

Escola Superior de Saúde Egas Moniz



**Mestrado em Biologia Molecular em Saúde**

**Estudo de padrões de expressão de  
transcritos alternativos do gene *tbccd1* em  
tecidos humanos e linhas celulares  
cancerígenas**

Dissertação para a obtenção de Grau de Mestre

Orientadora: Professora Doutora Helena Soares

Co-orientadora: Professora Doutora Luísa Cyrne

Ana Isabel Rodrigues Câmara

Setembro, 2013

Escola Superior de Saúde Egas Moniz

**Mestrado em Biologia Molecular em Saúde**

**Estudo de padrões de expressão de  
transcritos alternativos do gene *tbccd1* em  
tecidos humanos e linhas celulares  
cancerígenas**

Dissertação para a obtenção de Grau de Mestre

Orientadora: Professora Doutora Helena Soares

Co-orientadora: Professora Doutora Luísa Cyrne

Ana Isabel Rodrigues Câmara

Setembro, 2013

## **Agradecimentos**

Em primeiro lugar quero agradecer à professora Doutora Helena Soares, visto que foi uma grande impulsionadora do trabalho aqui apresentado. É admirável a sua experiência científica e por consequência a sua transmissão de conhecimentos permitiram-me de alguma forma crescer tanto a nível pessoal como científico.

Ao longo deste ano meio mostrou-se sempre disponível e teve sempre uma palavra amiga, inclusivamente nos momentos mais críticos, em que as suas palavras de força/incentivo foram muito importantes para que acreditasse em mim e nas minhas capacidades. Muito obrigada Helena pela confiança demonstrada!

Agradeço à professora Doutora Luísa Cyrne porque ao longo deste trabalho mostrou-se sempre muito disponível. Obrigado pelo interesse mostrado no decorrer do trabalho e pelas palavras de força.

Ao grupo de veterinária, por me ter recebido de uma forma tão simpática durante os meses que lá estive. À Doutora Sofia Nolasco, ao Samuel e à Alexandra Tavares pela disponibilidade no laboratório, pelas discussões construtivas e pela boa disposição.

À Alexandra de uma forma especial, por ser uma pessoa com uma disponibilidade inigualável para ajudar os outros. Obrigada por todo o apoio, pela boa disposição, pelas idas ao microscópio, pela transmissão de conhecimentos... Não há palavras!

Aos colegas do grupo de Bioquímica dos Oxidantes e Antioxidantes, Ruben Ramalho, Ana Jerónimo, Filipe Vilas-Boas e Ana Bagullho, pela ajuda na bancada e por se mostrarem sempre disponíveis perante eventuais dúvidas.

Um agradecimento especial à minha família pelo carinho e apoio demonstrado! Aos meus pais, os meus melhores amigos, por todo o apoio incondicional. Perante as dificuldades encontradas, sem o vosso incentivo, nada disto seria possível. Ao meu irmão porque, embora longe, está sempre presente. É uma pessoa cheia de força, a qual tenta sempre transmiti-la para mim!

À Rita, prima, verdadeira amiga e companheira de casa! Obrigada pela amizade e por teres tornado os últimos anos muito mais fáceis.

Aos meus amigos, por me proporcionarem muitos e bons momentos e por estarem sempre dispostos a fazer-me sorrir!

À Mariana, porque além de companheira de laboratório é uma verdadeira amiga. Obrigada pela amizade, pelos bons momentos, pelo companheirismo e por teres sempre uma palavra certa nos momentos mais difíceis.

À Patrícia, à Denise e à Sofia. Obrigada pelos bons momentos de distração proporcionados ao longo deste tempo! Obrigada pelas boas conversas, pelos preciosos conselhos e por serem pessoas tão especiais!

Ao Francisco, pela amizade, pelo apoio e por estar sempre lá quando é preciso!

## Resumo

O centrossoma é um organito essencial nos eucariotas sendo o principal centro organizador de microtúbulos nas células animais. É composto por um par de centríolos e rodeado por uma matriz pericentriolar. Em células em interfase, os centrossomas estão envolvidos na nucleação/organização dos microtúbulos, no posicionamento dos organitos, e.g. o complexo de Golgi, no estabelecimento da polaridade e ainda na migração e adesão, por sua vez em mitose facilitam a formação dos fusos mitóticos.

Estudos realizados pelo nosso grupo, identificaram uma nova proteína humana, que contém o domínio TBCC (TBCCD1), a qual está relacionada com o cofator C da tubulina, o qual participa na via de *folding* da tubulina apresentando uma atividade GAP (*GTPase activating protein*) para a  $\beta$ -tubulina. O TBCCD1 é um componente centrossomal, localizando-se também na zona mediana do fuso, corpo médio e corpos basais/zona de transição de cílios primários e móveis. O silenciamento do TBCCD1 em células RPE-1 provocou um aumento acentuado da distância núcleo-centrossoma, um atraso no ciclo celular, desorganização do complexo de Golgi e baixa eficiência para formar cílios primários. Através de técnicas de análise mutacional identificou-se o domínio mínimo necessário à localização do TBCCD1 no centrossoma, o qual corresponde aos 20 primeiros resíduos de aminoácidos da sua região N-terminal.

O *splicing* alternativo do pré-mRNA é um passo crítico para a expressão de genes sendo a principal fonte para a diversidade de proteínas nos eucariotas superiores. Atualmente pensa-se que ocorre em mais de 90% dos genes humanos. A proteína TBCCD1 humana é codificada por um gene localizado no cromossoma 3 (3q27.3) e apresenta a sua região codificante interrompida por 7 intrões. O presente estudo permitiu verificar que este gene origina três transcritos diferentes pelo processo de *splicing* alternativo. Um destes transcritos resulta do facto que existem dois primeiros exões alternativos, que originam duas proteínas putativas diferindo nos primeiros resíduos de aminoácidos da sua N-terminal. Esta sequência de aminoácidos alternativos corresponde no TBCCD1 ao domínio envolvido na sua localização centrossomal. De facto, as duas novas variantes apresentam uma localização citoplasmática não se localizando no centrossoma.

**Palavras-chave:** TBCCD1, centrossoma, isótipos do TBCCD1, *splicing* alternativo

## Abstract

The centrosome is a key organelle in eukaryotes being the major microtubule-organizing centre in animal cells. It consists of a pair of centrioles surrounded by a pericentriolar matrix. In interphase, by nucleating/organizing the distribution of microtubules, the centrosome is implicated in organelle positioning, such as the Golgi apparatus, cell polarity establishment, cell migration and adhesion, while in mitosis it assists spindle pole formation.

Recent studies in our group have identified a new TBCC domain-containing human protein (TBCCD1) that is related to tubulin cofactor C (TBCC) that participates in the tubulin folding pathway being a GTPase activating protein (GAP) for  $\beta$ -tubulin. TBCCD1 is a centrosomal component that also localizes to the spindle midzone, midbody and basal bodies/transition zone of primary and motile cilia. TBCCD1 silencing in RPE-1 cells causes a marked increase in the centrosome-nucleus distance, a cell cycle delay, GA disorganization and lower efficiency to assemble primary cilia. By performing mutational analysis we already identified the smallest domain required to localize TBCCD1 at the centrosome corresponding to the first 20 amino acids residues of its N-terminus region.

Pré-mRNA alternative splicing is a critical step for gene expression being the major source for protein diversity in higher eukaryotes. Actually it is known to occur in more than 90% of human genes. The human TBCCD1 is encoded by a gene localized at the chromosome 3 (3q27.3) being its coding region interrupted by 7 introns. Interestingly, this gene originates three distinct transcripts by alternative splicing process. One of these transcripts results from the fact that there are two alternative first exons, which generates two putative distinct proteins differing in the first N-terminal amino-acid residues. Striking, this alternative amino-acid sequence corresponds to the domain involved in the centrosomal localization of TBCCD1. In fact the two new variants do not localize at the centrosome.

**Keywords:** TBCCD1, centrosome, TBCCD1 isotypes, alternative splicing

# Índice Geral

Índice de Figuras.....	10
Índice de Tabelas .....	12
Lista de abreviaturas .....	13
Capítulo I - Introdução.....	16
Parte A – O mecanismo de <i>splicing</i> .....	16
A.I.1 <i>Splicing</i> : conceitos básicos .....	16
A.I.1.1 Reconhecimento exão-intrão .....	21
A.I.1.2 Papel da arquitetura do gene no reconhecimento do exão-intrão	22
A.I.1.3 Elementos reguladores do processo de <i>splicing</i> .....	23
A.I.1.4 Papel das estruturas secundárias de RNA no processo de <i>splicing</i>	27
A.I.2 <i>Splicing</i> Alternativo.....	29
Parte B – TBCCD1, uma proteína centrossomal .....	34
B.I.1 Citoesqueleto: Microtúbulos.....	34
B.I.1.1 Nucleação dos microtúbulos .....	37
B.I.2 O centrossoma.....	42
B.I.2.1 Estrutura e função do centrossoma .....	43
B.I.2.2 Ciclo de duplicação dos centrossomas.....	46
B.I.3 Via de <i>foldin</i> g da tubulina.....	47
B.I.3.1 Cofatores da tubulina .....	49
B.I.3.1.1 Cofator C (TBCC) .....	50
B.I.4 Proteínas relacionadas com os cofatores da tubulina.....	52
B.I.4.1 Proteínas que contém o domínio: TBCC, RP2 e TBCCD1 .....	52
B.I.4.1.1 Retinitis pigmentosa protein 2 (RP2).....	53
B.I.4.1.2 TBCC-domain containing protein 1 (TBCCD1) .....	55
Parte C – Objetivos .....	61
Capítulo II - Material e Métodos .....	63
II.1 Reagentes .....	63
II.2 Equipamento .....	64

II.3 Linhas celulares e cDNAs .....	65
II.3.1 Condições de cultura celular .....	65
II.3.1.1 Contagem de células .....	66
II.4 Extração de RNA total de linhas celulares.....	66
II.5 Síntese de cDNA .....	67
II.6 Pesquisa de possíveis transcritos alternativos do gene <i>tbccd1</i> amplificação específica de cDNAs por PCR .....	67
II.7 Estirpes bacterianas e condições de cultura .....	70
II.7.1 Preparação de células de <i>E. coli</i> competentes.....	71
II.8 Clonagem de fragmentos de DNA em vetores de expressão .....	72
II.8.1 Amplificação de sequências DNA por PCR .....	72
II.8.2 Análise de DNA por electroforese em gel de agarose .....	72
II.8.3 Hidrólise dos produtos de PCR com enzimas de restrição .....	73
II.8.4 Ligação do fragmento de DNA ao vector de expressão.....	73
II.8.5 Transformação de células competentes.....	73
II.8.6 Protocolo de <i>Cracking</i> .....	74
II.8.7 Extração e purificação de DNA plasmídico recombinante em pequena escala .....	75
II.8.8 Sequenciação automática de DNA.....	75
II.8.9 Produção de DNA plasmídico em grande escala .....	76
II.9 Clonagem dos cDNAs alternativos do gene <i>tbccd1</i> em vetores de expressão em células de mamíferos.....	76
II.10 Transfeção de células de mamífero com vetores de expressão contendo transcritos alternativos do gene que codifica para a proteína TBCCD1 .....	77
II.11 Microscopia de Imunofluorescência .....	78
Capítulo III – Resultados e Discussão .....	80
III.1 O gene que codifica para a proteína TBCCD1 apresenta um putativo codão de iniciação alternativo .....	80
III.2 O gene <i>tbccd1</i> apresenta um padrão de <i>splicing</i> alternativo que codifica para uma proteína diferente do TBCCD1 canônico .....	93
III.3 Pesquisa do transcrito alternativo do gene <i>tbccd1</i> que codifica para a variante identificada na base de dados por amplificação específica de cDNAs..	97
III.4 Análise da sequência de resíduos de aminoácidos da região N-terminal da variante do TBCCD1 .....	102

