Myc	Sigma-Aldrich	1:2000
Mouse Anti-FLAG	Sigma-Aldrich	1:500

3.5. Cloning and mutagenesis

The plasmids used on the overexpression assays were already available in the host lab, as previously referred, with the exception of a plasmid coding for hnRNP U. Therefore, pcDNA3.1-hnRNPU-V5, a gift from Dr. Susana Valente (Addgene plasmid #35974; http://n2t.net/addgene:35974; RRID:Addgene_35974), was used after being subcloned in pFLAG-CMV-2 (Sigma), constituting the henceforth referred pFLAG_CMV2_hnRNPU.

DNA cloning consists in the insertion of a DNA fragment into a self-replicating genetic element, through the action of restriction enzymes and a DNA ligase, forming a recombinant DNA molecule.[70,71]

Both the empty pFLAG-CMV-2 and the pcDNA3.1-hnRNPU-V5 were digested overnight in 10% (v/v) NEBuffer 2 (New England Biolabs®) and 10% (v/v) BSA (New England Biolabs®) at 37ºC, with the restriction enzymes BamHI (Promega) and HindIII (Promega). To the pcDNA3.1-hnRNPU-V5 an additional restriction enzyme PvuI (New England Biolabs®) was added to the mix, in order to cut the pcDNA3.1 plasmid, and consequently avoid religation of the original plasmid fragments.

The enzymatic cleavage was confirmed by running a sample of each digestion on a 1% agarose gel, and the remaining digested pFLAG-CMV-2 was incubated for 1 h at 65ºC with 1 U rAPid Alkaline Phosphatase (Sigma-Aldrich) to remove phosphate groups from the 5' region, avoiding self-ligation of the vector.[70] Both digestions were then purified using the NucleoSpin™ Gel and PCR Clean-up kit (Macherey-Nagel™) and the ligation was performed in a 1:7 ratio (v/v) of the vector and insert respectively, for 1 h at room temperature, using the Rapid DNA Dephos & Ligation Kit (Sigma-Aldrich). Subsequently, 5 µl of the ligation product were used to transform 50 µl of NZY5α Competent Cells (NZYTech) through a heat shock method consisting of the incubation of the bacteria for 2 h on ice, 45 sec at 42ºC, and then 2 min on ice. Afterwards, 250 µl of S.O.C. medium (Invitrogen™) were added and the bacteria were incubated at 37ºC for 2 h under agitation, before being spread onto an LB agar plate with 150 µg/ml ampicillin and incubated at 37ºC overnight.

From the obtained colonies, some were picked with a pipette tip and resuspended in 20 µl of water for further analysis through PCR screening, using 5 µl of sample and the primers FLAG1F (Table 3.3) and the previously described reverse primer for hnRNP U (Table 3.1). A replica LB agar plate was also made for the selected colonies. A volume of 5 µl from each positive colony suspension, meaning that there was amplification in the PCR screening of that colony, was subsequently grown in liquid LB medium with 100 µg/ml ampicillin, overnight at 37ºC under agitation. After DNA extraction from the

bacterial culture using the NZYMiniprep kit (NZYTech), DNA concentration was measured in a NanoDrop™ Spectrophotometer (Thermo Fisher Scientific) and DNA sequencing was performed.

To achieve the second goal of this work, two putative locations of ESRP1 binding were identified for further validation, based on the works of Dittmar *et al.* and Warzecha *et al.* where it was described that ESRP1 had a UG-rich binding motif.[50,72] Site directed mutagenesis was performed on the identified sites in a *RAC1* minigene that contains the region between exons 3 and 4 of the *RAC1* gene. This procedure involves the amplification of the DNA plasmid using oligonucleotides which are complementary to the region to be mutated but have an internal mismatch of one or more nucleotides, which will be specifically changed.[73] Two different sets of primers were used, originating the henceforth referred mutant 1 or mutant 2 with the Mut1_F (5'-CTTGGGTGTCCCGTGTAATGGGGCTGTCTG) and Mut1_R (5'-CAGACAGCCCCATTACACGGGACACCCAAG) or the Mut2_F (5'-CTAGGGCAG CGTGAGGGAAGTTGTCAGCTGTGAAG) and Mut2_R (5'-CTTCACAGCTGACAACTTCCCT CACGCTGCCCTAG) primers respectively. In this PCR reaction a Pfu DNA polymerase (Promega) was used in order to lower the error rate of the reaction and the used programme is described in Supplementary table I.

After the PCR reaction was completed and verified through agarose gel electrophoresis, the original methylated DNA strand was digested for 2 h at 37ºC with the DpnI enzyme (Thermo Fisher Scientific) and the resulting product was used to transform One Shot® Top 10 bacteria (Thermo Fisher Scientific), following the previously described cloning procedure. Automated DNA sequencing was performed to confirm the presence of the mutations.

To assure there were no further undesired mutations in the plasmids, a fragment containing the mutation was subcloned into the original *RAC1* minigene. The plasmids were digested overnight at 37ºC with ClaI (New England Biolabs®) and SacII (New England Biolabs®) restriction enzymes in buffer A (Sigma-Aldrich), or NotI (NZYTech) and SacII restriction enzymes in NZYSpeedyBuffer (NZYTech), in the case of mutant 1 or 2 respectively along with the original *RAC1* minigene for each. Digestion confirmation and vector dephosphorylation were performed following the above referred procedures for hnRNP U. Total volume of digested plasmids was run on a 1% low melting temperature agarose gel, from where the band which included the mutation and the one corresponding to the original vector were cut and, finally, purified using the NucleoSpin™ Gel and PCR Clean-up kit (Macherey-Nagel™). The final steps of ligation, bacterial transformation and liquid culture inoculation were also performed as previously described but using One Shot® Top 10 bacteria.

3.6. DNA sequencing

To confirm the correct hnRNP U subcloning in the pFLAG-CMV-2 and to confirm the presence of the desired mutations in the *RAC1* minigene, DNA samples were prepared for automated DNA sequencing at the in-house sequencing facility (*Unidade de Tecnologia e Inovação, Departamento de Genética Humana*).

The sequencing process is based on the chain termination method, consisting in the amplification of a DNA sample using a reaction mix that includes a small amount of ddNTPs associated with different fluorescent dyes.[74] During amplification, dNTPs are incorporated by DNA polymerase forming a new DNA strand; nevertheless, whenever a ddNTP is incorporated the elongation step terminates, so that this process results in the production of many fragments of the same DNA with different sizes, corresponding to all termination site locations.[74] The sequence of the resulting product is then identified by capillary gel electrophoresis. In this procedure, a laser is used to excite the dyes that correspond to one of the four ddNTP as the produced fragments exit the gel.[61,74] Because migration distance is inversely proportional to the size of the DNA molecules, smaller fragments are firstly

detected and the order by which the four nucleotides are identified corresponds to the sample DNA sequence.[61,74]

For each reaction an amount of 350 ng of DNA were used for 4 ρmol of the chosen primer, 1 μl of 1x BigDye™ terminator ready mix and 1 μl of sequencing buffer (BigDye™ Terminator v3.1 Cycle Sequencing Kit, Applied Biosystems™). The used primers are described on Table 3.3 and the primers for hnRNP U described on Table 3.1 were also used for the hnRNP U subcloning confirmation. The PCR programme that was used for these reactions is described in Supplementary table I.

The samples were analysed in an automatic DNA sequencer 3130xl Genetic Analyzer (Applied Biosystems™). The resulting DNA sequences, corresponding to the obtained chromatograms, were analysed with Nucleotide BLAST® (https://blast.ncbi.nlm.nih.gov/Blast.cgi), being compared to the expected reference DNA sequences.

Table 3.3 - List of primers used for automated DNA sequencing. Respective nucleotide sequence and melting temperature (Tm) in degrees Celsius are indicated. Tm was calculated using the ApE software (v 2.0.60).

Target	Primer	Primer sequence (5'→ 3')	Tm (ºC)
hnRNP U	pCMV5-1F	**F:** GGGACTTTCCAAAATGTCGTA	54
	FLAG1F	**F:** ATGGACTACAAAGACGATGA	52
	hGH-PA-R	**R:** CCAGCTTGGTTCCCAATAGA	56
	hnRNPU-1F	**F:** AAGAAGCGACGCCTTTCTGA	58
	hnRNPU-2R	**R:** TCTTACCTCCCGCCTGCT	59
	hnRNPU-3F	**F:** GATCATGGCCGTGGATATTT	54
	hnRNPU-4R	**R:** TTCTTCTTCAACAGGTGGCTG	56
	hnRNPU-5F	**F:** GGCGTTGCCTTCAAAATCAGT	58
	hnRNPU-6R	**R:** TGGCTTTTCCTTCTGACCAA	56
	hnRNPU-7R	**R:** CAGCTACCTCTGGGAGGGTAA	59
Mutant 1 and 2	RacRealEx3b-F	**F:** GGGCAAAGACAAGCCGATTG	58
	RacEx4Not-R	**R:** GCGGCCGCTTTGCACGGACATTTTCAAATG	68

3.7. Data analysis

Densitometric analyses of the RT-PCR gels and Western blot results were performed on the corresponding digital images using the *ImageJ* software (2.1.0/1.53c version). In the case of RT-PCR, the densitometric analysis of the control cDNA serial dilutions was used to build an equation of linear regression, which in turn was used to determine the concentration of the cDNA samples. Since the goal of these experiments was to assess variations on RAC1 alternative splicing, the RAC1B/RAC1 ratio was calculated for both the overexpression and depletion assays. For Western blot, the measured values were directly used to determine the RAC1B/RAC1 ratio.

Statistical analysis was performed using the *Microsoft excel* software (16.43 version) with the Real Statistics Resource Pack (https://www.real-statistics.com). Statistical significance between results was evaluated through a one-way ANOVA followed by a Dunnett's test, with significance set at $p < 0.05$. Values were presented as mean with respective standard error of the mean (SEM).

4. Results

4.1. Subcloning of hnRNP U for overexpression experiments

As described in *Chapter 3.5*, pcDNA3.1-hnRNPU-V5 was first subcloned into a pFLAG-CMV-2 expression plasmid to allow the detection of its expression in colorectal cell lines through Western blot with anti-epitope tag antibodies, comparable to the detection strategy used for the other four proteins of interest (see below).

Successful enzymatic digestion of both original vectors was confirmed through agarose gel electrophoresis (Figure 4.1) by comparing to 100 ng of undigested plasmids. As expected, pFLAG-CMV-2 digestion with BamHI and HindIII originated a fragment with 4622 bp of length, as the remaining 57 bp fragment is too small to be detected, whereas pcDNA3.1-hnRNPU-V5 digestion with BamHI, HindIII and PvuI resulted in three different DNA fragments with sizes of 4033 bp, 2481 bp and 1455 bp. The 2481 bp fragment was the one corresponding to the desired hnRNP U insert DNA.

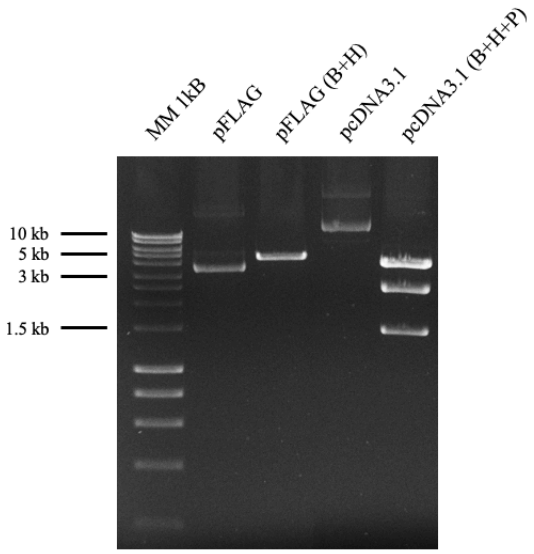


Figure 4.1 – Agarose gel electrophoresis of the digested pFLAG-CMV-2 and pcDNA3.1-hnRNPU-V5. DNA plasmids digested overnight with restriction enzymes were visualized in a 1% agarose gel. Undigested pFLAG-CMV-2 (pFLAG) consists of 4679 bp and originates one visible fragment of around 4600 bp after enzymatic digestion with BamHI (B) and HindIII (H), while undigested pcDNA3.1-hnRNPU-V5 (pcDNA3.1), with 7972 bp, leads to three fragments of about 4000 bp, 2500 bp and 1500 bp upon digestion with BamHI, HindIII and PvuI (P). The length of the obtained fragments was estimated through comparison with a molecular marker (MM 1kB) HyperLadder™ 1kb (Bioline).

After DNA purification, ligation and bacterial transformation, many colonies were obtained. A total of 32 colonies were then selected to assess the presence of the desired DNA plasmid. As described in *Chapter 3.5*, a sample of each of these colonies was submitted to a PCR screening using a forward primer with specificity for the FLAG expression tag and a reverse primer for the hnRNP U insert.

The resulting amplification product was expected to have a size of around 6150 bp, and two colonies were identified as positive using this methodology (Figure 4.2), therefore, the corresponding bacterial samples were grown in LB medium for plasmid purification, to be further studied.

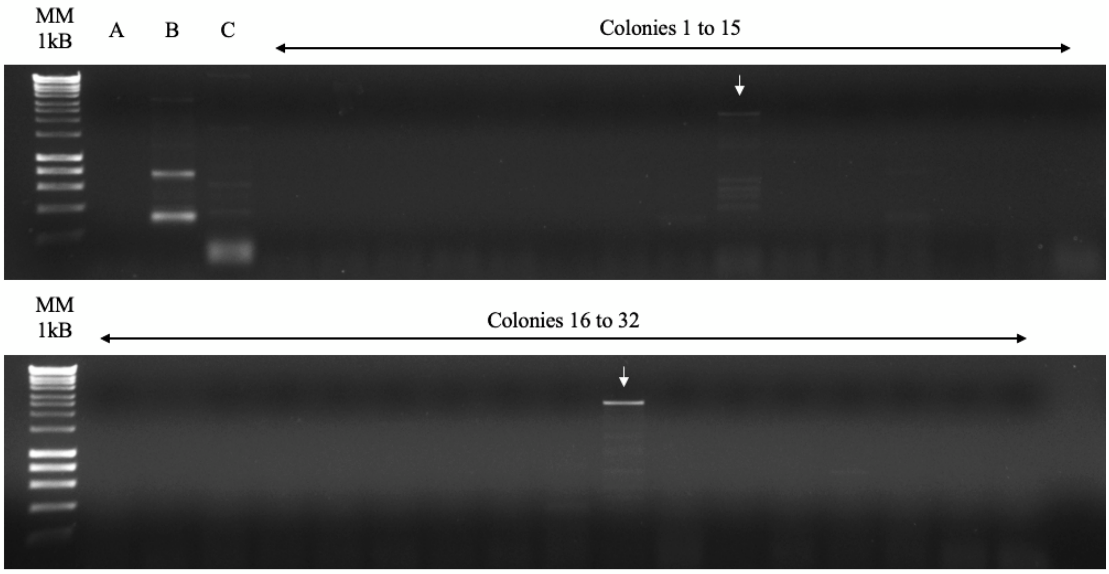


Figure 4.2 – PCR screening of colonies to detect hnRNP U subcloning. Agarose gel electrophoresis of the screening products from 32 chosen colonies. Experiment controls are presented on the upper panel, consisting in PCR mix without DNA (A) pcDNA3.1-hnRNPU-V5 (B) and pFLAG-CMV-2 (C). On colonies 9 and 25 there was DNA amplification (identified by arrows), corresponding to the desired PCR product. The length of the obtained fragments was estimated through comparison with a molecular marker (MM 1kB) HyperLadder™ 1kb (Bioline).

Prior to DNA sequencing, the purified DNA of these colonies was submitted to enzymatic digestion with BamHI and HindIII in order to confirm the presence of a 2500 bp fragment corresponding to the hnRNP U cDNA (Figure 4.3, left panel). A digestion with BamHI and MfeI was also performed in order to confirm if the insert was, in fact, subcloned into the pFLAG-CMV-2, given that this plasmid is not digested by MfeI unlike pcDNA3.1, where hnRNP U was initially cloned. The results of this latter experiment showed a unique DNA fragment for both samples, corresponding to the linearized plasmid due to the action of BamHI but not MfeI (Figure 4.3, right panel).

Following growth of the two positive colonies and plasmid DNA purification, automated sequencing was performed for the most concentrated DNA sample (colony 25) as described in *Chapter 3.6*, showing that the pFLAG_CMV2_hnRNPU sequence was correct, not being detected any undesired mutations when comparing to the expected DNA sequence. Thus, the desired expression vector tool was obtained and allowed to proceed in this work.

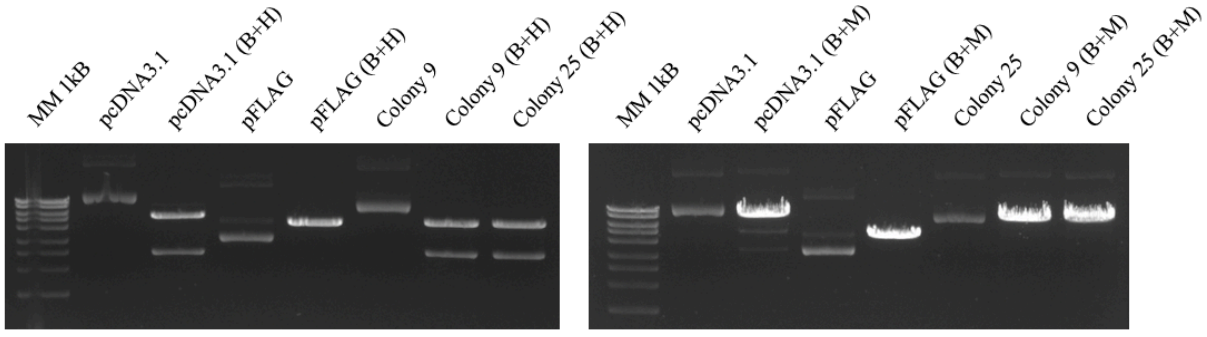


Figure 4.3 - Enzymatic digestion tests of the positive colonies obtained upon hnRNP U subcloning. Agarose gel electrophoresis of undigested pFLAG-CMV-2 (pFLAG), pcDNA3.1-hnRNPU-V5 (pcDNA3.1) and colony 9 or 25, in comparison with the same plasmids digested with the restriction enzymes BamHI (B) and HindIII (H) (left panel) or digested with BamHI and MfeI (M) (right panel). Both colonies presented a fragment with about 2500 bp upon digestion with BamHI and Hind III, and a single band with around 6150 bp after digestion with BamHI and MfeI. The length of the obtained fragments was estimated through comparison with a molecular marker (MM 1kB) HyperLadder™ 1kb (Bioline).

4.2. The effect of RANBP2, PTBP1, ESRP1, DIS3L2 or hnRNP U overexpression on RAC1B transcript and protein levels in colorectal cell lines

DNA plasmids coding for these splicing-related proteins were transfected into NCM460 and HT29 cells, as these cell lines endogenously express RAC1 and RAC1B (Supplementary figure I). Furthermore, the use of more than one cell line allows the distinction between cell line-dependent or -independent mechanisms, as they present different genetic properties. A vector for GFP expression was also used in these experiments as a control.

The protein expression from the transfected vectors in each cell line was confirmed through Western blot analysis using specific antibodies for the respective tags (Figure 4.4), except for GFP-tagged RANBP2 which was not detected in the HT29 cell line.

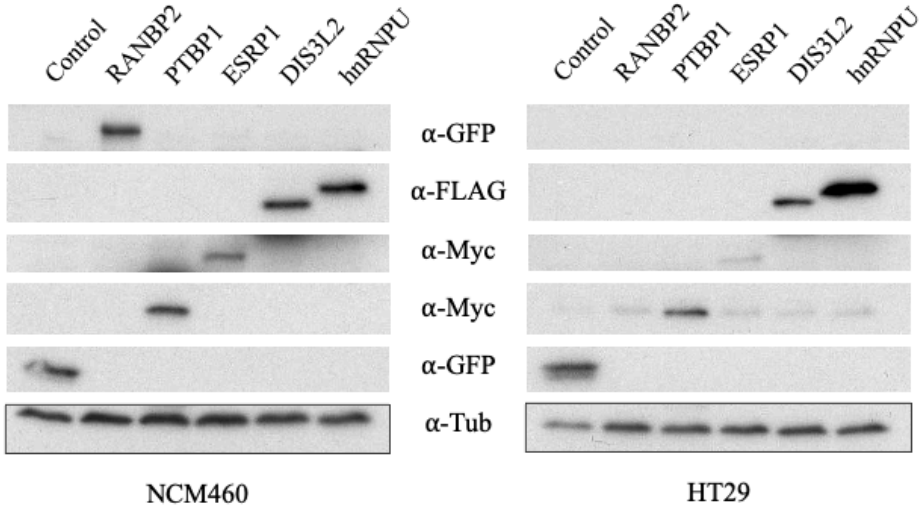


Figure 4.4 – Western blot analysis of the expression of different transfected vectors in NCM460 and HT29 cells. Expression vectors pcDNA3_GFP (30 kDa), pkTol2Chy_RANBP2_1-3224 (388 kDa), pcDNA_myc_PTBP1 (57 kDa), pcDNA_myc_ESRP1 (76 kDa), pFLAG_CMV2_DIS3L2 (100 kDa) and pFLAG_CMV2_hnRNPU (120 kDa) were transfected into NCM460 cells (left) and HT29 cells (right). Transfected plasmids were detected through anti-GFP, anti-Myc or anti-FLAG antibodies (center). Detection of tubulin (Tub), through an anti-Tubulin antibody, was used as loading control.