III.5 Estudo da localização celular da proteína que codifica para a variante do gene <i>tbccd1</i> por imunofluorescência nas células humanas HEK 293T .....	109
III.6 Estudo da localização celular do transcrito alternativo de maiores dimensões do gene <i>tbccd1</i> por imunofluorescência nas células humanas HEK 293T .....	111
Capítulo IV – Conclusões e Perspetivas Futuras .....	116
Referências bibliográficas: .....	119
Anexos .....	130
1. Sequência nucleotídica da região do gene <i>tbccd1</i> onde ocorre o <i>splicing</i> alternativo dos transcritos descritos ao longo do presente trabalho.....	130
2. pIC111 (6xHis-PreScission-GFP) – vetor construído no vetor pcDNA3.1+ - vetor de expressão em mamífero .....	134

# Índice de Figuras

Figura 1: Motivos conservados nas extremidades do intrão.....	18
Figura 2: Representação esquemática do processo de montagem do spliceossoma e excisão dos intrões.....	20
Figura 3: <i>Splicing code</i> . .....	25
Figura 4: Representação esquemática dos diferentes padrões para o processo de <i>splicing</i> alternativo.....	33
Figura 5: Representação esquemática da dinâmica de microtúbulos.. .....	36
Figura 6: Representação esquemática dos modelos propostos para a nucleação dos microtúbulos a partir do complexo $\gamma$ -TuRC.. .....	40
Figura 7: Representação esquemática do centrosoma de células animais. ....	43
Figura 8: Representação esquemática das várias fases do ciclo de duplicação dos centrosomas. ....	47
Figura 9: Representação esquemática da via de <i>folding</i> da tubulina. ....	48
Figura 10: Proteína humana TBCC. ....	50
Figura 11: Representação esquemática das proteínas que estão relacionadas com o TBCC. ....	53
Figura 12: Representação esquemática da proteína RP2.. .....	54
Figura 13: Proteína TBCCD1 humana. Representação esquemática dos seus domínios funcionais CARP e TBCC. ....	56
Figura 14: Determinação da localização celular do TBCCD1 em células humanas. ....	57
Figura 15: A região N-terminal do TBCCD1 é responsável pela sua localização centrossomal. ....	58
Figura 16: O silenciamento do gene <i>tbccd</i> .....	59
Figura 17: Sequência parcial do cromossoma 3 humano, onde está localizado o gene <i>tbccd1</i> .....	68
Figura 18: Representação gráfica da sequência conservada dos nucleótidos, calculada para os sete nucleótidos que rodeiam o triploto AUG .....	82
Figura 19: Análise de transcritos alternativos do gene <i>tbccd1</i> por PCR em cDNAs produzidos a partir de RNA proveniente de várias linhas celulares humanas.....	84

Figura 20: Análise de transcritos alternativos do gene <i>tbccd1</i> por PCR em cDNAs produzidos a partir de RNA proveniente de diferentes tecidos humanos.....	87
Figura 21: Transcrito alternativo de maiores dimensões do gene <i>tbccd1</i> .....	90
Figura 22: Representação esquemática do possível padrão de <i>splicing</i> alternativo que pode ocorrer no gene <i>tbccd1</i> .....	95
Figura 23: Alinhamento das sequências de resíduos de aminoácidos da proteína TBCCD1 canônico e das sequências de resíduos de aminoácidos previstas para a variante TBCCD1..	96
Figura 24: Representação esquemática do possível padrão de <i>splicing</i> alternativo que pode ocorrer no gene <i>tbccd1</i> .....	98
Figura 25: Amplificação de cDNA em diferentes tecidos e linhas celulares humanas..	99
Figura 26: Amplificação de cDNA produzido a partir de RNAs extraídos das linhas celulares HeLa e MCF7..	101
Figura 27: Electroferograma da sequenciação automática que comprova a existência do transcrito previsto e alternativo ao TBCCD1 canônico..	102
Figura 28: Domínio mínimo da proteína TBCCD1 humana, composto pelos 20 primeiros resíduos de aminoácidos.....	104
Figura 29: Perfil de hidrofobicidade da variante do TBCCD1. ....	105
Figura 30: Previsão da existência de um péptido sinal na proteína variante do TBCCD1.....	106
Figura 31: Análise de imunofluorescência indireta da proteína variante TBCCD1 humana em células de mamífero através da expressão de proteínas de fusão com proteínas fluorescentes.....	110
Figura 32: Análise de imunofluorescência indireta do transcrito alternativo de maiores dimensões do TBCCD1 em células de mamífero. ....	113

## Índice de Tabelas

Tabela 1: Sequências dos <i>primers</i> desenhados para a pesquisa de possíveis transcritos alternativos do gene <i>tbbcd1</i> . .....	69
Tabela 2: Condições das reações de amplificação por PCR, usadas na pesquisa de transcritos alternativos do gene <i>tbbcd1</i> . .....	69
Tabela 3: Sequências dos <i>primers</i> utilizados na clonagem dos transcritos alternativos do gene <i>tbbcd1</i> . .....	77
Tabela 4: Condições das reações de amplificação por PCR, usadas na estratégia de clonagem. ....	77
Tabela 5: Anticorpos primários e secundários usados na análise de microscopia de imunofluorescência. ....	79
Tabela 6: Representação do contexto nucleotídico considerado ótimo para um AUG iniciador e do contexto nucleotídico que rodeia o AUG descrito para o <i>tbbcd1</i> e o AUG putativo analisado neste estudo. ....	82
Tabela 7: Sequência de resíduos de aminoácidos que constituem a ORF gerada entre os dois AUGs presentes na 5' UTR. ....	93

## **Lista de abreviaturas**

**5'UTR** - 5' *Untranslated region*

**Arl2** - *ADP ribosylation factor-like protein 2*

**Arl3** - *ADP ribosylation factor-like protein 3*

**ATP** - 5'-trifosfato de adenosina

**ATPase** - enzima que catalisa a hidrólise do ATP

**BPS** - Branch point sequence

**BSA** - albumina do soro de bovino

**CAP** - *cyclase-associated proteins*

**CARP** - *domain in CAPs*

**C-Coil** - domínio *coiled-coil*

**CCT** - *cytosolic-chaperonin-containing TCP1*

**cDNA** - DNA obtido através de moléculas de RNA através de transcrição reversa

**C-terminal** - extremidade de uma cadeia polipeptídica cujo último resíduo de aminoácido apresenta um grupo carboxilo

**DAPI** - 4',6-diamidino-2-fenilindole

**DMEM** - meio de cultura *Dulbecco's Modified Eagle*

**DMSO** - dimetilsulfóxido

**DNA** - ácido desoxirribonucleico

**dNTPs** - 5'-trifosfato de desoxinucleótido

**dscam** - down syndrome cell-adhesion molecule

**EDTA** - ácido etileno diamino tetra-acético

**FBS** - soro fetal bovino

**GAP** - *GTPase activating protein*

**GCPs** - *γ-tubulin complex proteins*

**GCP-WD** - *γ-tubulin complex proteins-WD40 domain*

**GDP** - 5'-difosfato de guanina

**GFP** - proteína fluorescente verde

**GTP** - 5'-trifosfato de guanina

**hnRNP** - *heteronuclear ribonuclear protein*

**ISE** - *Intronic splicing enhancer*

**ISS** - *Intronic splicing silencer*

**kb** - 10<sup>3</sup> pares de bases

**LB** - meio Luria Bertani ou *Lysogeny Broth*

**MAPs** - proteínas associadas a microtúbulos

**mRNA** - RNA mensageiro

**MTOCs** - centros organizadores de microtúbulos

**NDK1** - *nucleoside diphosphate kinase 1*

**NDPk** - *nucleoside diphosphate kinase*

**N-terminal** - extremidade de uma cadeia polipeptídica cujo primeiro resíduo de aminoácido apresenta um grupo amina livre

**ORF** - *Open Reading frame*

**pb** - pares de bases

**PBS** - solução tampão fosfato-salino

**PCM** - material pericentriolar

**PCR** - *polymerase chain reaction*

**PPT** - *Polypyrimidine tract*

**Pré-mRNA** - Pré RNA mensageiro

**RNA** - ácido ribonucleico

**RNAi** - RNA de interferência

**RP2** - *Retinitis pigmentosa protein 2*

**RT-PCR** - *reverse transcription polymerase chain reaction*

**SF1** - *Splicing factor 1*

**siRNA** - pequeno RNA interferente (de *short interfering RNA*)

**snRNP** - small nuclear Ribonucleoprotein

**SR** - *Serine-Arginine rich*

**ss** - *splice site*

**TBCA** - cofactor da tubulina A (de *tubulin cofactor A*)

**TBCB** - cofactor da tubulina B (de *tubulin cofactor B*)

**TBCC** - cofactor da tubulina C (de *tubulin cofactor C*)

**TBCCD1** - *TBCC-domain containing protein 1*

**TBCD** - cofactor da tubulina D (de *tubulin cofactor D*)

**TBCE** - cofactor da tubulina E (de *tubulin cofactor E*)

**Tween 20** - polioxietileno(sorbitan)monolaurato

**uORF** - *upstream open reading frame*

**$\gamma$ -TuRC** -  *$\gamma$ -tubulin ring complex*

**$\gamma$ -TuSC** -  *$\gamma$ -tubulin small complex*

# Capítulo I - Introdução

## Parte A – O mecanismo de *splicing*

### A.I.1 *Splicing*: conceitos básicos

O controlo da expressão genética é um processo biológico essencial a todos os organismos, sendo crucial na definição do fenótipo de uma célula ou organismo. Este controlo é conseguido através da interação de proteínas reguladoras com motivos de DNA específicos, nas regiões controlo dos genes que regulam. Após a ligação ao DNA e através de interações específicas proteína-proteína, estas proteínas reguladoras transmitem sinais à maquinaria basal de transcrição, contendo as respetivas RNA polimerases e por isso resultando em níveis particulares da expressão de genes (Yilmaz and Grotewold 2010).

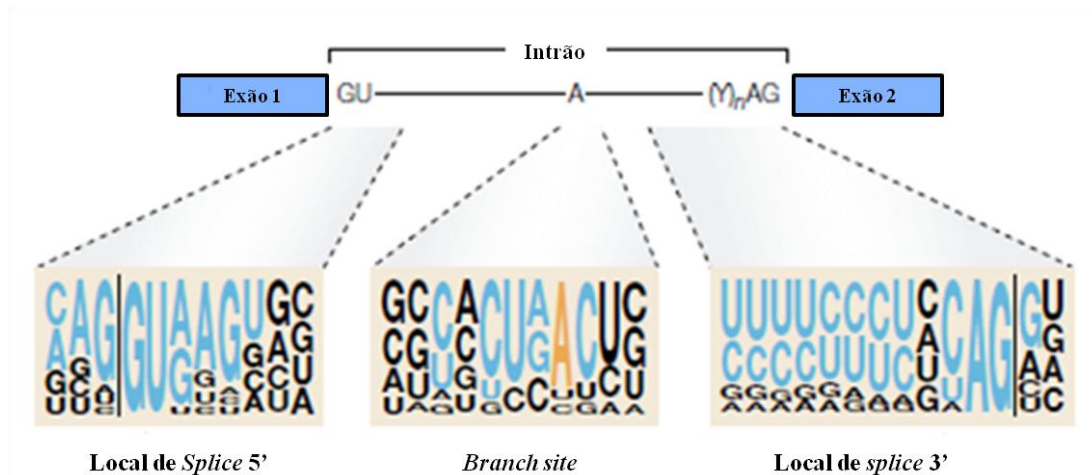
Sabe-se que para uma correta execução de processos biológicos, entre os quais, o desenvolvimento, a proliferação, a apoptose e o envelhecimento, é necessário um preciso e cuidadoso conjunto de passos bem coordenados, que dependem da adequada expressão de genes, no espaço e no tempo (Maston, Evans et al. 2006).

Uma eficaz expressão dos genes eucarióticos é um processo que compreende vários passos. Neles incluem-se a transcrição, o *splicing*, a poliadenilação do transcrito primário, o transporte do RNA mensageiro (mRNA) para o citoplasma e por fim, a tradução deste em proteína (Singer and Green 1997; Misteli and Spector 1998)

As regiões codificantes da maioria dos genes eucariotas são interrompidas por vários intrões, que são removidos pelo mecanismo de *splicing*, de forma a conseguir uma correta expressão do gene, e conseqüentemente, uma falta de rigor neste processo tem a capacidade de causar ruturas ao nível das funções celulares (Pleiss, Whitworth et al. 2007). No entanto, para que haja um aumento da diversidade proteómica é também necessário uma variação no mesmo. A regulação do *splicing* do pré-mRNA pode ocorrer em várias etapas e conta com a coordenação de múltiplos fatores. Este processo, pelo qual os intrões são removidos a partir do mRNA precursor, ocorre

através de uma série de interações de diferentes naturezas coordenadas pelo spliceossoma, sendo elas RNA-RNA, RNA-proteína e proteína-proteína (Hastings and Krainer 2001).