In order to access the effect of RANBP2, PTBP1, ESRP1, DIS3L2 and hnRNP U overexpression in the alternative splicing of *RAC1*, the inclusion of RAC1 exon 3b was studied in both cell lines at the endogenous transcript level through semi-quantitative RT-PCR, using specific primers for RAC1B and RAC1, as well as at the protein level by Western blot, using specific antibodies for RAC1B and RAC1. Densitometric quantification and statistical analysis of the experimental results were performed as described in *Chapter 3.7*.

Overexpression of GFP-RANBP2 in the NCM460 cell line led to higher RAC1B expression relatively to RAC1 at the protein level, but not at the transcript level (Figure 4.5 A, B). Overexpression of the remaining proteins in NCM460 cells did not show a statistically significant effect on RAC1B expression levels, neither through RT-PCR nor Western blot.

Concerning the HT29 cell line, none of the overexpressed proteins caused a statistically significant variation of RAC1B expression relatively to RAC1, both at the transcript and protein levels (Figure 4.5 C, D).

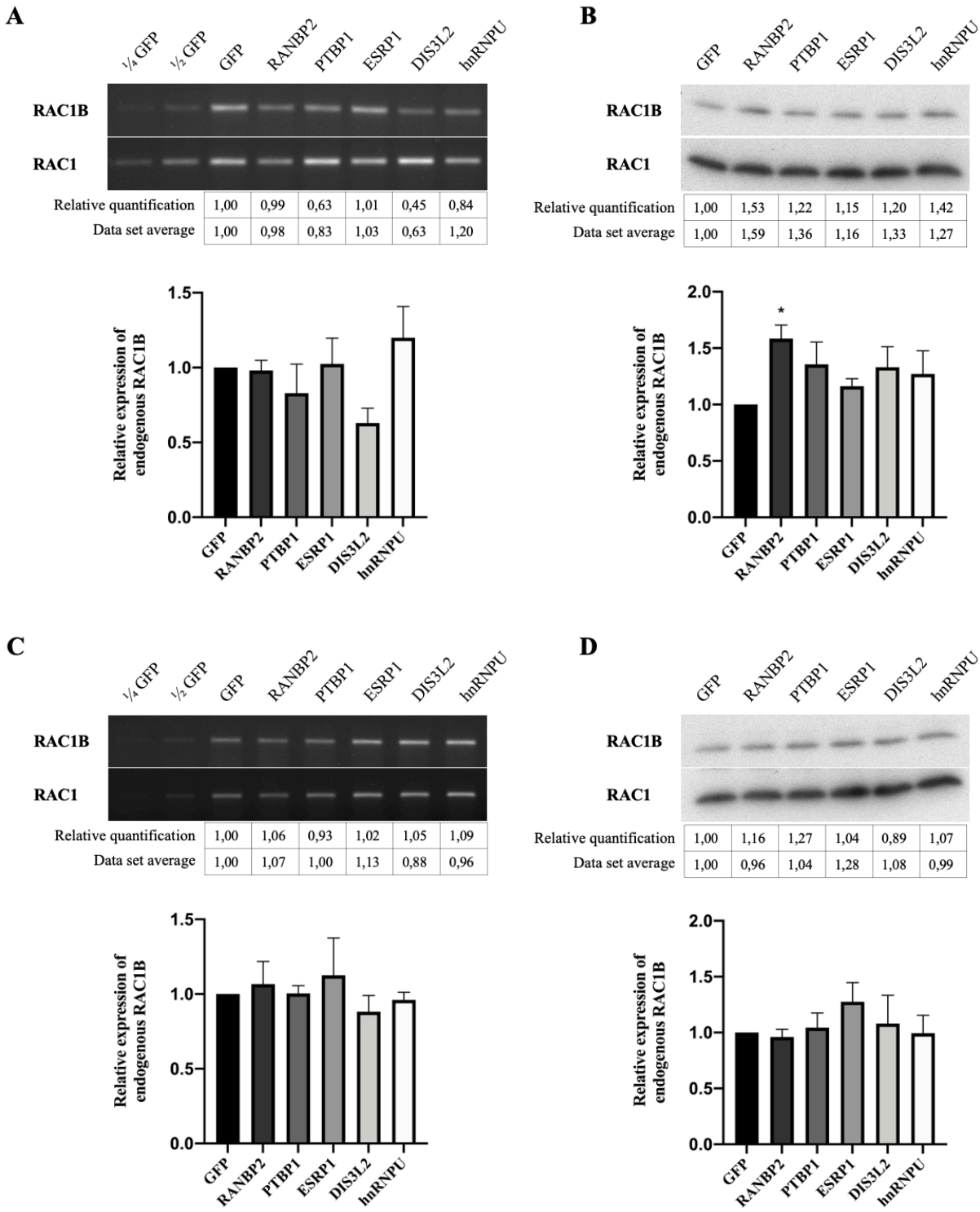


Figure 4.5 – Analysis of RAC1B expression relatively to RAC1 upon RANBP2, PTBP1, ESRP1, DIS3L2 or hnRNP U overexpression in NCM460 and HT29 cells. (A) Representative semi-quantitative RT-PCR of the transfection effect on RAC1 exon 3b inclusion using NCM460 cells (upper panel), **(B)** Representative Western blot of the transfection effect on RAC1B protein levels using NCM460 cells (upper panel), **(C)** Representative semi-quantitative RT-PCR of the transfection effect on RAC1 exon 3b inclusion using HT29 cells (upper panel) and **(D)** Representative Western blot of the transfection effect on RAC1B protein levels using HT29 cells (upper panel). Respective relative quantifications of the shown image are presented below, as well as the average values obtained from all replicate assays, both normalized to the control. Graphical representation of the obtained results (created with GraphPad software 9.0.0 version) is presented as mean ± error bars (SEM) of at least three independent experiments (lower panels). Statistically significant results were considered to have p-value ≤ 0.05 and represented as *p≤0.05; **p≤0.01; ***p≤0.001; **** p≤0.0001.

4.3. The effect of RANBP2, PTBP1, ESRP1, DIS3L2 or hnRNP U depletion on RAC1B transcript and protein levels in colorectal cell lines

Given that in pilot experiments the endogenous expression of RANBP2, PTBP1, ESRP1, DIS3L2 and hnRNP U was detected in the used cell lines, the effect of the depletion of these splicing-related proteins was also studied in this work in order to complement the results obtained from the overexpression experiments. Thus, commercially available siRNAs were transfected into NCM460 and HT29 cells, with an siRNA for luciferase being used as a control for these experiments.

The siRNAs for RANBP2, ESRP1 and DIS3L2 were previously tested in the host lab, being most effective when cells were analyzed at 48 h or 72 h after cell transfection, while the siRNAs for PTBP1 and hnRNP U were tested for the first time in this work. NCM460 cells were lysed 24 h, 48 h or 72 h after transfection with the new siRNAs and the lysates were analyzed through Western blot and, after RNA extraction, by RT-PCR. The analysis showed that siPTBP1 was more efficient at the 48 h time point and the sihnRNPU was efficient in all incubation times. Taking this into account, in all the experiments performed in this work cells were lysed 48 h after transfection with siRNAs, as this time point was considered as efficient for all of them.

The depletion caused by each of the siRNAs was assessed in each cell line through RT-PCR analysis using specific primers for the cDNA of the protein in study, and also by Western blot analysis using specific antibodies for the proteins (Figure 4.6). The used siRNAs were quite efficient in both cell lines, with the exception of siRANBP2 which did not produce a great difference between the controls and the RANBP2 RNA or protein levels in the HT29 cell line.

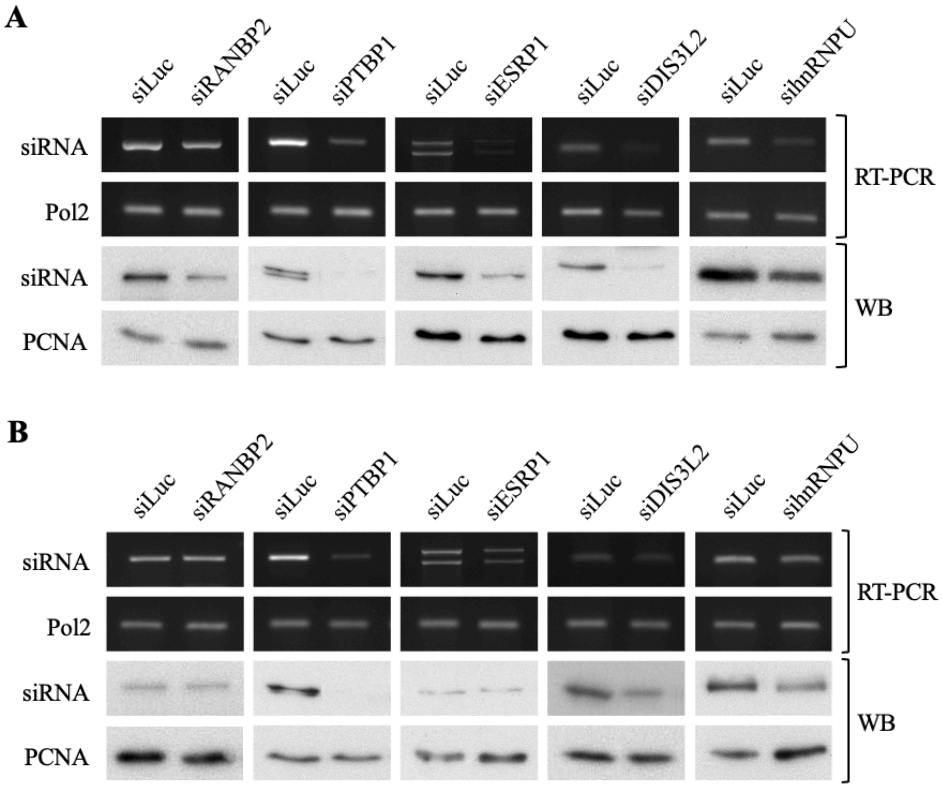


Figure 4.6 – Analysis of endogenous RANBP2, PTBP1, ESRP1, DIS3L2 or hnRNP U expression levels in NCM460 and HT29 cells after siRNA-mediated depletion. (A) Analysis in NCM460 and **(B)** HT29 cell lysates 48 h after transfection with the indicated siRNAs. Representative semi-quantitative RT-PCR (top panels) and Western blot analyses (bottom panels) show the siRNA's effect on the corresponding target RNA, using specific primers for each cDNA, or protein levels, using specific antibodies for each protein. Detection of polymerase II (Pol2) with specific primers was used as control for the RT-PCR assays, while detection of PCNA was used as a control for the Western blot experiments.

In this case, the used loading control for the Western blot assays was PCNA because the molecular weight of endogenous PTBP1 was coincidental with the molecular weight of tubulin, interfering with its detection.

The depletion effect of these splicing-related proteins in the RAC1B expression relatively to RAC1, was studied in both cell lines at the transcript level by semi-quantitative RT-PCR and also at the protein level through Western blot. Densitometric quantification and statistical analysis of the experimental results were performed as described in *Chapter 3.7*.

Regarding the NCM460 cell line (Figure 4.7), when comparing to the control, the siRANBP2 caused an increased inclusion of the RAC1 exon 3b both at the transcript and, even more drastically, at the protein level. Conversely, a decrease in the inclusion of this exon was observed upon PTBP1 depletion, also through RT-PCR and Western blot analysis, with a greater variation at the protein level.

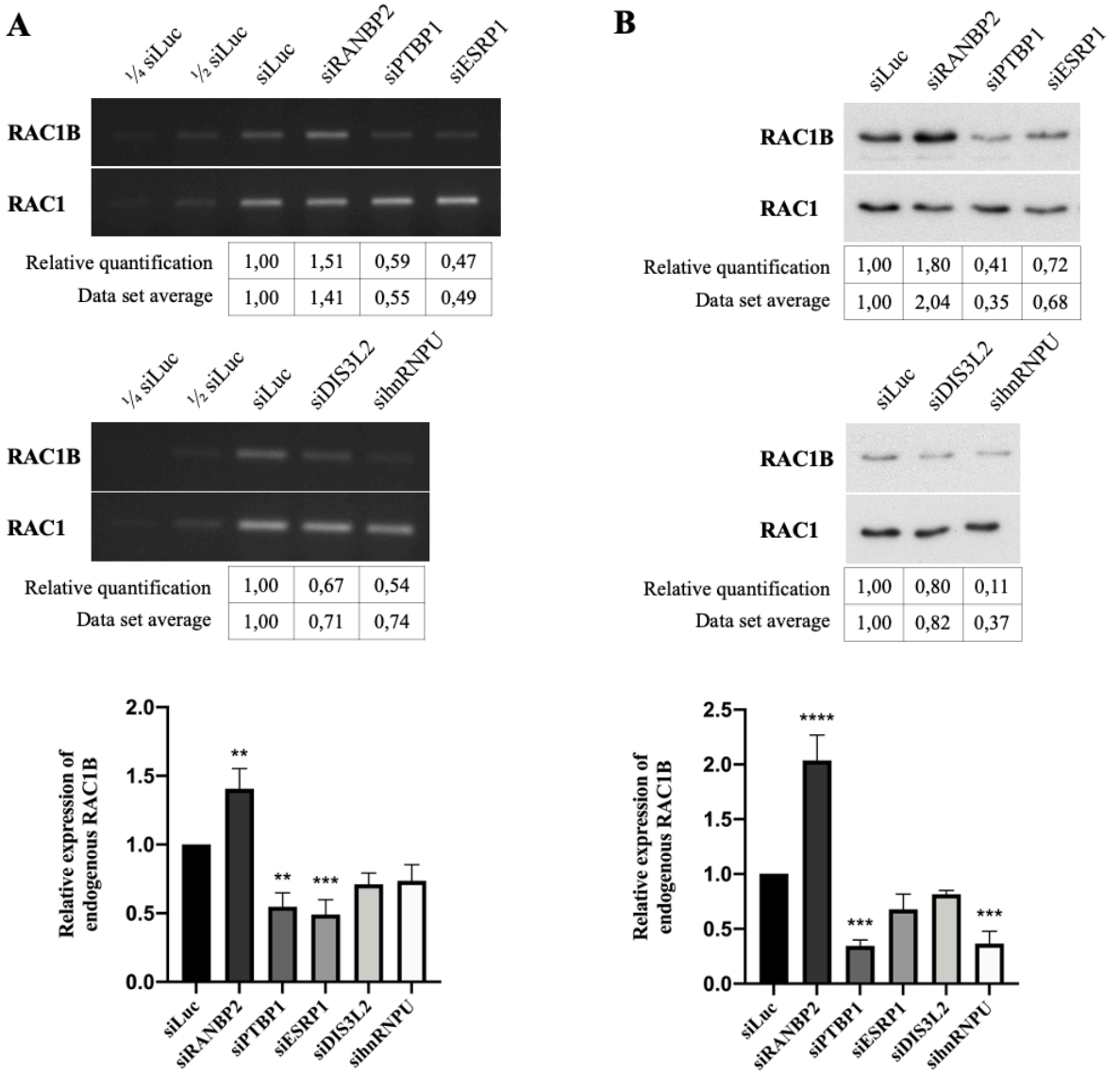


Figure 4.7 – Analysis of RAC1B expression relatively to RAC1 upon RANBP2, PTBP1, ESRP1, DIS3L2 or hnRNP U depletion in NCM460 cells. (A) Representative semi-quantitative RT-PCR of the transfection effect on RAC1 exon 3b inclusion and **(B)** Representative Western blot of the transfection effect on RAC1B protein levels using NCM460 cells (upper and middle panels). Respective relative quantifications are presented below, as well as the average values of all replicate assays, both normalized to the control. Graphical representation of the obtained results (created with GraphPad software 9.0.0 version) is presented as mean ± error bars (SEM) of at least three independent experiments (lower panels). Statistically significant results were considered to have p-value ≤ 0.05 and represented as *p≤0.05; **p≤0.01; ***p≤0.001; **** p≤0.0001.

The siESRP1 transfection in NCM460 cells led to a decrease of RAC1B transcript relatively to RAC1, but in terms of protein expression there was not a statistically significant variation. On the contrary, the sihnRNPU showed an effect at the protein level, decreasing RAC1B expression, but no statistically significant changes were observed in relation to RNA splicing.

Concerning the depletion of DIS3L2, it did not show any statistically significant effect on RAC1 exon 3b inclusion levels, neither through RT-PCR nor Western blot analysis.

For equivalent experiments using the HT29 cell line (Figure 4.8), the transfection of siPTBP1 led to a decrease of RAC1 exon 3b inclusion when comparing to the siLuc control. Similarly, for sihnRNPU the obtained results showed a lower RAC1B expression relatively to RAC1, both through RT-PCR and Western blot analysis.

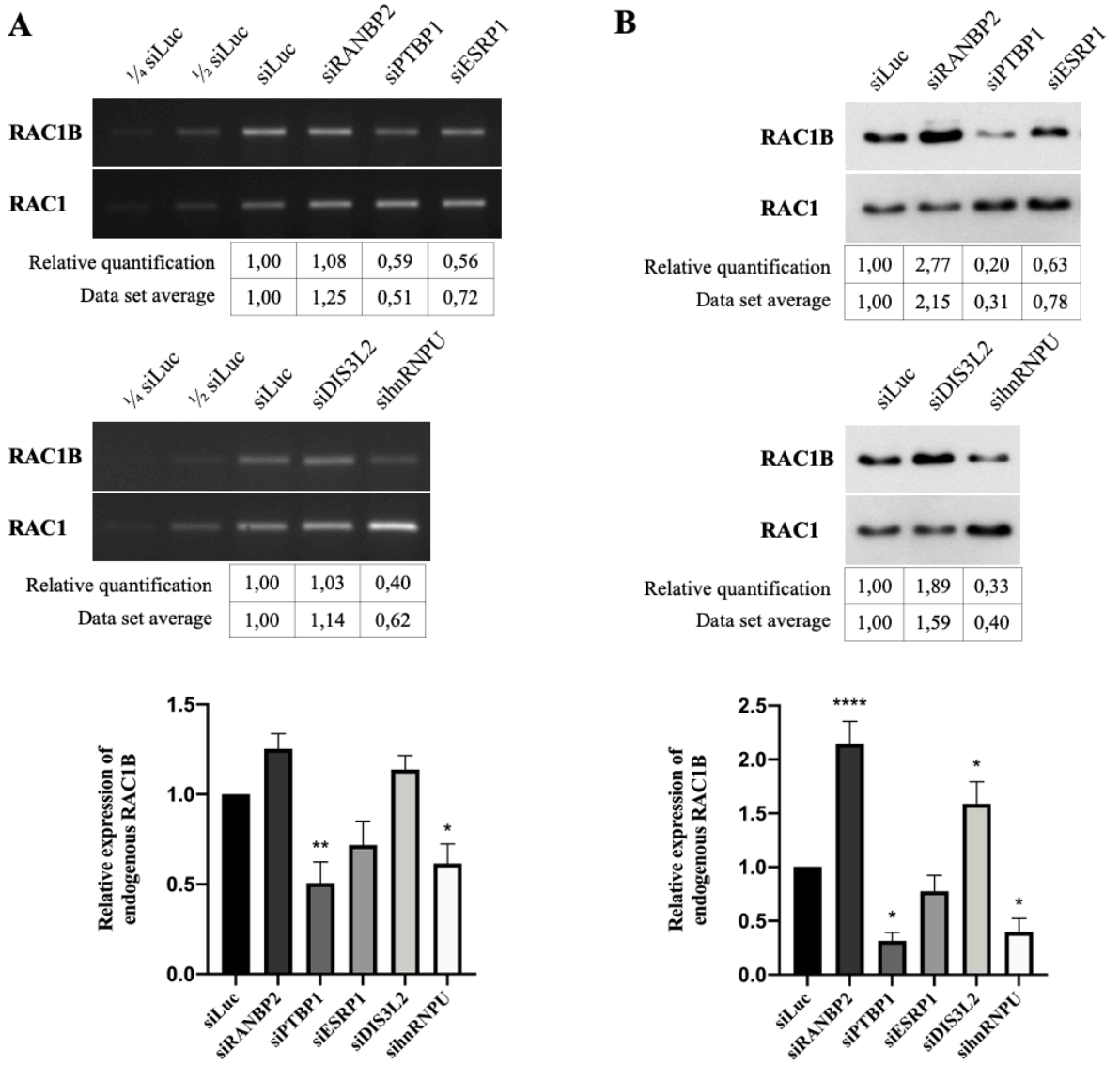


Figure 4.8 – Analysis of RAC1B expression relatively to RAC1 upon RANBP2, PTBP1, ESRP1, DIS3L2 or hnRNP U depletion in HT29 cells. (A) Representative semi-quantitative RT-PCR of the transfection effect on RAC1 exon 3b inclusion and **(B)** Representative Western blot of the transfection effect on RAC1B protein levels using HT29 cells (upper and middle panels). Respective relative quantifications are presented below, as well as the average values of all replicate assays, both normalized to the control. Graphical representation of the obtained results (created with GraphPad software 9.0.0 version) is presented as mean ± error bars (SEM) of at least three independent experiments (lower panels). Statistically significant results were considered to have p-value ≤ 0.05 and represented as *p≤0.05; **p≤0.01; ***p≤0.001; **** p≤0.0001.