O spliceossoma é um complexo macromolecular dinâmico, o qual é altamente conservado entre os eucariotas. Os principais componentes do spliceossoma incluem cinco pequenos RNAs (snRNAs) – U1, U2, U4, U5 e U6 – sendo que cada um funciona como um complexo RNA-proteína, designado como pequeno complexo nuclear ribonucleoproteico (snRNP - *small nuclear ribonucleoprotein complexes*). Além destes, requer ainda um grande número de proteínas, muitas das quais não são snRNP (Hastings and Krainer 2001; Pleiss, Whitworth et al. 2007; Crabb, Lam et al. 2010). Os snRNAs e os componentes proteicos do spliceossoma interagem com sequências definidas nos limites exão-intrão para direcionar a excisão dos intrões e consequente junção dos exões ao nível dos locais de *splice*, de forma dar origem ao mRNA maduro (Goren, Ram et al. 2006). Assim, para que o *splicing* ocorra de forma eficiente, é necessário que existam na maioria dos intrões quatro locais, entre eles o local de *splice* 5' (5'ss – 5' *splice site*), que marca a junção exão-intrão na extremidade 5' do intrão (inclui o dinucleótido conservado GU na extremidade do intrão abrangido por uma maior e menos conservada sequência consenso). Na outra extremidade do intrão, o local de *splice* 3' (3'ss – 3' *splice site*) apresenta três sequências conservadas: a *branch point sequence* (BPS), seguida do *polypyrimidine tract* (PPT) e do dinucleótido conservado AG terminal na extremidade 3' do intrão (Hastings and Krainer 2001; Black 2003) (Figura 1 e 2A).

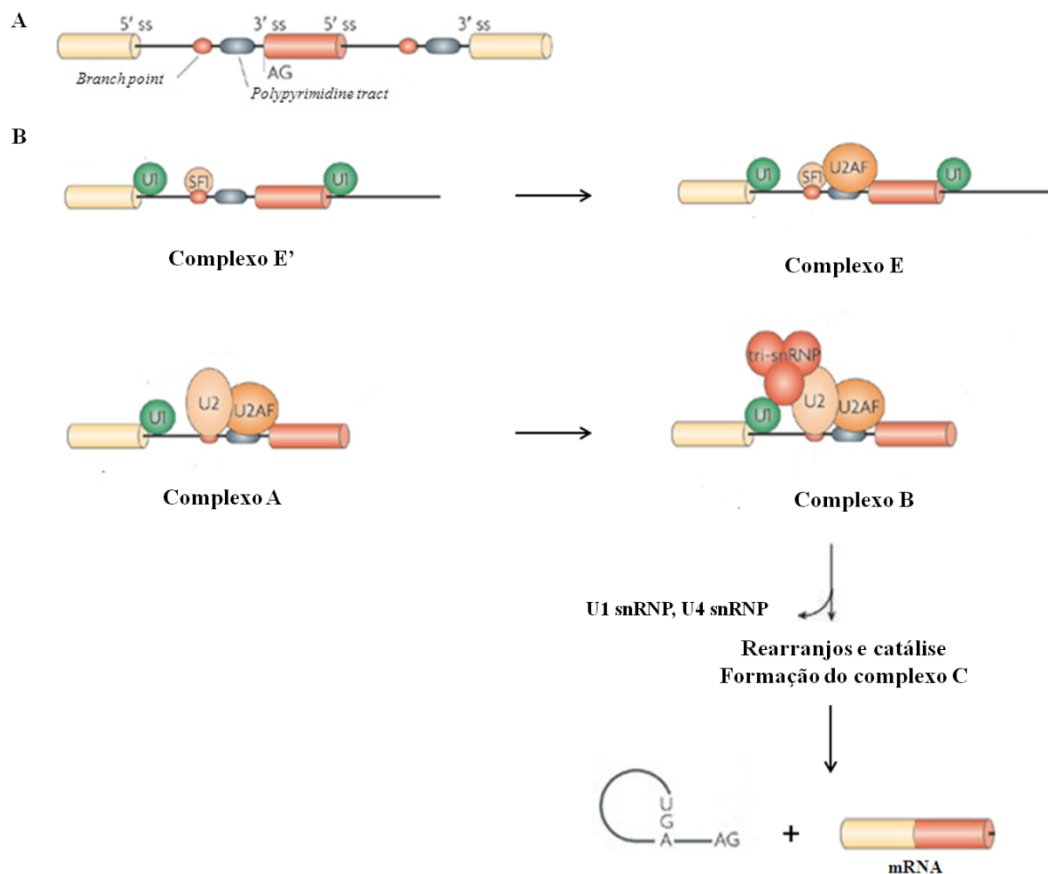


**Figura 1: Motivos conservados nas extremidades do intrão, ou relativamente perto.** Na ilustração acima, estão representados os dinucleótidos conservados GU e AG nas extremidades do intrão, o *polypyrimidine tract* (Y)<sub>n</sub> e o resíduo A que integra o *branch point*, sendo que as sequências motivo que rodeiam estes nucleótidos conservados estão representadas abaixo. Para cada sequência motivo, o tamanho de um nucleótido numa dada posição é proporcional à frequência deste naquela posição, num alinhamento de sequências conservadas a partir de intrões humanos. Os nucleótidos que fazem parte dos motivos consenso clássicos estão ilustrados a azul, exceto o *branch point* (A), que está a laranja (adaptado de Cartegni, Chew et al. 2002).

Os mecanismos básicos da montagem do spliceossoma no pré-mRNA encontram-se bem estabelecidos e envolvem o reconhecimento dos elementos consenso em ambas as extremidades de um intrão (Smith and Valcarcel 2000).

Sendo o *splicing* um processo ordenado é normal que existam diferentes intermediários, assim a excisão do intrão inicia-se com o reconhecimento do local de *splice* 5' pelo U1 snRNP e a ligação da proteína *splicing factor 1* (SF1) à região BPS numa forma independente de ATP para formar o complexo E' (Figura 2B). Este complexo pode ser convertido no complexo E, através do recrutamento do fator auxiliar heterodimérico U2 (U2AF), que compreende as subunidades U2AF65 e U2AF35, ligando-se respetivamente ao PPT e à sequência 3' na região terminal do intrão (3'ss), AG, este complexo compreende ainda, pelo menos, um membro das proteínas SR (arginina-serina). De uma forma dependente de ATP, ao nível da *branch point*, dá-se a substituição do SF1 pelo U2 snRNP, formando-se o complexo A (pré-spliceossoma). Além disso, o recrutamento do tri-snRNP U4/U5-U6 conduz à

formação do complexo B, o qual contém todas as subunidades do spliceossoma que realizam o *splicing* do pré-mRNA. Seguidamente ocorrem extensas remodelações e alterações conformacionais, incluindo a perda dos U1 e U4 snRNPs resultando finalmente, na formação do complexo C, que consiste no spliceossoma cataliticamente ativo (Hastings and Krainer 2001; Chen and Manley 2009; Wang, Mueller et al. 2011). Uma vez formado, este complexo vai promover a excisão dos intrões, que ocorre através de duas reações de transesterificação, cada uma delas envolvendo um ataque nucleofílico nas ligações fosfodiéster na extremidade terminal do intrão. Na primeira reação, o grupo 2'-hidroxilo (geralmente de uma adenosina) na BPS “ataca” o fosfato no local de *splice 5'*, o que leva à clivagem do exão 5' a partir do intrão e a correta ligação da extremidade 5' do intrão ao 2'-hidroxilo na BPS. Este passo produz duas reações intermédias, um exão 5' independente e um intrão-exão 3' com formação do *lariat* (apresenta estrutura em laço), contendo o nucleótido A ramificado na BPS. A segunda reação de transesterificação consiste no “ataque” do fosfato na extremidade 3' do intrão pelo grupo 3'-hidroxilo do exão individual. Este passo liga os dois exões e liberta o intrão, ainda em forma de *lariat* (Black 2003; Chen and Manley 2009). Após a ligação dos exões, ocorre a desmontagem do spliceossoma que vai libertar os seus componentes, ficando estes, assim disponíveis, para a formação de um novo spliceossoma, processo denominado por “ciclo de *splicing*” (Jurica and Moore 2003).



**Figura 2: Representação esquemática do processo de montagem do spliceossoma e excisão dos intrões.** (A) Locais de *splicing* situados próximos das extremidades do intrão (local de *splice* 5'; *Branch point*; *Polypyrimidine tract*; local de *splice* 3'). (B) Remoção dos intrões através do spliceossoma, que começa a sua montagem em ambas as extremidades do intrão. Inicia-se pelo reconhecimento do local de *splice* 5' pelo U1 snRNP, através de combinações RNA-RNA e RNA-proteína (Complexo E'); o U2AF liga-se ao local de *splice* 3' (complexo E); o U2AF, por sua vez recruta U2 snRNP que emparelha com a *branch point* (complexo A). Seguidamente, de uma forma dependente de ATP ocorre a ligação do U4/U5-U6 tri-snRNP (complexo B). Ocorrem rearranjos que levam à perda do U1 e U4 snRNP, que origina o spliceossoma cataliticamente ativo (complexo C). Este por sua vez realiza duas reações de transesterificação nos locais de *splice*, que resultam na ligação dos exões e excisão do intrão numa conformação em laço (adaptado de Chen and Manley 2009).

### A.I.1.1 Reconhecimento exão-intrão

Devido à degeneração dos motivos consenso nos locais de *splice* 5' e 3', um dos principais desafios do spliceossoma consiste em distinguir os locais de *splice* “*bona fide*” dos inúmeros segmentos que se assemelham a exões (pseudo-locais) encontrados num transcrito de pré-mRNA, juntar os exões contíguos sem saltos inadvertidos e regular apropriadamente o *splicing* alternativo (que será explicado em maior detalhe na secção A.I.2) atendendo às necessidades fisiológicas das células e tecidos (Wang and Cooper 2007). A eficiência com que o spliceossoma se liga a um exão é determinada por um equilíbrio de várias características, incluindo a força de um local de *splice* (ou seja, a sua concordância com sequências consenso dos locais de *splice*), o tamanho do exão e a presença de elementos auxiliares *cis-acting*, (situados aproximadamente dentro de 50 nucleótidos a partir dos limites do exão). Os elementos auxiliares *cis-acting* comunicam entre si, direcionam o spliceossoma para os nucleótidos corretos e servem ainda como locais de ligação, recrutando fatores auxiliares de *splicing* (positivos e negativos) que regulam o *splicing* de forma equilibrada. Estes elementos compõem, por isso, o que é hoje conhecido como “*splicing code*”, que parece ser particularmente denso, dentro e ao redor dos exões (Ladd and Cooper 2002; Wang and Cooper 2007).

Os exões com um tamanho ideal (normalmente 50 a 300 nucleótidos) e com sequências consenso fortes são eficientemente reconhecidos pela maquinaria de *splicing* e são incluídos constitutivamente no transcrito, ao passo que exões em condições menos ótimas necessitam de elementos auxiliares para o reconhecimento. No entanto, tem-se verificado que muitos exões constitutivos também usam elementos auxiliares para assegurar o seu reconhecimento. Mais ainda, sabe-se que muitos dos locais de *splice* normalmente reconhecidos pela maquinaria de *splicing* apresentam sequências altamente divergentes das sequências consenso “ideais”. Seria então de esperar que estes locais de *splice* fossem pouco reconhecidos sob qualquer condição, o que não se verifica (Ladd and Cooper 2002; De Conti, Baralle et al. 2013).