Although the siRANBP2 was inefficient (Figure 4.6), the depletion of this protein in HT29 cells caused an increase in RAC1B at the protein level; however, at the transcript level there was not a statistically significant variation. The results for siDIS3L2 were very similar, leading to an increase in the inclusion levels of exon 3b in terms of protein.

In the case of ESRP1 depletion, the obtained results did not reflect any statistically significant effect on RAC1 exon 3b inclusion levels, neither through RT-PCR nor Western blot analysis.

4.4. Mutagenesis of putative ESRP1 RNA-binding motifs in the RAC1 transcript sequence using a *RAC1* minigene

As described in *Chapter 3.5*, two possible binding sites for the ESRP1 splicing factor located in intron 3b, were mutated in a *RAC1* minigene and automated DNA sequencing was performed to confirm the presence of the mutations (Figure 4.9). The obtained results showed that the sequences were successfully mutated, as in mutant 1 two guanine residues were changed to two adenines and in mutant 2 a thymine and a guanine were also changed to two adenine residues.

Although site directed mutagenesis was not performed in the case of PTBP1, two possible RNA-binding sites were also identified for this splicing factor (Figure 4.9).

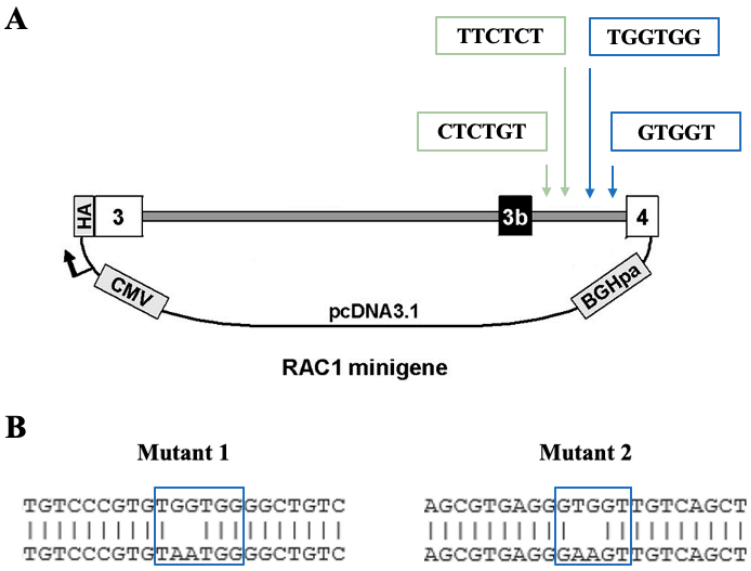


Figure 4.9 – Selection of putative binding motifs for PTBP1 and ESRP1 on the RAC1 minigene and analysis of mutated DNA sequencing results. (A) Schematic representation of the RAC1 minigene, indicating the location of possible binding sites for PTBP1 (green) or ESRP1 (blue) in the RAC1 3b intron. (Adapted from Gonçalves *et al.* 2009) **(B)** Confirmation of the wanted mutations for ESRP1 in intron 3b through the alignment of the original *RAC1* minigene sequence (top sequence) with the automated DNA sequencing results of mutants 1 or 2 (bottom sequence). Selected RNA-binding motifs are indicated in blue squares. Images obtained with Nucleotide BLAST® (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

To avoid any undesired mutations that could have occurred during the mutagenesis process, the mutated regions of the plasmid should be subcloned back into the original *RAC1* minigene. The overnight digestion of the plasmids was visualized through agarose gel electrophoresis, together with the undigested plasmids as controls (Figure 4.10). For mutant 1, the digestions with the ClaI and SacII restriction enzymes resulted in two fragments, one of them with 10751 bp and another with 588 bp, while the digestions with the NotI and SacII enzymes for mutant 2 resulted in a fragment with 10324 bp and a second one with 1015 pb. The mutated sites were located on the smaller fragment of each mutant.

The subsequent steps were made as described in *Chapter 3.5*, however, no colonies or a low number of small colonies was obtained after bacterial transformations and the DNA extracted from the bacterial cultures never presented a satisfactory concentration to be used in further experiments.

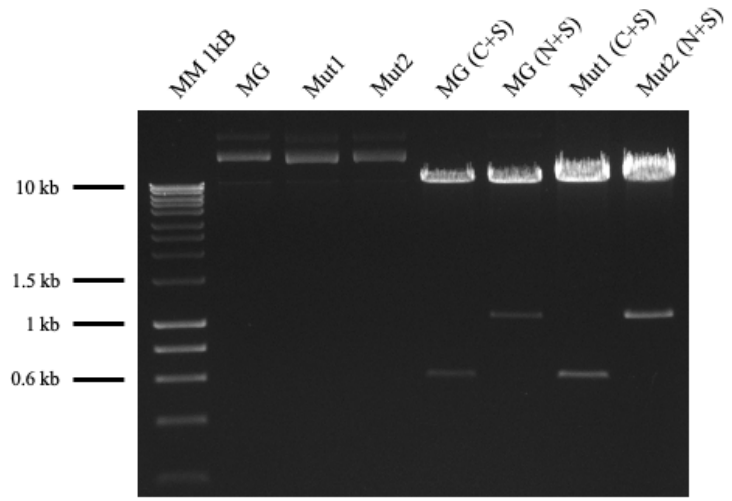


Figure 4.10 – Agarose gel electrophoresis of the digested RAC1 minigene and mutant plasmids. DNA plasmids digested overnight with restriction enzymes were visualized in a 1% agarose gel. The original minigene (MG) and the mutant 1 (Mut1) or mutant 2 (Mut2) plasmids, were correctly digested with the restriction enzymes SacII (S) and ClaI (C) or NotI (N) respectively. Undigested plasmids were used as a control. The length of the obtained fragments was estimated through comparison with a molecular marker (MM 1kB) HyperLadder™ 1kb (Bioline).

5. Discussion and Conclusions

Colorectal cancer is a common malignancy with significant mortality, which presents different subtypes due to genetic heterogeneity.[4] The focus of this work was on a subtype that is characterized by MSI and mutations in the *BRAF* gene, namely $BRAF^{V600E}$, which is the most common BRAF mutation and is known to cause an overstimulation of the MAPK pathway, being associated, in most cases, with RAC1B overexpression.[3,5,6] RAC1B is a splicing variant of the *RAC1* gene and is characterized by the inclusion of an additional exon, designated as exon 3b, resulting in some regulation and signalling differences when compared to RAC1.[24,28] Besides CRC, RAC1B upregulation was previously found to be associated with breast, lung, thyroid and pancreatic cancers.[10,31–34] RAC1B has been described to promote tumour progression, as it is an hyperactive form of RAC1 able to stimulate the NF-κB and Wnt pathways, to increase cell survival and cell cycle progression, to activate EMT, to cause ROS-mediated DNA damage and to decrease cell adhesion.[6,24,28] For these reasons, RAC1B is considered as a potential therapeutic target, as it may allow the design of targeted therapies for patients with this subtype of CRC, but for this, understanding in more detail the mechanisms of RAC1B regulation is essential. Previous research projects in the host lab had already identified SRSF1 and SRSF3 as two splicing factors involved in the regulation of RAC1B splicing in CRC.[14] SRSF1 overexpression led to an increase in the RAC1B isoform relatively to RAC1, while its depletion resulted in a decrease of exon 3b inclusion.[14] In contrast, SRSF3 showed the opposite effect in this specific splicing event, as its overexpression caused a decrease in RAC1B expression and its silencing increased the inclusion of this alternative exon.[14]

The first goal of this work was to determine the effect of a nucleoporin, RANBP2, and four splicing factors, PTBP1, ESRP1, DIS3L2 and hnRNP U, on the expression of RAC1B in colorectal cell lines. Two colorectal cell lines that express RAC1B endogenously were selected for this work. HT29 cells have their origin in a human colon adenocarcinoma with $BRAF^{V600E}$ mutation, and NCM460 cells were derived from an immortalized normal colon mucosa and showed in previous studies a higher transfection efficiency compared to other colorectal cell lines.[61]

The effect of the previously referred proteins on RAC1B alternative splicing was assessed through their overexpression and depletion in the chosen cell lines. Overexpression was achieved by transfecting expression vectors coding for the respective cDNAs to be transcribed by the cell machinery, while depletion was accomplished through transfection of siRNAs, which lead to degradation of their specific complementary mRNAs that are transcribed in the cells, thus preventing their translation into protein. Analysis of RAC1 exon 3b inclusion at the mRNA level was made by RT-PCR, while for protein expression, the used methods were SDS-PAGE and Western blot.

Regarding RANBP2, previous studies by Navarro *et al.* described it as a necessary protein for the correct function of kinetochores, mitotic spindles and centrosomes, acting as a tumour suppressor, as it prevents mitotic and chromosomal abnormalities.[40] However, Vecchione *et al.* described RANBP2 as an essential gene for the survival of BRAF-mutant cancer cells, as its depletion was shown to increase the time spent in mitosis, eventually leading to cell death.[43] Given that RAC1B was previously shown to synergize with BRAF to promote cell survival and cell cycle progression, the effect of RANBP2 in RAC1B expression was of great interest for this study.

RANBP2 depletion led to an increase in RAC1 exon 3b inclusion in NCM460 cells at the transcript and protein levels and the same effect was observed in HT29 cells (Figures 4.7 and 4.8) although at the transcript level the effect did not reach statistical significance but rather a tendency for the increase in these cells. Given that the siRNA efficiency of RANBP2 in the HT29 cell line was low, it could explain the lack of statistical significance (Figure 4.6 B). It is, however, also possible that at the time of cell lysis, at 48 h post-transfection, the siRNAs were already losing their effect on the transcript so that endogenous RAC1B mRNA levels were recovering although its protein levels were still depleted. To clarify this situation, this siRNA should be tested in the HT29 cell line with several incubation periods through RT-PCR and Western blot analyses.

These results indicate that RANBP2 normally acts as a silencer for RAC1B expression, causing exon skipping, possibly through its role in the regulation of nucleocytoplasmic transport of phosphorylated SR proteins into the nucleus, which consequently affects alternative splicing, as described by Saitoh *et al.*.[41] Specifically for RAC1B, the nuclear levels of the SRSF1 protein were already shown to determine the amount of RAC1B expressed.[21,35]

Another explanation behind these observations may be related to the enzymatic function of RANBP2. First, the sumoylation of TCF-4 by RANBP2 was shown to enhance its ability to bind β-catenin, which led to an increase of their transcriptional activity in CRC cell lines.[75] As described by Gonçalves *et al.*, the activation of β-catenin/TCF-4 increases the expression of SRSF3, a RAC1B silencer; so that RANBP2 depletion may be decreasing SRSF3 expression via β-catenin/TCF-4, resulting in the increase of exon 3b inclusion.[14] Second, RANBP2 is also responsible for IRAK2 sumoylation upon Toll-like receptor stimulation, which is important for IRAK2 nuclear translocation.[76] Moreover, Zhou *et al.* showed that IRAK2 may phosphorylate SRSF1 which facilitates its dissociation from target mRNAs and their consequent nuclear export.[76] In this case, as SRSF1 is a RAC1B enhancer, its dissociation from it could prevent the exon 3b inclusion. Another pathway to be considered here, is the PI3K/AKT/mTOR pathway which was previously described by Renner *et al.* as a transcriptional activator of RANBP2 in a prostate cancer model.[77] Given that Gonçalves *et al.* have previously documented in colorectal cell lines that upon PI3K inhibition there is an increase in SRSF1 and RAC1B levels, it is plausible that RANBP2 may act as a mediator in this mechanism.[14]

Unfortunately, these observations could not be confirmed through RANBP2 overexpression (Figure 4.5). Upon RANBP2 plasmid transfection, an increase of the RAC1B protein was observed in NCM460 cells, while at the transcript level there were no significant differences in the RAC1 splicing ratio; moreover, in the HT29 cell line there were no splicing variations. The difficulties in the obtention of statistically significant results for HT29 cells may be partly explained by the low transfection

efficiency of the expression vector pkTol2Chy_RANBP2_1-3224, which was not detected in HT29 cells through Western blot (Figure 4.4). Furthermore, given the high endogenous expression of RANBP2 in colorectal cells (https://www.proteinatlas.org/ENSG00000099901-RANBP1/tissue), the pathways in which it is involved may already be saturated and no more affected by the overexpression, which does not constitute a problem through the use of siRNAs. To overcome these difficulties, the transfection experiments could be further optimized, through the testing of other colorectal cell lines, transfection conditions or even a different transfection method.

In further experiments, the mechanisms by which RANBP2 is affecting RAC1B expression should be clarified, for example through the determination of how RANBP2 overexpression and depletion affect the expression levels of SRSF1 and SRSF3 in colorectal cell lines. It would also be interesting to study the RANBP2 role in the subcellular localization of these two SR proteins, or even other splicing factors, through RANBP2 overexpression or depletion followed by immunofluorescence assays targeting SRSF1 or SRSF3 and analysis by confocal microscopy.

PTBP1 is a major regulator of alternative splicing, despite having many other cellular functions, and it is known to be overexpressed in several tumours.[44] It was described by Fu *et al.* that PTBP1 silencing in HT29 cells influenced their invasiveness in a dose-dependent manner, so that the influence of this factor on RAC1B expression was investigated in this thesis.[48]

The transfection of an siRNA for PTBP1 in NCM460 and HT29 cells caused a decrease in the RAC1B expression relatively to RAC1 (Figures 4.7 and 4.8), indicating that this splicing factor acts as a RAC1B enhancer. These findings are in agreement with the results of Hollander *et al.* that used bioinformatic tools and RT-PCR to show that PTBP1 overexpression, through the MAPK pathway, acts as a stimulator of the alternative splicing isoform RAC1B in different human tissues and cell lines.[45] However, in this thesis work PTBP1 was overexpressed in the same two cell lines, but no significant variations were observed in the RAC1 exon 3b inclusion levels (Figure 4.5), although the plasmid transfection and PTBP1 expression seemed to be efficient (Figure 4.4).

Similarly to RANBP2, a high expression of PTBP1 was described for colorectal cell lines (https://www.proteinatlas.org/ENSG00000011304-PTBP1/tissue), which may be interfering with the overexpression results, as the pathways which are activated by the expression of this protein may already be saturated. As it was previously described that the effects of PTBP1 silencing may be different depending on its downregulation level, this may also happen for its overexpression, given that the obtained results for RAC1B relative expression were highly variable between experiments.[48]

Although it was not possible to solve these issues in this work, future experiments should clarify how PTBP1 affects this specific splicing event. Through crosslinking immunoprecipitation and computational models, Han *et al.* predicted several binding motifs and splicing sites of PTBP1.[78] Besides the previously described binding preference for polypyrimidine-rich sequences, PTBP1 binding sites were enriched within the downstream intron, for exon sets that are enhanced by this splicing factor, as in the case of RAC1 exon 3b.[78]

With this type of model, putative PTBP1 binding sites can be identified in the exon 3b proximities and further validated. One of the goals of this work was to obtain a RAC1 minigene in which possible RNA-binding sites for PTBP1 were mutated, to be used in further experiments; however, as this procedure was not successful when it was applied to ESRP1 possible binding sites (further discussed in this chapter), site directed mutagenesis was not performed in the case of PTBP1.

With regard to ESRP1, this RNA-binding protein was previously described as upregulated during carcinogenesis and downregulated upon EMT, leading to genome-wide differences in splicing patterns, and RAC1B was identified as one of its target transcripts in oral squamous cancer cells.[49] Therefore, its role in RAC1B alternative splicing in colorectal cells was investigated.[49,50]

Statistically significant results were only obtained for the experiments of ESRP1 depletion in NCM460 colonocytes at the transcript level, where a decrease of RAC1B expression relatively to RAC1 was documented (Figure 4.7 A). In the remaining depletion assays, although the results were not considered statistically significant, a tendency for exon 3b exclusion may also be observed in both cell lines (Figures 4.7 B and 4.8). Once more, these results could not be complemented through the overexpression experiments, as the detected variations were not statistically significant (Figure 4.5), which can be partly explained by the fact that the expression of the tagged protein from the plasmid was not very pronounced (Figure 4.4). Nonetheless, these results seem to show a tendency for RAC1 exon 3b inclusion upon ESRP1 overexpression.

If the observed trends were confirmed in future experiments, the results for ESRP1 depletion and overexpression would be in accordance to one another, indicating that this splicing factor acts as a RAC1B enhancer in colorectal cancer. These findings contradict what was previously described for oral squamous cancer cells and lung cancer cells, where siESRP1 caused an increase in RAC1 exon 3b inclusion, however this is not an abnormal situation, given that many regulatory proteins function in a cell- or tissue-specific manner, causing wide variations in alternative splicing.[15,29,49] To clarify this complex regulatory process, the mechanisms by which ESRP1 influences RAC1B alternative splicing should be further studied.

In this work, two predicted ESRP1 binding motifs in the RAC1 transcript were identified and mutated in intron 3b, in order to verify if any of these changes could avoid ESRP1 binding and consequently affect the splicing process. If the binding site was, in fact, one of the selected ones there would be a variation in the inclusion levels of exon 3b in the final RAC1 transcript, when comparing the co-transfection of an ESRP1 expression vector with a mutated *RAC1* minigene or the wild type *RAC1* minigene.

Unfortunately, this goal was not accomplished yet because it was not possible to reclone the mutant 1 and mutant 2 fragments into the original minigene vector, to assure they did not have undesired mutations. As the selected strategy of subcloning did not produce the expected results, an alternative would be to sequence the mutated plasmids completely, specially focusing on the RAC1 gene fragment and promoter regions. If no extra mutations are found upon sequencing, the plasmids may be used for the planned experiments. Furthermore, a double mutant may be constructed to be similarly studied, through a new subcloning of the mutated region of one of the mutants into the other mutant plasmid, using restriction enzymes, and subsequent sequencing of the mutated regions. Alternatively, if subcloning in the *RAC1* minigene remains difficult to achieve, the double mutant may be constructed through a new mutagenesis of one of the mutants using the primers corresponding to the other mutant and with subsequent sequencing of the obtained DNA.

DIS3L2 is an RNA-binding protein and its role in cancer progression seems variable, when different types of cancer are considered.[23,52] It was previously described by Xing *et al.* that in HCC cells, DIS3L2 promotes RAC1 exon 3b inclusion, therefore, it was identified as an interesting candidate to be further investigated in this work, in a CRC context.[23]

Overexpression of DIS3L2 in NCM460 and HT29 cell lines did not result in any statistically significant variations in the RAC1B expression levels (Figure 4.5), however, its depletion allowed to draw some conclusions. In HT29 cells, the transfection of siDIS3L2 promoted exon 3b inclusion, significantly increasing the expression of the RAC1B protein but not its transcript, as the RT-PCR results showed a similar tendency for exon inclusion but were not considered statistically significant (Figure 4.8). Similarly to what was proposed for RANBP2, it is possible that at the time of cell lysis the endogenous RAC1B mRNA levels were already recovering from the siRNA transfection, although its protein levels were still affected. These results are contrary to what was previously documented for the DIS3L2 action on RAC1 splicing in HCC; however, as RAC1B was previously implicated in the

promotion of cell cycle progression, through the NF-κB pathway, and DIS3L2 overexpression was shown to suppress cell cycle progression, it is not unreasonable to suggest that, in CRC, this factor may act as a RAC1B inhibitor and consequently a tumour suppressor, as different cell lines present variable molecular mechanisms acting on alternative splicing events.[6,52]

The results for the impact of DIS3L2 depletion on RAC1B expression using the NCM460 cell line were not statistically significant (Figure 4.7), but intriguingly, a tendency for a decrease of exon 3b inclusion was observed, which antagonizes with the findings for HT29 cells, suggesting that these cell lines may behave differently. Taking into account the work of Xing *et al.*, where DIS3L2 was found to promote RAC1B expression, NCM460 cells seem to present a similar mechanism.[23] In HT29 cells, which present different genetic properties for being carcinogenic cells, DIS3L2 may play a different role, through the degradation of RNAs that promote RAC1B expression due to its RNB, activation of different signalling pathways, binding in other regions of the *RAC1* gene or inability to interact with hnRNPU, affecting the regulation of this alternative splicing event.

These experiments should be repeated in the future, in order to clarify if there is indeed a difference in the cells' mechanisms, even both being colorectal cell lines, and other colorectal cell lines should be tested. Moreover, the transfection protocols may be further optimized, although the transfection of the pFLAG_CMV2_DIS3L2 vector seems to be rather efficient in both cell lines.

The last splicing factor studied in this work was hnRNP U that has been previously described as a DNA- and RNA-binding protein involved in alternative splicing regulation, most frequently with cassette exon inclusion, and in RNA stability control.[37,58]

The experiments performed in HT29 cells showed that hnRNP U overexpression did not influence RAC1B expression relatively to RAC1 in a measurable way (Figure 4.5 C, D), which may be due to the very high endogenous expression of this protein in colorectal cells (https://www.proteinatlas.org /ENSG00000153187-HNRNPU/tissue), so that the pathways in which it is involved may already be saturated and no longer affected by its overexpression. Nonetheless, hnRNP U depletion significantly decreased exon 3b inclusion, both at the transcript and protein level, in this cell line (Figure 4.8) indicating that it may have a role as a RAC1B enhancer in colorectal cell lines.