Além dos exões que são processados constitutivamente, existem muitos outros que são processados alternativamente. Resumidamente, no *splicing* constitutivo todos os exões de um dado pré-mRNA vão estar presentes e na mesma ordem no mRNA produzido a partir dele, por outro lado no *splicing* alternativo várias formas do mRNA são geradas a partir de um único gene (Nilsen and Graveley 2010; Will and Luhrmann 2011). Este último, além da importância na regulação da expressão genética assume um papel fundamental na diversidade do proteoma (Zheng, Fu et al. 2005).

### **A.I.1.2 Papel da arquitetura do gene no reconhecimento do exão-intrão**

Normalmente, nos transcritos humanos, os exões são por norma pequenos, quando comparados aos intrões que são consideravelmente maiores (podem ir das centenas aos milhares de pares de bases) e em média representam mais de 90% do transcrito primário (Wang and Burge 2008). Posto isto, é de esperar que a maquinaria que envolve o *splicing* encontre uma grande dificuldade em localizar os locais de *splice* através de intrões tão longos. Sabe-se que, quando os exões são pequenos e os intrões são longos, é mais provável que a maquinaria de *splicing* se organize através de um exão do que através de um intrão, visto que parece mais fácil unir exões pequenos em vez de intrões de grandes unidades. Estes factos levaram à explicação através de dois modelos, o reconhecimento dos locais de *splice* pelo spliceossoma, sendo eles (1) modelo *exão definition* (definição de exão) e (2) *intron definition* (definição de intrão). No primeiro, a maquinaria spliceossomal irá procurar e por sua vez ligar-se aos locais de *splice* na extremidade de um exão interno, ou seja, os locais *splice* são inicialmente reconhecidos e emparelhados em conjunto através do exão. A montagem do spliceossoma prossegue através de unidades de exões subsequentemente emparelhadas, sendo que uma mutação numa das extremidades deste leva ao salto (*skipping*) do exão. Quanto ao modelo alternativo *intron definition*, deriva do fato de alguns transcritos (especialmente nos invertebrados) serem muito curtos em relação aos exões e por isso a maquinaria de *splicing* tende a emparelhar diretamente nos locais de *splice* através do intrão em detrimento dos exões. Neste caso, sendo os intrões as unidades de reconhecimento pelo

spliceossoma, o qual identifica os locais de *splice* 5' e 3' que flanqueiam o intrão, uma mutação no local de *splice* 5' de um pequeno intrão leva à retenção do intrão em alternativa ao salto do exão, por sua vez o acréscimo de intrões curtos inibe o processo de *splicing* a eles inerente, *in vitro* e *in vivo*. Desta forma, este tipo de acontecimentos origina diferentes padrões de *splicing* alternativo (Lim and Burge 2001; Catania and Lynch 2013).

Verifica-se portanto, que o *splicing* do pré-mRNA é um aspeto fundamental da expressão de genes sendo surpreendente a precisão e a complexidade da remoção dos intrões a ele inerente (Faustino and Cooper 2003).

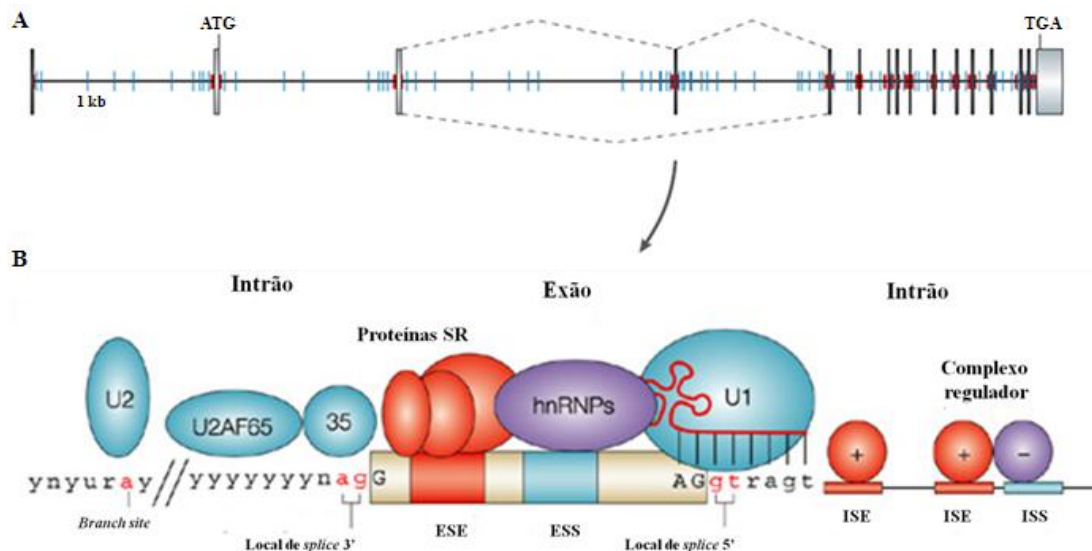
### **A.I.1.3 Elementos reguladores do processo de *splicing***

Tanto o *splicing* constitutivo como o *splicing* alternativo requerem a montagem da maquinaria basal de *splicing* nos complexos do spliceossoma que se liga às sequências consenso presentes em todos os limites intrão-exão (locais de *splice* 5' e 3') (Ladd and Cooper 2002).

Vários estudos genéticos e bioquímicos revelam que são necessários rearranjos conformacionais em ambos os componentes do RNA e das proteínas, essenciais ao spliceossoma, de forma a processar com fidelidade os transcritos do pré-RNA mensageiro (pré-mRNA) (Pleiss, Whitworth et al. 2007). Nos eucariotas superiores, nos quais os genes são frequentemente interrompidos por vários intrões, sabe-se que o spliceossoma pode utilizar sequências específicas presentes nos transcritos individuais para regular tanto quantitativamente como qualitativamente aspetos da expressão do gene. No entanto, sabe-se que a eficiência deste processo depende da complementaridade entre o local de *splice* 5' e a extremidade 5' do U1 snRNA, bem como da presença de elementos reguladores de *splicing* (*SREs* de *Splicing Regulatory Elements*) nas proximidades. Verifica-se por isso, que os principais locais de *splice* por si só não dão informação suficiente e conseqüentemente existem informações adicionais em pequenos elementos *cis*, degenerados, encontrados tanto nos exões como nos intrões flanqueantes (Pleiss, Whitworth et al. 2007; Ke and Chasin 2011; Erkelenz, Mueller et al. 2013). No geral, os elementos auxiliares que

regulam o uso dos locais de *splice* alternativos apresentam várias características em comum, entre elas, o tamanho, geralmente são pequenos, variáveis na sequência, fracos individualmente e presentes em várias cópias. Os elementos auxiliares são no geral de cadeia única, embora estruturas secundárias têm sido envolvidas na função de alguns elementos. Os elementos auxiliares são, frequentemente, conservados entre as espécies e possivelmente entre os genes regulados de forma idêntica. No entanto como contêm motivos com sequências degeneradas, tornam-se difíceis de identificar. Estes podem ser exônicos ou intrônicos, sendo que quando são intrônicos podem estar a montante, jusante ou flanquear ambos os lados do exão regulado. Os elementos intrônicos podem ainda ser proximais ou distais, embora normalmente se encontrem localizados perto do exão (Ladd and Cooper 2002).

Assim, dependendo da sua posição e função, os elementos *cis-acting* são divididos em quatro categorias, sendo elas *Exonic Splicing Enhancers* (ESEs) ou *Exonic Splicing Silencers* (ESSs), se a partir de uma localização exónica têm como função promover ou inibir, respetivamente, a inclusão do exão onde se encontram. Alternativamente, podem ser *Intronic Splicing Enhancers* (ISEs) ou *Intronic Splicing Silencers* (ISSs), se ativam ou inibem o uso dos locais de *splice* adjacentes ou exões a partir da localização intrónica (Figura 3A e B) (Chasin 2007; Wang and Burge 2008). Além dos componentes *cis*, a regulação do *splicing* envolve elementos *trans*, estes são fatores celulares (RNA ou proteínas) que são recrutados pelos elementos reguladores do *splicing* (SREs) e que, dependente do contexto, ativam ou reprimem o reconhecimento dos locais de *splice* ou a montagem do spliceossoma, por diferentes mecanismos (Matlin, Clark et al. 2005; Wang and Burge 2008).



**Figura 3: Splicing code.** (A) Representação esquemática do pré-mRNA, tal como o spliceossoma o encontra. As linhas vermelhas indicam os locais de *splice* consenso nos limites intrão-exão. A azul encontram-se representados os elementos auxiliares *cis-acting* intrônicos que compõem o *splicing code*. (B) Representação esquemática dos vários elementos que regulam o processo de *splicing*. Para além dos clássicos locais de *splicing*, estão ainda representados os elementos auxiliares *cis-acting* (ESEs e ESEs – *Exonic splicing enhancers* e *silencers*; ISEs e ISSs – *Intronic splicing enhancers* e *silencers*), assim como fatores *trans-acting*, as proteínas SR que interagem com elementos *cis-acting enhancers* e as hnRNP, que interagem com os *cis-acting silencers* (adaptado de Pagani and Baralle 2004; Wang and Cooper 2007).

Geralmente, os ESEs são reconhecidos por membros da família de proteínas SR. (Chen and Manley 2009; Ke and Chasin 2011). Todas as proteínas SR têm uma organização modular e contêm na sua N-terminal um domínio de reconhecimento do RNA (RRM – *RNA recognition motif*), que medeia a ligação ao pré-mRNA. Estas proteínas contêm ainda um domínio rico em arginina/serina na C-terminal, necessário para as interações proteína-proteína. Assim, quando ligadas aos ESEs, as proteínas SR recrutam e facilitam a ligação de diferentes componentes do spliceossoma às proximidades do intrão (Graveley 2000; Buvoli, Buvoli et al. 2007). Além disso, estas promovem a montagem do spliceossoma, contactando diretamente com o mRNA, bem como a seleção dos locais de *splice*, trazendo os exões para as proximidades. Mesmo que a maioria destes SREs tenha sido encontrada em intrões dependentes de um ativador contendo um local de *splice* fraco, a sua descoberta nos

exões constitutivos em conjunto com o facto de as sequências que codificam proteínas serem enriquecidas em potenciais locais de ligação SR, sugere que os ESEs podem estar presentes na maioria dos exões (Buvoli, Buvoli et al. 2007). Verifica-se por isso, que as proteínas SR associam-se aos ESEs para promoverem tanto o *splicing* constitutivo como o alternativo através da formação de redes de interação, entre si bem como com as proteínas associadas às snRNP e proteínas relacionadas com as SR (Blencowe 2000). Por exemplo, as proteínas SR associam-se a SREs exónicos para facilitar o recrutamento do U1 snRNP para o local de *splice 5'* através do seu domínio RS ou através de interações do domínio de ligação ao RNA com a proteína 70K específica do U1 snRNP (composto pelo U1 snRNA, proteínas específicas U1, sendo uma delas a U1 70K e proteínas spliceossomais que também estão presentes nos complexos U2, U5 e U6 snRNP). No entanto, em alguns casos as proteínas SR parecem interferir com a seleção do exão quando ligadas às posições intrónicas, possivelmente devido ao impedimento estérico (Erkelenz, Mueller et al. 2013).

Por outro lado, as ISSs e ESSs, são reconhecidas por ribonucleoproteínas heterogéneas nucleares (hnRNP - *heterogeneous nuclear ribonuclearprotein*), as quais têm um ou mais domínios de ligação ao RNA, domínios de interação proteína-proteína e às vezes domínios inibidores do *splicing*, tais como motivos ricos em glicina (Wang and Burge 2008; Chen and Manley 2009). A repressão mediada pelas hnRNP a partir de posições exónicas tem sido foco de estudo. No entanto, em alguns casos provou-se que as hnRNP intrónicas podem ativar o *splicing* (Erkelenz, Mueller et al. 2013).

As hnRNP funcionam através de um largo número de mecanismos, por exemplo a PTB (*polypyrimidine tract binding protein* ou hnRNP I) pode bloquear interações essenciais entre os U1 e U2 snRNPs. Por sua vez, a hnRNP A1 (fator de *splicing* constitutivo) pode ligar-se a ambos os lados dos exões formando um *loop*, o qual contém o local de *splice 5'* e desta forma mediar a repressão do mecanismo de *splicing* (Wang and Burge 2008).

De um ponto de vista evolutivo, parece que os locais de ligação intrónicos das hnRNP têm desempenhado um papel importante no panorama dos SREs de

mamíferos. Em conjunto com mapas de *splicing* do RNA definidos recentemente (correlacionam perfis de interações RNA-proteína e consequente efeito na ativação do *splicing*), estes aspetos sugerem que a atividade das proteínas reguladoras de *splicing* podem ser altamente dependentes da posição (Wang and Burge 2008; Wang, Mueller et al. 2011; Erkelenz, Mueller et al. 2013).

Como já referido, as proteínas SR podem ligar-se a um ESE favorecendo o *splicing* pelo recrutamento de fatores de *splicing*, mas também podem antagonizar os repressores como a hnRNP A1. Assim, os efeitos das proteínas SR no *splicing* alternativo podem ser antagonizados pelo fator de *splicing* constitutivo hnRNP A1. Em alguns casos, as hnRNP A1 ligam-se cooperativamente e podem antagonizar o *splicing* pois competem diretamente pela ligação das proteínas SR aos ESEs ou do U2 snRNP ao *branch point*, próximo de um ISS (Ladd and Cooper 2002; Singh and Valcarcel 2005).

De uma forma geral, estes dois grandes grupos de proteínas, as SR e hnRNP, são capazes de modular a atividade de *splicing*, permitindo ao spliceossoma gerar várias proteínas distintas a partir de um único *locus* genómico e consequentemente, aumentar significativamente a diversidade proteómica (Kim, Goren et al. 2008).

Quanto aos ISEs, estes não estão tão bem estabelecidos como os outros três tipos de elementos, mas estudos recentes indicam, que várias proteínas tais como hnRNP F, hnRNP H, antigénio 1 neuro-oncológico ventral (NOVA1), NOVA2, FOX1 e FOX2 (também conhecidos por RBM9) são reconhecidas por estes e desta forma estimulam o processo de *splicing* (Chen and Manley 2009).

#### **A.I.1.4 Papel das estruturas secundárias de RNA no processo de *splicing***

Embora o pré-mRNA seja normalmente representado de forma linear, existe uma grande probabilidade de o RNA de cadeia simples adquirir estruturas secundárias

locais e interações terciárias que podem ser constituídas por centenas de nucleótidos (Shepard and Hertel 2008).

A maioria dos fatores proteicos de ligação ao RNA interagem com a sua cadeia simples, no entanto as sequências alvo de ligação destes fatores proteicos encontram-se normalmente presentes na forma de *loop* (Pagani and Baralle 2004). Para além da força de um local de *splice*, dos reguladores de *splicing*, da arquitetura exão-intrão e do processo de transcrição pela RNA polimerase II, estas estruturas secundárias do mRNA parecem assumir um papel importante, tanto no *splicing* alternativo como no *splicing* constitutivo. Sabe-se que influenciam um número de passos do processamento, onde se inclui o *splicing* alternativo, uma vez que dependendo da estabilidade termodinâmica, podem permanecer o tempo suficiente para interferir ou modular o reconhecimento dos locais de *splice* (ou motivos regulatórios).

As estruturas secundárias do RNA podem, em princípio, inibir ou ativar a montagem do spliceossoma, porque o reconhecimento dos locais de *splice* (*branch sites*, ativadores e repressores) depende geralmente da interação entre fatores proteicos, numa determinada zona do pré-mRNA de cadeia simples e, conseqüentemente, a formação de cadeias duplas pode evitar o reconhecimento e, portanto, a ligação dos mesmos (Maris, Dominguez et al. 2005; Hiller, Zhang et al. 2007; Shepard and Hertel 2008). De uma forma geral, as estruturas de RNA podem interferir com a montagem do spliceossoma ou com o reconhecimento dos locais de *splice*, caso ocultem os mesmos, ou por sua vez ocultem sequências ativadoras, ambos dentro de hélices estáveis. Por outro lado, as estruturas locais de RNA podem também promover a montagem do spliceossoma, mascarando locais de ligação repressores do *splicing* (Shepard and Hertel 2008; Jin, Yang et al. 2011).

As alterações nas estruturas secundárias do mRNA têm um papel essencial em muitos processos biológicos. O frequente envolvimento de elementos das estruturas de RNA com o *splicing* alternativo mostra que a formação destas é um importante mecanismo que regula tanto a expressão dos genes como o desenvolvimento de algumas doenças (Shepard and Hertel 2008). Por exemplo, no exão 10 do gene *tau*

existe uma região *stem-loop* que regula a acessibilidade do local de *splice 5'* pelo U1 snRNP e conseqüentemente leva a alterações no processo de *splicing*. As mutações que quebram este elemento estrutural resultam num aumento da acessibilidade do local de *splice 5'* e como consequência o reconhecimento do exão é aumentado, sendo que o resultado destas alterações tem sido relacionado com a demência frontotemporal e parkinsonismo (Buratti and Baralle 2004; Donahue, Muratore et al. 2006).

### **A.I.2 *Splicing* Alternativo**

O *splicing* alternativo é o principal mecanismo que modula a expressão de genes celulares e virais e permite o aumento da capacidade de codificação de um único gene, proporcionando a síntese de vários isótipos de proteínas estruturalmente e funcionalmente diferentes (Caceres and Kornblihtt 2002). A partir do *splicing* alternativo, um mRNA precursor pode dar origem a mais do que um transcrito de mRNA devido às variações na incorporação das regiões de codificação. O *splicing* alternativo das regiões não codificantes (UTRs de *untranslated regions*) pode também determinar a localização e estabilidade do mRNA, bem como a eficiência da tradução (Kim, Goren et al. 2008).

Além das bases de dados, estudos genômicos de larga escala permitiram saber que o *splicing* alternativo pode ocorrer em mais do que 90% dos genes humanos, sugerindo um papel de grande importância na produção da complexidade funcional no genoma humano, visto que este é limitado a apenas 31,000 – 39,000 genes (Cartegni, Chew et al. 2002; Luco, Allo et al. 2011).

O *splicing* alternativo apresenta uma forte especificidade para um tecido em particular ou para um estadio do desenvolvimento embrionário, modulando as características funcionais dos isótipos de proteínas em tecidos específicos. O mRNA e os isótipos de proteínas produzidas pelo processamento alternativo podem diferir na estrutura, função, localização e outras propriedades (Wang, Sandberg et al. 2008). Estima-se também que aproximadamente 15% das mutações que causam doença nos humanos envolvem uma desregulação do *splicing* alternativo e erros no

processamento do mRNA têm sido associados com o cancro e outras doenças humanas (Xu, Modrek et al. 2002). Em média um gene humano pode originar dois a três transcritos. No entanto, devido ao fato de, um único transcrito primário poder conter várias regiões que sofrem *splicing* alternativo, o resultado de efeitos combinatórios na seleção dos diferentes locais de *splice* pode ser muito acentuado e serem comuns os genes que codificam para entre dez a centenas de isótipos diferentes (Cartegni, Chew et al. 2002).

Um ótimo exemplo desta diversidade proteômica que é atribuída ao *splicing* alternativo é o gene *dscam* (*Down syndrome cell-adhesion molecule*) da *Drosophila melanogaster*, que codifica para uma proteína de superfície celular envolvida na conectividade neuronal. O pré-mRNA do gene *dscam* é processado alternativamente e pode originar 38 016 isótipos de proteínas diferentes, duas vezes mais o número de genes que completam o genoma da *Drosophila* (Schmucker, Clemens et al. 2000; Caceres and Kornblihtt 2002).

O *splicing* alternativo tem sido descrito numa grande variedade de grupos filogenéticos eucariotas incluindo as plantas, apicomplexas, diatomáceas, amebas, animais e fungos. No entanto, é incerto e difícil de estimar se este processo apareceu independentemente nos diferentes grupos, como sugerem alguns autores, ou se é um antigo processo evolucionário e provavelmente teve origem num ancestral comum dos eucariotas (Irimia, Rukov et al. 2007b; Kelemen, Convertini et al. 2013). Estudos indicaram que a maquinaria spliceossomal é largamente conservada ao longo dos eucariotas e ainda que os eucariotas ancestrais contêm genes ricos em intrões com os limites consenso fracos no local de *splice* 5', característica que está ligada à presença de *splicing* alternativo nos organismos recentes. Estes dados favorecem a hipótese de que algumas formas de *splicing* alternativo apareceram relativamente cedo num ancestral comum unicelular das plantas, animais e fungos, e bastante cedo na evolução dos eucariotas existentes (Irimia, Penny et al. 2007a; Irimia, Rukov et al. 2007b). A função dos exões individuais pode ser determinada por comparação evolucionária. Todas as comparações dependem da hipótese que os exões alternativos, funcionalmente importantes serão conservados na evolução, enquanto

os exões não funcionais serão eliminados por seleção negativa (Kelemen, Convertini et al. 2013).

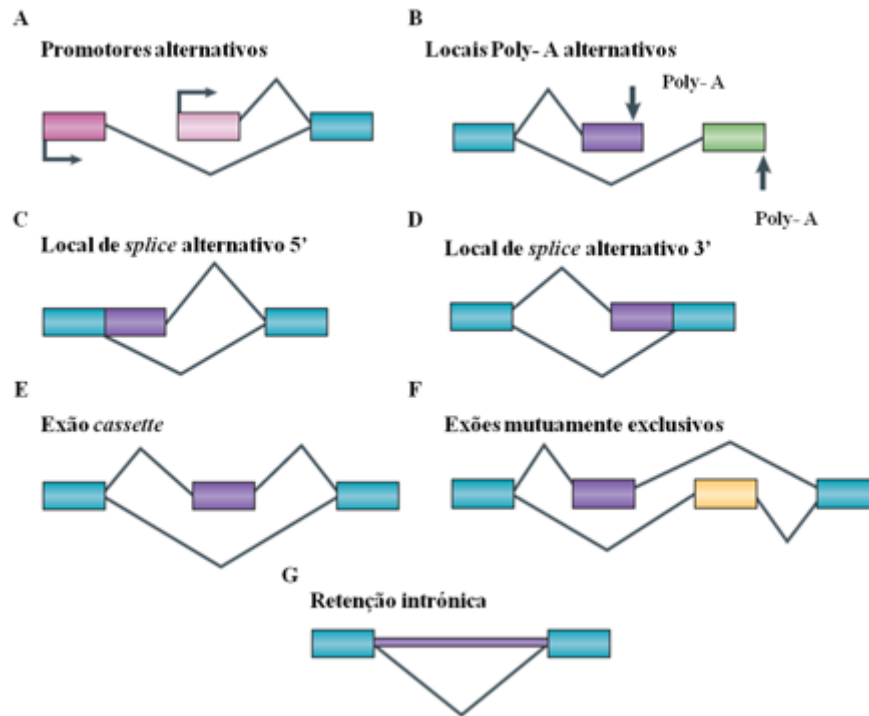
A produção de diferentes isótipos a partir de um mesmo transcrito por *splicing* alternativo é um processo bastante comum em células de mamíferos e pode ocorrer por diversos padrões, os quais variam entre as espécies (Figura 4). Os padrões de *splicing* alternativo podem ser classificados em quatro subgrupos principais, sendo eles o uso dos locais de *splice* alternativos 5' (Figura 4C), o uso dos locais de *splice* alternativos 3' (Figura 4D), a inclusão controlada de um exão *cassette* (Figura 4E), e a retenção de um intrão (Figura 4G). Outros eventos, menos frequentes, que originam transcritos alternativos, incluem o uso de promotores alternativos (Figura 4A), a poliadenilação alternativa (Figura 4B) e exões mutuamente exclusivos (Figura 4F) (Zahler 2005; Li, Lee et al. 2007; Keren, Lev-Maor et al. 2010). Nos pré-mRNAs complexos, mais do que um destes modelos de *splicing* alternativo pode aplicar-se em diferentes zonas de um transcrito (Cartegni, Chew et al. 2002).

Apesar de ser extremamente raro em eucariotas inferiores, análises em bases de dados permitiram saber que o padrão mais comum de *splicing* alternativo é o de exão *cassette*, o qual ocorre em aproximadamente 40% dos eventos de *splicing* alternativo nos eucariotas superiores. Neste tipo de *splicing* alternativo, a totalidade do exão pode ser incluído no mRNA, ou por sua vez excluído (saltado) juntamente com os seus intrões flanqueantes e por isso é também designado de saltar (*skipping*) de exões (Blencowe 2006; Keren, Lev-Maor et al. 2010).