As previously referred, Xing *et al.* have described for liver cell lines, that DIS3L2 promotes RAC1B expression via hnRNP U.[23] However, from the experiments described above in HT29 cells, only hnRNP U was identified as a putative RAC1B enhancer. This suggests that either a different mechanism may be used in this cell line where hnRNP U acts independently of DIS3L2 in the alternative splicing process, or the DIS3L2 depletion experiments led to wrong conclusions, reinforcing the need to better study these two splicing factors, also including other colorectal cell lines. In the future, the interaction between these two factors should be assessed in colorectal cells, for instance through co-immunoprecipitation experiments.

Unlike what was observed for DIS3L2, the depletion of hnRNP U in NCM460 cells had a similar effect to its depletion in HT29 cells, causing a decrease in the RAC1B protein expression and a similar tendency for decrease of RNA expression, although the variation did not present statistical significance in the latter case (Figure 4.7). As hypothesized in previous cases, at the time of cell lysis the siRNAs could be losing their effect at the transcript but not at the protein level. The corresponding overexpression experiments were not statistically significant neither through RT-PCR nor Western blot, though a tendency for exon 3b inclusion was observed (Figure 4.5 A, B), which would support the results obtained for the siRNA transfections and indicate that in these cells hnRNP U acts as a RAC1B enhancer.

Overall, the overexpression experiments performed in this work did not lead to clear conclusions about the effect of the proteins of interest on RAC1B expression. It is important to notice that although in most of the cases the transfected plasmids were being efficiently expressed in the cells, their detection was based on their epitope tag, which did not allow comparing the expression levels to those of the corresponding endogenous proteins. Thus, it is not known if the total amount of the respective proteins were increased by transfection, as there may be cellular feedback mechanisms regulating their expression levels. This is especially relevant for the HT29 cell line, because RAC1B may be essential for the survival of the cancer cells and, consequently, be more tightly regulated. Therefore, it could be interesting to assess the total amount of the proteins of interest in the overexpression assays, using specific antibodies for the protein and not the plasmid tag.

The siRNA assays, on the other hand, were rather successful, resulting in several statistically significant results regarding variations on the inclusion levels of RAC1 exon 3b. Nevertheless, the protocol may be further optimized, namely through the testing of each siRNA efficiency in the HT29 cell line, when cell lysis is performed at different time points after transfection. Here, these tests were made only for the NCM460 cell line and the period which was thought to be more adequate was applied to the HT29 cell line. In order to facilitate the experiments, a single time point was chosen at 48h and used in all siRNA transfections, which was effective for most cases; however, although all siRNAs were effective at 48h, this may not be the perfect incubation time for some of them, so a more systematic approach for each of them should be applied in further studies, through the use of different incubation periods.

The main goal of this thesis work was to investigate the role of RANBP2, PTBP1, ESRP1, DIS3L2 and hnRNP U in the alternative splicing of exon 3b in the *RAC1* pre-mRNA in colorectal cancer cells. The obtained results provided strong evidence that, in fact, these proteins influence RAC1B expression, both at the transcript and protein level. Taking into account the results obtained through depletion experiments, RANBP2 seems to be a RAC1B inhibitor in NCM460 cells, while PTBP1, ESRP1, DIS3L2 and hnRNP U are probably enhancers of that RAC1 isoform. In the HT29 cell line, RANBP2 and DIS3L2 were pointed as inhibitors of exon 3b inclusion, while PTBP1, ESRP1 and hnRNP U are most likely enhancers of that alternative splicing event.

Technical issues did not allow to achieve the second goal of this work in time, which was to identify, if and where PTBP1 and ESRP1 bind the RAC1 pre-mRNA in order to exert their effect on splicing regulation. Further studies should clarify not only these mechanisms, but also through which molecular pathways RANBP2 is regulating RAC1B expression, and whether the mechanism described for the regulation of RAC1B expression by DIS3L2 and hnRNP U in HCC is also true for CRC. The comprehension of how the alternative splicing of RAC1 is regulated, besides being scientifically interesting to understand fundamental molecular principles, may be useful to help in the development of novel therapies for patients with colorectal cancer, or even other cancer types, which present RAC1B overexpression, as the pathways which affect this exon inclusion may be pharmacologically inhibited, restoring a normal and healthy cell signalling.

6. Bibliography

1. Keum, N. N. & Giovannucci, E. Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 713–732 (2019).
2. W. Wong, R. & D'Angelo, M. Linking nucleoporins, mitosis, and colon cancer. *Cell Chem. Biol.* **176**, 139–148 (2017).
3. Matos, P., Gonçalves, V. & Jordan, P. Targeting the serrated pathway of colorectal cancer with

mutation in BRAF. *Biochim. Biophys. Acta - Rev. Cancer* **1866**, 51–63 (2016).
4. Arnold, M. *et al.* Global patterns and trends in colorectal cancer incidence and mortality. *Gut* **66**, 683–691 (2016).
5. Matos, P. & Jordan, P. Targeting colon cancers with mutated BRAF and microsatellite instability. *Adv. Exp. Med. Biol.* **1110**, 7–21 (2018).
6. Matos, P. *et al.* B-RafV600E Cooperates With Alternative Spliced Rac1b to Sustain Colorectal Cancer Cell Survival. *Gastroenterology* **135**, 899–906 (2008).
7. Roth, A. D. *et al.* Prognostic role of KRAS and BRAF in stage II and III resected colon cancer: Results of the translational study on the PETACC-3, EORTC 40993, SAKK 60-00 trial. *J. Clin. Oncol.* **28**, 466–474 (2010).
8. Fang, M., Ou, J., Hutchinson, L. & Green, M. R. The BRAF Oncoprotein Functions through the Transcriptional Repressor MAFG to Mediate the CpG Island Methylator Phenotype. *Mol. Cell* **55**, 904–915 (2014).
9. Sanz-Garcia, E., Argiles, G., Elez, E. & Tabernero, J. BRAF mutant colorectal cancer: Prognosis, treatment, and new perspectives. *Ann. Oncol.* **28**, 2648–2657 (2017).
10. Jordan, P., Brazão, R., Boavida, M. G., Gespach, C. & Chastre, E. Cloning of a novel human Rac1b splice variant with increased expression in colorectal tumors. *Oncogene* **18**, 6835–6839 (1999).
11. Matlin, A. J., Clark, F. & Smith, C. W. J. Understanding alternative splicing: Towards a cellular code. *Nat. Rev. Mol. Cell Biol.* **6**, 386–398 (2005).
12. Lee, Y. & Rio, D. C. Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annu. Rev. Biochem.* **84**, 291–323 (2015).
13. Wang, Y. *et al.* Mechanism of alternative splicing and its regulation. *Biomed. Reports* **3**, 152–158 (2015).
14. Gonçalves, V., Matos, P. & Jordan, P. Antagonistic SR proteins regulate alternative splicing of tumor-related Rac1b downstream of the PI3-kinase and Wnt pathways. *Hum. Mol. Genet.* **18**, 3696–3707 (2009).
15. Ule, J. & Blencowe, B. J. Alternative Splicing Regulatory Networks: Functions, Mechanisms, and Evolution. *Mol. Cell* **76**, 329–345 (2019).
16. De Conti, L., Baralle, M. & Buratti, E. Exon and intron definition in pre-mRNA splicing. *Wiley Interdiscip. Rev. RNA* **4**, 49–60 (2013).
17. Di, C. *et al.* Function, clinical application, and strategies of Pre-mRNA splicing in cancer. *Cell Death Differ.* **26**, 1181–1194 (2019).
18. Corbo, C., Orrù, S. & Salvatore, F. SRp20: An overview of its role in human diseases. *Biochem. Biophys. Res. Commun.* **436**, 1–5 (2013).
19. Staknis, D. & Reed, R. Direct interactions between pre-mRNA and six U2 small nuclear ribonucleoproteins during spliceosome assembly. *Mol. Cell. Biol.* **14**, 2994–3005 (1994).
20. Kim, E., Goren, A. & Ast, G. Alternative splicing: Current perspectives. *BioEssays* **30**, 38–47 (2008).
21. Gonçalves, V. *et al.* Phosphorylation of SRSF1 by SRPK1 regulates alternative splicing of tumor-related Rac1b in colorectal cells. *Rna* **20**, 474–482 (2014).
22. Bessa, C., Matos, P., Jordan, P. & Gonçalves, V. Alternative splicing: Expanding the landscape of cancer biomarkers and therapeutics. *Int. J. Mol. Sci.* **21**, 1–23 (2020).
23. Xing, S. *et al.* DIS3L2 promotes progression of hepatocellular carcinoma via hnRNP U-mediated alternative splicing. *Cancer Res.* **79**, 4923–4936 (2019).
24. Matos, P., Collard, J. G. & Jordan, P. Tumor-related Alternatively Spliced Rac1b Is Not Regulated by Rho-GDP Dissociation Inhibitors and Exhibits Selective Downstream Signaling. *J. Biol. Chem.* **278**, 50442–50448 (2003).
25. Sahai, E. & Marshall, C. J. RHO-GTPases and cancer. *Nat. Rev. Cancer* **2**, 133–142 (2002).
26. Joyce, D. *et al.* Integration of Rac-dependent regulation of cyclin D1 transcription through a nuclear factor-κB-dependent pathway. *J. Biol. Chem.* **274**, 25245–25249 (1999).
27. Zondag, G. C. M. *et al.* Oncogenic Ras downregulates Rac activity, which leads to increased Rho activity and epithelial-mesenchymal transition. *J. Cell Biol.* **149**, 775–781 (2000).
28. Melzer, C., Hass, R., Lehnert, H. & Ungefroren, H. RAC1B: A Rho GTPase with Versatile Functions in Malignant Transformation and Tumor Progression. *Cells* **8**, 21 (2019).