Um caso especial de exões *cassette* duplos mostra um *splicing* mutuamente exclusivo, em que um ou outro exão ou são incluídos, mas não ambos (Li, Lee et al. 2007). Estudos identificaram um grande número de genes humanos e animais contendo pares duplicados de exões, que na sua maioria são mutuamente exclusivos, pois a inclusão de ambos leva a alterações na grelha de leitura (Letunic, Copley et al. 2002).

A iniciação da transcrição em diferentes promotores origina exões com uma extremidade terminal 5' alternativa que podem ser ligados a jusante a um exão 3' comum. De uma forma idêntica exões com sequências 3' alternativas, com locais de poliadenilação diferentes podem-se ligar a montante a um exão comum. Quanto ao uso alternativo dos 5' e 3' locais de *splice*, sabe-se que este tipo de *splicing* alternativo ocorre quando dois ou mais locais de *splice* são reconhecidos numa das extremidades de um exão. Assim, os exões podem ser maiores ou mais pequenos (Li, Lee et al. 2007; Keren, Lev-Maor et al. 2010). Este tipo de *splicing* alternativo é capaz de introduzir alterações subtis nas sequências codificantes, mesmo divergindo em apenas um único codão (Blencowe 2006).

Na retenção do intrão, como o próprio nome indica, um intrão permanece no transcrito de mRNA maduro, sendo o evento mais raro tanto nos vertebrados como nos invertebrados. À medida que se avança na escala filogenética dos eucariotas, maior a prevalência do salto de exão, o que pode sugerir que o salto de exões é o tipo de *splicing* alternativo que contribui em maior escala para a complexidade fenotípica. Pensa-se que os locais de *splice* 5' e 3' alternativos possam ser subfamílias do salto de exão, podendo representar um estadio intermediário evolutivo (Keren, Lev-Maor et al. 2010).



**Figura 4: Representação esquemática dos diferentes padrões para o processo de *splicing* alternativo.** (A) Os locais de iniciação da transcrição (representados pelas duas tonalidades rosa) originam uma extremidade 5' alternativa que se vai ligar ao exão 3' comum a jusante (representado com a cor azul). (B) Exões 3' alternativos (representados a roxo e verde) com locais de poliadenilação (poly A) alternativos podem unir-se a um exão comum a montante (representado com a cor azul) (C) Uso de local de *splice* 5' alternativo (D) Uso de local de *splice* 3' alternativo (E) Exão *cassette*, o qual pode ou não ser incluído no transcrito de mRNA. (F) Exões mutuamente exclusivos, no qual há a seleção de apenas um exão a partir de dois ou mais exões existentes. (G) Retenção de um intrão (adaptado de Li, Lee et al. 2007).

## Parte B – TBCCD1, uma proteína centrossomal

### B.I.1 Citoesqueleto: Microtúbulos

O citoesqueleto é uma rede de polímeros filamentosos e proteínas reguladoras ligados entre si, que estão em constante fluxo, sendo portanto uma estrutura dinâmica e adaptável. Não é por isso de estranhar, que a capacidade que uma célula eucariótica tem para resistir à deformação, para realizar o transporte intracelular e para alterar a sua forma durante o movimento depende do citoesqueleto. Assim, as principais funções realizadas pelo citoesqueleto são a organização espacial do conteúdo da célula, que permite que a célula conecte, fisicamente e bioquimicamente, com o ambiente externo, e gerar forças coordenadas que permitem à célula mover-se e mudar de forma. Para a realização destas funções, o citoesqueleto integra ainda a atividade de uma variedade de proteínas citoplasmáticas e organelos (Fletcher and Mullins 2010).

Os principais componentes do citoesqueleto nas células eucarióticas são os filamentos de actina, os filamentos intermediários e os microtúbulos. Os filamentos de actina e os microtúbulos são polímeros polares, compostos, respetivamente, por subunidades de actina globular (G-actina), que se ligam e hidrolisam ATP e por heterodímeros de  $\alpha$ - e  $\beta$ -tubulina que se ligam e hidrolisam GTP (Li and Gundersen 2008; Wade 2009).

Os microtúbulos estão presentes em todas as células eucarióticas estando envolvidos em diversos processos celulares tais como a mitose, mobilidade celular, transporte intracelular, secreção, manutenção da forma celular e polaridade celular. Estas funções envolvem a interação dos microtúbulos com uma variedade de proteínas a eles associadas (MAPs – *microtubule associated proteins*), que são importantes para a regulação da dinâmica e distribuição dos microtúbulos na célula (Nogales 2001; Conde and Caceres 2009). Os microtúbulos são também os principais componentes do fuso mitótico, uma estrutura altamente dinâmica, fatores-chave na organização da

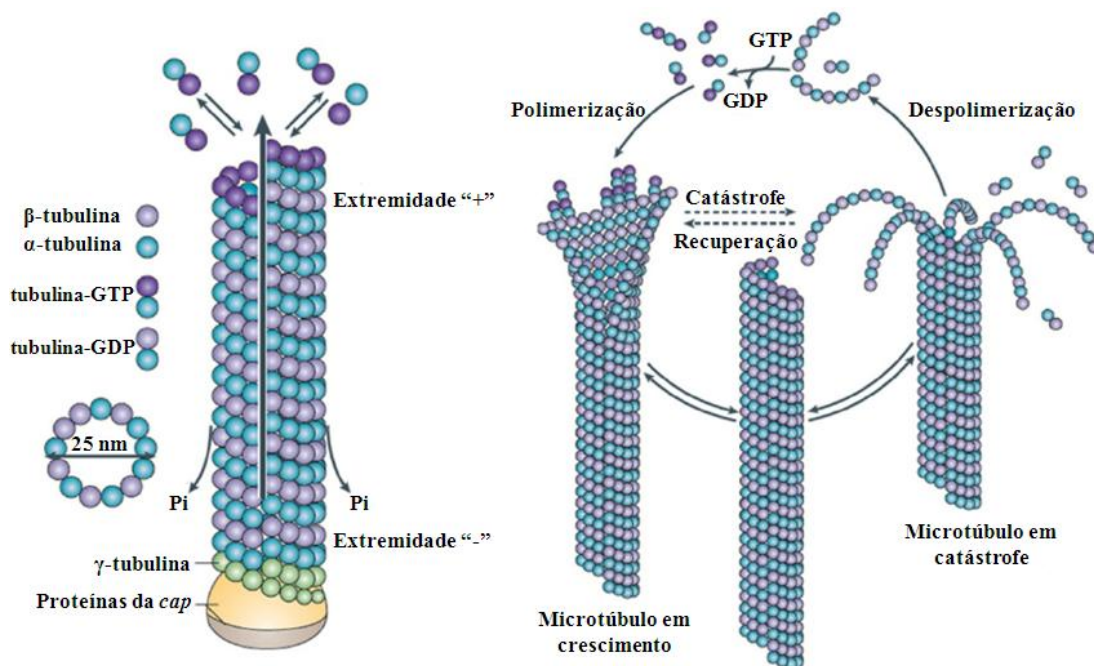
distribuição espacial dos organelos nas células em interfase e ainda componentes extremamente estáveis dos cílios/flagelos e centríolos (Wade 2009).

Em termos estruturais, os microtúbulos são tubos ocos, com cerca de 25 nm de diâmetro e as suas paredes são constituídas, como já foi dito, por heterodímeros de proteínas globulares, a  $\alpha$ - e  $\beta$ -tubulina. Estes dímeros encontram-se associados topo a topo, por ligações não covalentes, formando protofilamentos lineares, geralmente 13 nas células de mamífero, dispendo-se longitudinalmente ao longo da parede (Conde and Caceres 2009; Wade 2009). O modo de associação dos heterodímeros faz com que o microtúbulo apresente duas extremidades diferentes, designadas de extremidade “+” (com crescimento mais rápido) e extremidade “-” (com crescimento mais lento), consoante apresente exposta a  $\beta$ -tubulina ou a  $\alpha$ -tubulina e por isso os microtúbulos apresentem diferentes taxas de polimerização ao nível das suas duas extremidades (Desai and Mitchison 1997; Amos and Schlieper 2005; Conde and Caceres 2009). Consequentemente, os microtúbulos apresentam polaridade estrutural.

Quer a  $\alpha$ - quer a  $\beta$ -tubulina possuem locais de ligação para moléculas de GTP, no entanto, apenas a  $\beta$ -tubulina tem a capacidade de hidrolisar o GTP em GDP e posteriormente permutá-lo novamente por GTP, apresentando uma atividade de GTPase (Amos and Schlieper 2005). Apenas os dímeros de tubulina-GTP podem polimerizar, mas logo após a polimerização, o GTP é hidrolisado em GDP e deixa de ser permutável devido ao fato do seu local de ligação se situar na interface entre dois dímeros subsequentes e por consequência inacessível. A ligação e hidrólise de GTP pela tubulina, permite uma polimerização dinâmica e não equilibrada dos microtúbulos em que cada extremidade do microtúbulo alterna de forma estocástica, entre fases de polimerização (crescimento) e despolimerização (encurtamento), comportamento designado por instabilidade dinâmica (Figura 5) (Desai and Mitchison 1997).

Esta instabilidade dinâmica dos microtúbulos, observada tanto *in vivo* como *in vitro*, pode ser explicada pelo modelo “GTP cap”. Este modelo propõe que a extremidade “+” do microtúbulo tem um cap de tubulina-GTP o qual estabiliza a estrutura do microtúbulo ocorrendo a polimerização, ao passo que o microtúbulo composto por

subunidades de tubulina-GDP é instável. Por isso, quando este *cap* é estocasticamente perdido, as interações entre os protofilamentos diminuem e o microtúbulo rapidamente despolimeriza (Nogales 2001; Conde and Caceres 2009). Esta dinâmica de polimerização dos microtúbulos é crucial para as suas funções biológicas pois permite uma rápida reorganização dos microtúbulos, uma diferenciação espacial e temporal de acordo com o contexto celular, e gerar forças que empurram e puxam durante a polimerização e despolimerização, respetivamente (Conde and Caceres 2009).



**Figura 5: Representação esquemática da dinâmica de microtúbulos.** A polaridade dos microtúbulos é fundamental para a capacidade das proteínas motoras moverem carga para determinados locais, *in vivo*. Esta polaridade resulta da associação da  $\alpha$ - e  $\beta$ -tubulina topo a topo, formando os protofilamentos que constituem a parede do tubo. A instabilidade dinâmica dos microtúbulos é caracterizada pela polimerização e rápida despolimerização dos microtúbulos, em que a transição de polimerização para despolimerização é denominada por catástrofe, ocorrendo a perda da *cap* GTP na extremidade “+” dos microtúbulos. Por sua vez a transição oposta é denominada por recuperação (adaptado de Conde and Caceres 2009).

### **B.I.1.1 Nucleação dos microtúbulos**

Embora estudos *in vitro* mostrem que os microtúbulos podem polimerizar quando a tubulina está presente em grandes concentrações, dentro da célula estes polimerizam a concentrações muito mais baixas. Assim, a concentrações baixas de tubulina a nucleação de novos microtúbulos é cineticamente limitante. De forma a ultrapassar esta barreira cinética, a nucleação *in vivo*, ocorre principalmente em estruturas morfológicamente diferentes, as quais se denominam centros organizadores de microtúbulos (MTOCs – *microtubule organizing centers*) (Wiese and Zheng 2006). Os MTOCs variam muito em tamanho, forma e ocorrência, tanto dentro de uma determinada célula como entre espécies. Por exemplo, o principal MTOC nas células animais é o centrossoma. Estes são organelos citoplasmáticos, constituídos por um par de centríolos e rodeados por uma matriz pericentriolar (PCM). Por outro lado, em leveduras é o corpo polar do fuso (SPM – *spindle pole body*), que está incorporado no envelope nuclear. Existem ainda outros organitos/estruturas capazes de nuclear os microtúbulos, entre eles os corpos basais ou o corpo médio (*midbody*) (Pereira and Schiebel 1997; Wiese and Zheng 2006) e o complexo de Golgi (Chabin-Brion, Marceiller et al. 2001).

No material pericentriolar (PCM), que rodeia os centríolos, encontra-se uma proteína crítica para a nucleação dos microtúbulos, a  $\gamma$ -tubulina. Esta proteína faz parte de um grande complexo multiproteico, que se assemelha a um anel e onde tem origem a nucleação dos microtúbulos, chamado de  $\gamma$ -TuRC ( *$\gamma$ -tubulina ring complex*) (Heald and Nogales 2002).

A  $\gamma$ -tubulina pertence às superfamílias das tubulinas apresentando uma grande homologia com a  $\alpha$ - e  $\beta$ - tubulina (Fuller, Gowen et al. 1995). Além de ser uma proteína ubíqua é filogeneticamente conservada, mas não é incorporada nos protofilamentos que formam as paredes do microtúbulo, no entanto associa-se à sua extremidade “-” a qual se liga ao centrossoma. Esta tubulina além de estar presente na PCM encontra-se ainda no citoplasma (Wiese and Zheng 2006) e nos centríolos do centrossoma (Fuller, Gowen et al. 1995).

As proteínas que integram os complexos nos quais a  $\gamma$ -tubulina se encontra são denominadas por GCPs ( *$\gamma$ -tubulina complex proteins*). Além do  $\gamma$ -TuRC, já referido, a  $\gamma$ -tubulina encontra-se associada a outro complexo, no entanto de menores dimensões, o  $\gamma$ -TuSC ( *$\gamma$ -tubulina small complex*), o qual é composto por duas moléculas de  $\gamma$ -tubulina, cada uma associada com uma molécula de GCP2 e GCP3 (sendo que as GCPs estão numeradas por ordem crescente de massa molecular e a GCP1 corresponde a  $\gamma$ -tubulina). Por sua vez, o  $\gamma$ -TuRC, além de ser composto por múltiplas cópias das proteínas do  $\gamma$ -TuSC, contém uma variedade de proteínas adicionais, incluindo a GCP4, 5 e 6, a GCP-WD (ou NEDD1), a MOZART 1, MOZART 2 e a NME7. Estas quatro últimas, não estão relacionadas com a família GCP (Raynaud-Messina and Merdes 2007; Remy, Merdes et al. 2013). Tanto a posição como a função das GCP4, 5 e 6 no  $\gamma$ -TuRC ainda não se encontram bem estabelecidas, no entanto um alinhamento das suas sequências com a GCP2 e a GCP3 identificou duas regiões conservadas entre estas proteínas, as quais foram denominadas de motivos GRIP1 e GRIP2. Para além disso, dados indicam que a GCP4 provavelmente liga-se à  $\gamma$ -tubulina através do seu domínio GRIP2, sugerindo um papel direto da GCP4 na organização da  $\gamma$ -tubulina dentro do  $\gamma$ -TuRC (Guillet, Knibiehler et al. 2011).

Aparentemente, a presença do  $\gamma$ -TuSC sozinho é suficiente para reunir os complexos de ordem superior que servem de molde para os microtúbulos, visto que os genes que codificam para as proteínas  $\gamma$ -TuSC são essenciais em todos os eucariotas ao passo que os que codificam para as proteínas específicas do  $\gamma$ -TuRC, GCP4, 5 e 6, não são essenciais numa variedade de organismos, onde se inclui a *D. Melanogaster* (Verollet, Colombie et al. 2006; Kollman, Polka et al. 2010; Remy, Merdes et al. 2013).

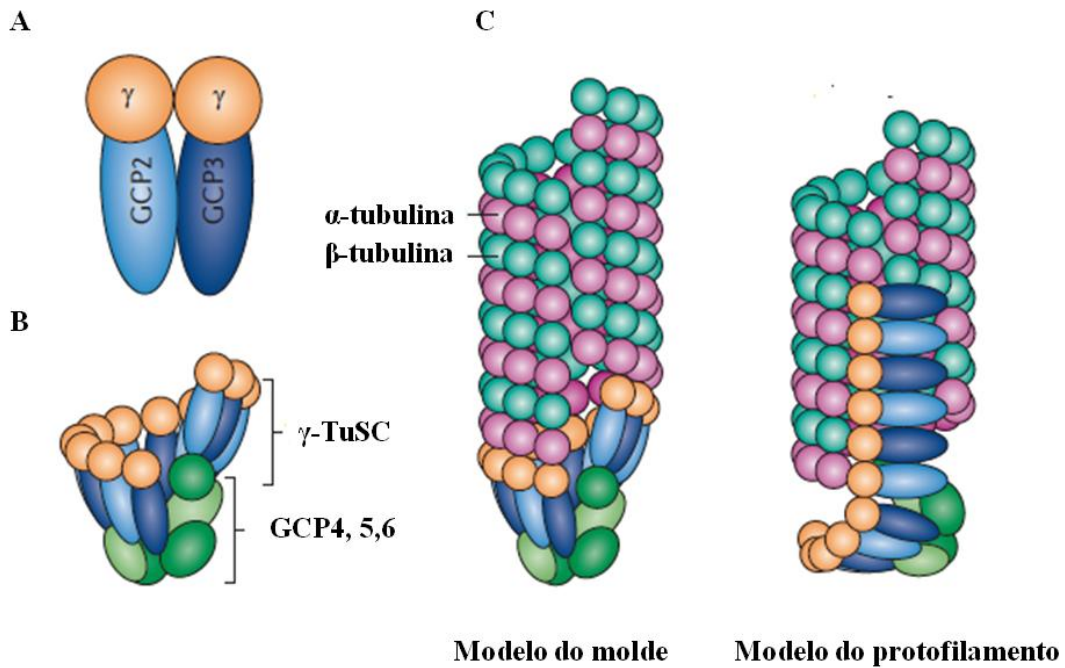
Os contactos laterais entre a N-terminal do domínio GRIP1 da GCP2 e GCP3 levam à formação de vários oligómeros de  $\gamma$ -TuSC, em forma de espiral, sendo que cada volta da hélice apresenta um formato em *lock-washer*, sem contacto entre as camadas. A simetria helicoidal origina à volta de seis e meio  $\gamma$ -TuSCs, ou seja 13  $\gamma$ -tubulinas, por volta com metade de um  $\gamma$ -TuSC sobreposto. A oligomerização dos  $\gamma$ -TuSC produz uma pequena quantidade de complexos em forma de anel que são altamente semelhantes aos  $\gamma$ -TuRCs purificados (Kollman, Polka et al. 2010).

Recentemente, estudos baseados em dados bioquímicos e cristalográficos da proteína GCP4 ditam que a GCP4, 5 e 6, apresentam estruturas semelhantes às GCP2 e 3 e estão integradas na parede da hélice do  $\gamma$ -TuRC (Guillet, Knibiehler et al. 2011). Portanto, é possível que os  $\gamma$ -TuRCs se formem pelas associações laterais do  $\gamma$ -TuSC em hélice e por isso as GCP 4, 5 e 6, devido à sua baixa abundância, ocupem posições específicas no início e no final desta hélice, o que limita o seu crescimento a uma volta completa. Além disso, as GCP4, 5 e 6 podem estabilizar o  $\gamma$ -TuRC e prevenir a sua desmontagem (Kollman, Merdes et al. 2011).

Por sua vez, as proteínas GCP-WD, MOZART 1 e 2 e a NME7 podem ligar-se à base do  $\gamma$ -TuRC e mediar a ligação deste aos MTOCs (Kollman, Polka et al. 2010; Guillet, Knibiehler et al. 2011).

De forma a explicar a nucleação dos microtúbulos, a partir dos complexos  $\gamma$ -TuRC, onde a  $\gamma$ -tubulina assume um papel crítico, existem dois modelos propostos, o modelo do molde e o modelo do protofilamento (Figura 6). O modelo do protofilamento sugere uma disposição perpendicular das subunidades de  $\gamma$ -tubulina no  $\gamma$ -TuRC. Consequentemente, as subunidades de  $\gamma$ -tubulina contactam as proximidades através de interações topo a topo, formando uma espécie de protofilamento de  $\gamma$ -tubulina, que posteriormente serve de base para a formação dos protofilamentos de  $\alpha$ - e  $\beta$ -tubulina (Wiese and Zheng 2006).

Quanto ao modelo do molde, tal como o nome sugere, propõe que o  $\gamma$ -TuRC serve como um molde para as associações longitudinais dos heterodímeros de  $\alpha$ - e  $\beta$ -tubulina. Neste modelo as subunidades adjacentes de  $\gamma$ -tubulina, interagem lateralmente umas com as outras, reunindo-se em forma de anel com 13 moléculas, que por sua vez, formam um molde para a nucleação dos microtúbulos, com 13 protofilamentos de tubulina (Kollman, Polka et al. 2010).



**Figura 6: Representação esquemática dos modelos propostos para a nucleação dos microtúbulos a partir do complexo  $\gamma$ -TuRC. (A) Complexo  $\gamma$ -TuSC.** Apresenta duas cópias de  $\gamma$ -tubulina e cada uma associada a uma molécula de GCP2 e GCP3 **(B) Complexo  $\gamma$ -TuRC.** Modelo atual da organização das subunidades no complexo  $\gamma$ -TuRC, unidade funcional dos microtúbulos. Em vários eucariotas, múltiplos  $\gamma$ -TuSC associam-se com a GCP4, GCP6 e GCP6, formando o  $\gamma$ -TuRC. Por sua vez as proteínas “CAP” (GCP 4-6 e o GCP-WD) são responsáveis por manter este complexo unido. **(C) Modelos propostos para a nucleação dos microtúbulos.** No modelo do molde as subunidades adjacentes de  $\gamma$ -tubulina associam-se lateralmente, formando um anel, a partir do complexo  $\gamma$ -TuRC, que servirá de molde para a nucleação. Por sua vez no modelo do protofilamento as subunidades de  $\gamma$ -tubulina associam-se topo a topo, formando assim um protofilamento, que pela adição de heterodímeros de  $\alpha$ - e  $\beta$ -tubulina será prolongado (adaptado de Kollman, Merdes et al. 2011).

É de notar que, apesar do modelo do protofilamento ainda não ter sido definitivamente descartado, o modelo do molde tem sido descrito como o mais favorável para o mecanismo de nucleação dos microtúbulos, o que é suportado pelos dados de que, *in vitro*, o  $\gamma$ -TuSC de leveduras purificado reúne-se em filamentos do tipo espiral, com 13 moléculas de  $\gamma$ -tubulina, por volta e a  $\gamma$ -tubulina nestes complexos está orientada de forma a expor o seu putativo local de contacto à  $\alpha$ - e  $\beta$ -tubulina longitudinalmente. Mais ainda, estudos da estrutura cristal da  $\gamma$ -tubulina indicam que em ambas as formas cristalinas da  $\gamma$ -tubulina ( $\gamma$ -tubulina-GDP e  $\gamma$ -

tubulina-GTP), as  $\gamma$ -tubulinas individuais associam-se lateralmente através da mesma região de contato que a  $\alpha$ - e  $\beta$ -tubulina usam nas interações laterais dos microtúbulos o que apoia o modelo molde, no qual se prevê interações laterais entre as  $\gamma$ -tubulinas e associações longitudinais ente a  $\gamma$ -tubulina e a  $\alpha/\beta$ -tubulina (Kollman, Polka et al. 2010; Kollman, Merdes et al. 2011).

Nas células, a maioria da  $\gamma$ -tubulina encontra-se solúvel no citoplasma. Esta proteína, ainda que na forma de  $\gamma$ -TuRC, é incapaz de nuclear os microtúbulos. Por isso, é necessário o recrutamento para os MTOCs e consequente ativação. Uma forma de ativar os  $\gamma$ -TuRC pode envolver proteínas adaptadoras que medeiam o recrutamento do  $\gamma$ -TuRC para o centrossoma e microtúbulos do fuso (Remy, Merdes et al. 2013).