29. Seiz, J. R. *et al.* Different signaling and functionality of Rac1 and Rac1b in the progression of lung adenocarcinoma. *Biol. Chem.* **401**, 517–531 (2020).
30. Radisky, D. C. *et al.* Rac1b and reactive oxygen species mediate MMP-3-induced EMT and genomic instability. *Nature* **436**, 123–127 (2005).
31. Schnelzer, A. *et al.* Rac1 in human breast cancer: Overexpression, mutation analysis, and characterization of a new isoform, Rac1b. *Oncogene* **19**, 3013–3020 (2000).
32. Zhou, C. *et al.* The Rac1 splice form Rac1b promotes K-ras-induced lung tumorigenesis. **32**, 903–909 (2013).
33. Silva, A. L., Carmo, F. & Bugalho, M. J. RAC1b overexpression in papillary thyroid carcinoma: A role to unravel. *Eur. J. Endocrinol.* **168**, 795–804 (2013).
34. Mehner, C. *et al.* Tumor cell-derived MMP3 orchestrates Rac1b and tissue alterations that promote pancreatic adenocarcinoma. *Mol. Cancer Res.* **12**, 1430–1439 (2014).
35. Gonçalves, V. & Jordan, P. Posttranscriptional regulation of splicing factor SRSF1 and its role in cancer cell biology. *Biomed Res. Int.* **2015**, (2015).
36. Matos, P. & Jordan, P. Rac1, but not Rac1B, stimulates RelB-mediated gene transcription in colorectal cancer cells. *J. Biol. Chem.* **281**, 13724–13732 (2006).
37. Huelga, S. C. *et al.* Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. *Cell Rep.* **1**, 167–178 (2012).
38. Huang, X. *et al.* The association of RAN and RANBP2 gene polymerphisms with Wilms tumor risk in Chinese children. *J. Cancer* **11**, 804–809 (2020).
39. Hashizume, C., Kobayashi, A. & Wong, R. W. Down-modulation of nucleoporin RanBP2/Nup358 impaired chromosomal alignment and induced mitotic catastrophe. *Cell Death Dis.* **4**, e854-12 (2013).
40. Navarro, M. S. & Bachant, J. RanBP2: A Tumor Suppressor with a New Twist on TopoII, SUMO, and Centromeres. *Cancer Cell* **13**, 293–295 (2008).
41. Saitoh, N. *et al.* The distribution of phosphorylated SR proteins and alternative splicing are regulated by RANBP2. *Mol. Biol. Cell* **23**, 1115–1128 (2012).
42. R. Gareau, J. & D. Lima, C. The SUMO pathway: emerging mechanisms that shape specificity, conjugation and recognition. *Nat. Rev. Mol. Cell Biol.* **11**, 357–361 (2011).
43. Vecchione, L. *et al.* A Vulnerability of a Subset of Colon Cancers with Potential Clinical Utility. *Cell* **165**, 317–330 (2016).
44. Zhu, W. *et al.* Roles of PTBP1 in alternative splicing, glycolysis, and oncogensis. *J. Zhejiang Univ. Sci. B* **21**, 122–136 (2020).
45. Hollander, D. *et al.* A network-based analysis of colon cancer Splicing changes reveals a tumorigenesis-favoring regulatory pathway emanating from ELK1. *Genome Res.* **26**, 541–553 (2016).
46. Li, X., Han, F., Liu, W. & Shi, X. PTBP1 promotes tumorigenesis by regulating apoptosis and cell cycle in colon cancer. *Bull. Cancer* **105**, 1193–1201 (2018).
47. Wang, Z. N. *et al.* High expression of PTBP1 promote invasion of colorectal cancer by alternative splicing of cortactin. *Oncotarget* **8**, 36185–36202 (2017).
48. Fu, X. *et al.* Suppression of PTBP1 signaling is responsible for mesenchymal stem cell induced invasion of low malignancy cancer cells. *Biochim. Biophys. Acta - Mol. Cell Res.* **1865**, 1552–1565 (2018).
49. Ishii, H. *et al.* Epithelial splicing regulatory proteins 1 (ESRP1) and 2 (ESRP2) suppress cancer cell motility via different mechanisms. *J. Biol. Chem.* **289**, 27386–27399 (2014).
50. Warzecha, C. C. *et al.* An ESRP-regulated splicing programme is abrogated during the epithelial-mesenchymal transition. *EMBO J.* **29**, 3286–3300 (2010).
51. Ala, U. *et al.* Proteomics-based evidence for a pro-oncogenic role of ESRP1 in human colorectal cancer cells. *Int. J. Mol. Sci.* **21**, 1–15 (2020).
52. Luan, S., Luo, J., Liu, H. & Li, Z. Regulation of RNA decay and cellular function by 3′-5′ exoribonuclease DIS3L2. *RNA Biol.* **16**, 160–165 (2019).
53. Pashler, A. L., Towler, B. P., Jones, C. I. & Newbury, S. F. The roles of the exoribonucleases DIS3L2 and XRN1 in human disease. *Biochem. Soc. Trans.* **44**, 1377–1384 (2016).
54. Towler, B. P., Jones, C. I., Harper, K. L., Waldron, J. A. & Newbury, S. F. A novel role for the 3′-5′ exoribonuclease Dis3L2 in controlling cell proliferation and tissue growth. *RNA Biol.* **13**,

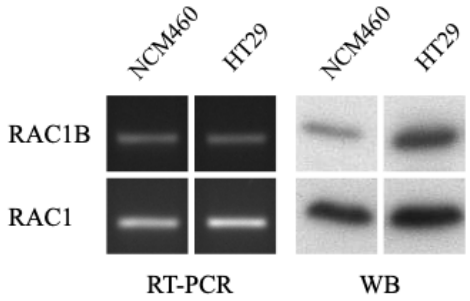
1286–1299 (2016).
55. Liu, W. *et al.* Knockdown of a DIS3l2 promoter upstream long noncoding RNA (AC105461.1) enhances colorectal cancer stem cell properties in vitro by down-regulating DIS3l2. *Onco. Targets. Ther.* **10**, 2367–2376 (2017).
56. Geuens, T., Bouhy, D. & Timmerman, V. The hnRNP family: insights into their role in health and disease. *Hum. Genet.* **135**, 851–867 (2016).
57. Vu, N. T. *et al.* HnRNP U enhances caspase-9 splicing and is modulated by AKT-dependent phosphorylation of hnRNP L. *J. Biol. Chem.* **288**, 8575–8584 (2013).
58. Yugami, M., Kabe, Y., Yamaguchi, Y., Wada, T. & Handa, H. hnRNP-U enhances the expression of specific genes by stabilizing mRNA. *FEBS Lett.* **581**, 1–7 (2007).
59. Zhang, L., Song, D., Zhu, B. & Wang, X. The role of nuclear matrix protein HNRNPU in maintaining the architecture of 3D genome. *Semin. Cell Dev. Biol.* **90**, 161–167 (2019).
60. Li, L. *et al.* Long noncoding RNA SFTA1P promoted apoptosis and increased cisplatin chemosensitivity via regulating the hnRNP-U-GADD45A axis in lung squamous cell carcinoma. *Oncotarget* **8**, 97476–97489 (2017).
61. Rodrigues, C. Regulation of the alternative splicing of RAC1b in tumour cells. (Faculty of Sciences, University of Lisbon, 2018).
62. Moyer, M. P., Manzano, L. A., Merriman, R. L., Stauffer, J. S. & Tanzer, L. R. NCM460, a normal human colon mucosal epithelial cell line. *Vitr. Cell. Dev. Biol. - Anim.* **32**, 315–317 (1996).
63. Martínez-Maqueda, D., Mirallez, B. & Recio, I. HT29 Cell Line. in *The Impact of Food Bioactives on Health: In Vitro and Ex Vivo Models* 113–124 (2015). doi:10.1007/978-3-319-16104-4_11
64. Kim, T. K. & Eberwine, J. H. Mammalian cell transfection: The present and the future. *Anal. Bioanal. Chem.* **397**, 3173–3178 (2010).
65. Haiyong, H. RNA Interference to Knock Down Gene Expression. *Methods Mol. Biol.* 1–9 (2018). doi:10.1007/978-1-4939-7471-9_16
66. Ramesh, R., Munshi, A. & Panda, S. K. Polymerase chain reaction. *Natl. Med. J. India* **5**, 115–119 (1992).
67. Videira, A. Metodologia de Análise de Genes e seus Produtos. in *Engenharia Genética, Princípios e aplicações* 49–74 (2011).
68. Bose, U., Wijffels, G., Howitt, C. A. & Colgrave, M. L. Proteomics: Tools of the Trade. in *Advances in Experimental Medicine and Biology* **1073**, 1–22 (2019).
69. Yang, L. *et al.* Study on enhancement principle and stabilization for the luminol-H2O2-HRP chemiluminescence system. *PLoS One* **10**, 1–14 (2015).
70. Videira, A. Clonagem de Genes. in *Engenharia Genética, Princípios e aplicações* 33–39 (2011).
71. Alberts, B. *et al.* Manipulating Proteins, DNA and RNA. in *Molecular Biology of the Cell* 540–552 (2008).
72. Dittmar, K. A. *et al.* Genome-Wide Determination of a Broad ESRP-Regulated Posttranscriptional Network by High-Throughput Sequencing. *Mol. Cell. Biol.* **32**, 1468–1482 (2012).
73. Carter, P. Site-directed mutagenesis. *Biochem. J.* **237**, 1–7 (1986).
74. Yang, D. L., Sauvageot, R. & Pentoney, S. L. DNA sequencing by capillary electrophoresis. *Handb. Capill. Microchip Electrophor. Assoc. Microtech. Third Ed.* 467–514 (2007).
75. Shitashige, M. *et al.* Regulation of Wnt Signaling by the Nuclear Pore Complex. *Gastroenterology* **134**, 1961–1971 (2008).
76. Zhou, H. *et al.* IRAK2 directs stimulus-dependent nuclear export of inflammatory mRNAs. *Elife* **6**, 1–25 (2017).
77. Renner, O. *et al.* Mst1, RanBP2 and eIF4G are new markers for in vivo PI3K activation in murine and human prostate. *Carcinogenesis* **28**, 1418–1425 (2007).
78. Han, A. *et al.* De Novo Prediction of PTBP1 Binding and Splicing Targets Reveals Unexpected Features of Its RNA Recognition and Function. *PLoS Comput. Biol.* **10**, (2014).

7. Supplementary data

Supplementary table I – PCR programmes used for RAC1, RAC1B, RANBP2, PTBP1, ESRP1, DIS3L2, hnRNP U and polymerase II amplification, for ESRP1 mutagenesis and DNA sequencing reactions.

RAC1		**RAC1B**		**RANBP2**		**PTBP1**		**ESRP1**	
94º - 5 min		94º - 5 min		94º - 5 min		94º - 5 min		94º - 5 min	
94º - 30 sec		94º - 30 sec		94º - 30 sec		94º - 30 sec		94º - 30 sec	
60º - 30 sec	x32	60º - 30 sec	x37	60º - 30 sec	x26	60º - 30 sec	x29	64º - 30 sec	x35
72º - 30 min		72º - 30 min		72º - 45 min		72º - 30 min		72º - 60 min	
72º - 5 min		72º - 5 min		72º - 5 min		72º - 5 min		72º - 5 min	

DIS3L2		**hnRNP U**		**Pol2**		**Mutagenesis**		**Sequencing**	
94º - 5 min		94º - 5 min		94º - 5 min		95º - 5 min		96º - 10 min	
94º - 30 sec		94º - 30 sec		94º - 30 sec		95º - 1 min		96º - 10 sec	
60º - 30 sec	x35	60º - 30 sec	x24	64º - 30 sec	x28	54º - 1 min	x20	54º - 5 sec	x25
72º - 30 min		72º - 30 min		72º - 30 min		68º - 24 min		60º - 4 min	
72º - 5 min		72º - 5 min		72º - 5 min		12º - ∞		4º - ∞	



Supplementary figure I – Endogenous expression of RAC1B and RAC1 in NCM460 and HT29 cell lines. Expression at the transcript level was assessed through RT-PCR (left), and at the protein level it was detected by Western blot (WB) (right).