Estudos em *D. melanogaster*, por técnica do RNA de interferência (RNAi) mostraram um papel essencial dos componentes do  $\gamma$ -TuSC na mitose e na progressão do ciclo celular, por sua vez, estudos onde se realizou a depleção da GCP-WD, também por RNAi, em células de mamíferos mostraram que os  $\gamma$ -TuRCs, na ausência de GCP-WD, não se acumulam no centrossoma, havendo uma supressão na nucleação dos microtúbulos e um aumento fusos anormais na mitose (Raynaud-Messina and Merdes 2007). Estes resultados mostram que a GCP-WD, como possível proteína adaptadora, é requerida para o recrutamento da  $\gamma$ -tubulina para o centrossoma.

Embora os mecanismos da ligação do  $\gamma$ -TuRC aos centrossomas e microtúbulos estão, em parte, compreendidos, permanece obscuro se a ligação pode diretamente ativar a nucleação dos microtúbulos ou se requer mecanismos reguladores adicionais. Além disso, não é claro quais as subunidades do  $\gamma$ -TuRC que estão envolvidas no processo de ativação. Alterações conformacionais dentro do  $\gamma$ -TuRC são necessárias para ativar a nucleação dos microtúbulos. No complexo de forma helicoidal é necessário que as  $\gamma$ -tubulinas sejam alinhadas lateralmente, de forma a servirem de molde para a polimerização dos 13 protofilamentos dos microtúbulos. Estudos de crio-microscopia eletrônica das hélices dos  $\gamma$ -TuSC mostraram que as moléculas de  $\gamma$ -tubulina ligadas aos GCP3 estão desalinhadas com a GCP2, ligada à  $\gamma$ -tubulina.

Portanto, para uma eficiente nucleação dos microtúbulos tem de ocorrer uma alteração conformacional na GCP3, de forma a posicionar corretamente a  $\gamma$ -tubulina. O domínio GRIP2 da GCP3 atua como um braço flexível, que deve rodar em torno de um ponto de articulação, de forma a alinhar as  $\gamma$ -tubulinas associadas. Assim, todas as moléculas de  $\gamma$ -tubulina são posicionadas num círculo com um diâmetro idêntico ao típico dos microtúbulos, com 13 protofilamentos. No entanto, ainda não é claro como esta alteração conformacional na GCP3 é induzida (Kollman, Polka et al. 2010; Remy, Merdes et al. 2013).

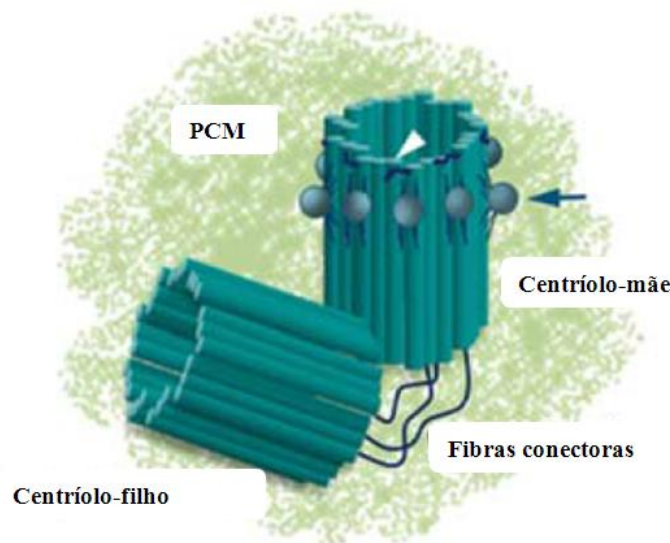
Estudos de localização dos complexos da  $\gamma$ -tubulina aquando da mitose, revelaram que estes não estão apenas ligados ao centrossoma, mas também distribuídos ao longo das fibras do fuso e em alguns casos perto da extremidade “-” dos microtúbulos periféricos. O  $\gamma$ -TuRC, *in vitro*, interage especificamente com a extremidade “-” dos microtúbulos, na qual funciona como um *cap*, de forma a prevenir o crescimento nesta direção. Assim, existem fortes evidências que os  $\gamma$ -TuRCs associados aos microtúbulos influenciam a sua dinâmica, atuando como um fator de estabilidade, o que parece ser independente da função de nucleação dos microtúbulos (Raynaud-Messina and Merdes 2007; Bouissou, Verollet et al. 2009).

## **B.I.2 O centrossoma**

Os centrossomas são considerados organitos invulgares, visto que na maioria das células dos vertebrados, são os únicos não membranares, formando-se através da reunião de componentes à escala nanométrica. Na célula, ocupam um volume pequeno, próximos do centro e em estreita relação com o núcleo (Doxsey 2001). Como já referido anteriormente, é o principal centro organizador de microtúbulos (MTOC) nas células animais, regulando a organização espacial e nucleação dos microtúbulos (Bettencourt-Dias and Glover 2007; Lawo, Hasegan et al. 2012).

### B.I.2.1 Estrutura e função do centrossoma

O centrossoma das células animais é composto por dois centríolos orientados perpendicularmente, o centríolo mãe (ou maduro) e o centríolo filho (ou imaturo), ligados por fibras conectoras nas extremidades proximais. Como referido anteriormente são rodeados por uma matriz proteica densa, designado por material pericentriolar (PCM) (Figura 7) (Bettencourt-Dias and Glover 2007; Luders and Stearns 2007).



**Figura 7: Representação esquemática do centrossoma de células animais.** O centrossoma é constituído por dois centríolos, orientados ortogonalmente, os quais se encontram ligados por fibras conectoras e rodeados pelo material pericentriolar (PCM). A seta azul indica as apêndices distais apresentadas pelo centríolo-mãe, enquanto a seta branca indica os apêndices sub-distais (adaptado de Gonçalves, Tavares et al. 2010b).

Os centríolos apresentam uma estrutura cilíndrica, em forma de barril (*barrel-shaped*), que é constituída por nove tripletos (cada triplete composto por um microtúbulo completo e dois incompletos) de microtúbulos dispostos radialmente e de forma simétrica, estando estruturalmente polarizados ao longo do eixo proximal-distal (Doxsey 2001; Bettencourt-Dias and Glover 2007; Luders and Stearns 2007).

Nas células durante o ciclo celular, os centríolos recrutam a matriz pericentriolar, da qual provêm os microtúbulos. Os centríolos, herdados da divisão celular anterior, são morfologicamente distintos, o centríolo-mãe é o mais velho dos dois e contém

apêndices distais, que se pensa serem críticos para o recrutamento do centríolo e posterior ancoramento à membrana plasmática durante a sua transformação em corpo basal e posterior formação do cílio primário. Existem ainda os apêndices sub-distais, que assumem um papel importante na ligação dos microtúbulos ao centrossoma (Kobayashi and Dynlacht 2011).

Os cílios e flagelos são projeções celulares baseados em microtúbulos, que podem ser móveis e estão envolvidos numa variedade de processos, desde sensores de sinais extracelulares ao movimento fluido e mobilidade celular. Os centríolos são chamados de corpos basais, quando ancorados na membrana para nuclear o axonema, a estrutura central de microtúbulos dos cílios. As propriedades dos corpos basais definem várias características dos cílios, tais como a sua localização e simetria (Debec, Sullivan et al. 2010).

Como já foi dito, os centríolos são rodeados pelo PCM, uma matriz electrodensa que é constituída por uma variedade de proteínas, em que uma grande proporção delas contém regiões *coiled-coil*, no entanto, e apesar de estudos de proteómica identificarem cada vez mais proteínas, permanece bastante obscura de que forma as proteínas *coiled-coil* cooperam para formar esta estrutura (Andersen, Wilkinson et al. 2003; Lawo, Hasegan et al. 2012).

No entanto, quanto às suas propriedades bioquímicas e composição estão bem conhecidas. Sabe-se que existem, pelo menos, quatro classes diferentes de proteínas centrossomais, sendo elas (1) proteínas que mantêm a estrutura do centrossoma e servem de suporte para a formação de outras proteínas centrossomais; (2) proteínas que funcionam na nucleação dos microtúbulos ( $\gamma$ -tubulina juntamente com componentes do  $\gamma$ -TuRC); (3) elementos de ancoragem, que formam uma interface entre proteínas nucleadoras de microtúbulos e os elementos reguladores; (4) as cinases e fosfatases e moléculas sinalizadoras (Lange 2002). Na terceira classe de proteínas, descrita acima, encontra-se a PCNT (pericentrina), uma proteína *coiled-coil* e uma das proteínas de ancoragem melhor caracterizada, a qual se localiza com a

$\gamma$ -tubulina numa estrutura em forma de rede no centrossoma (Delaval and Doxsey 2010).

Sabe-se que durante a mitose os níveis de  $\gamma$ -tubulina e outros componentes do PCM, incluindo a GCP-WD (o principal fator para o recrutamento do  $\gamma$ -TuRC), aumentam inúmeras vezes. As proteínas implicadas no recrutamento do  $\gamma$ -TuRC além da PCNT incluem a CEP192 e a CDK5RAP2, sendo que todas elas são reguladas pela cinase PLK1 e necessárias para a maturação do centrossoma (Haren, Stearns et al. 2009; Lawo, Hasegan et al. 2012).

Estudos recentes mostraram que a PCNT está ancorada aos centríolos através da sua região C-terminal e que a sua região N-terminal se alonga para a periferia do centrossoma. Assim, a PCNT é uma molécula de alongamento, necessária para a formação da PCM na proximidade dos centríolos e para a eficiente formação de centrossomas mitóticos. Na sequência do mesmo estudo mostraram que nas células em interfase, as componentes da PCM adquirem uma distribuição concêntrica toroidal em torno da extremidade do centríolo mãe e difundidos na periferia, onde a sua organização é mantida. Por sua vez, em mitose, a PCNT e a CDK5RAP2 promovem, juntamente com a CEP192, a expansão da rede da PCM, pois estabelecem uma base para o recrutamento de fatores de nucleação dos microtúbulos (Lawo, Hasegan et al. 2012).

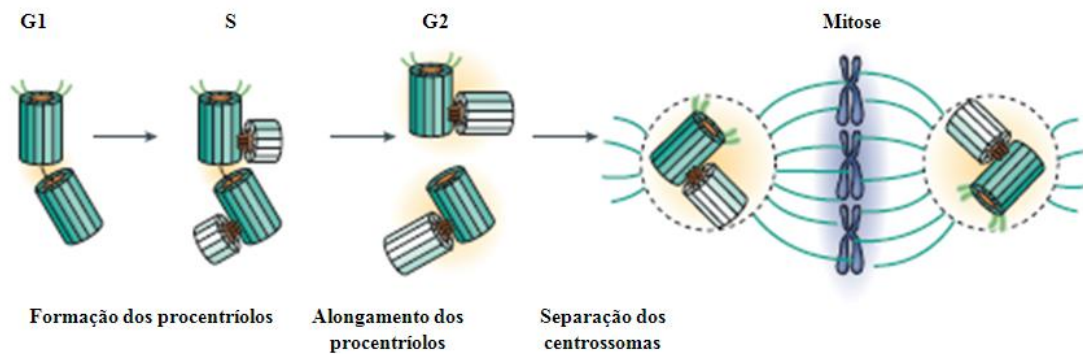
Devido à sua função nucleadora de microtúbulos, os centrossomas regulam diversas funções, mais concretamente, a mobilidade celular, adesão e polaridade na interfase e facilitam a formação dos fusos mitóticos durante a mitose (Andersen, Wilkinson et al. 2003; Chae, Yun et al. 2005; Bettencourt-Dias and Glover 2007).

A identificação de componentes de controlo do ciclo celular e de transdução de sinal nos centrossomas tem sustentado a ideia de que este é mais do que um MTOC. E de facto, evidências moleculares sugerem que o centrossoma assume um papel importante na coordenação da duplicação dos centrossomas, na progressão do ciclo celular, na resposta ao stresse e no controlo do *checkpoint* do ciclo celular. Para

alcançar todas estas funções, o centrossoma interage com uma variedade de moléculas de transdução de sinal (Lange 2002).

### **B.I.2.2 Ciclo de duplicação dos centrossomas**

Em preparação para a mitose, o centrossoma tem que obrigatoriamente duplicar, de forma a garantir os dois pólos do fuso mitótico. De acordo com o ciclo celular, quatro fases consecutivas têm sido descritas, incluindo-se, o afastamento dos dois centríolos (*disengagement*), a nucleação dos centríolos-filhos (também designados de procentríolos, antes de adquirirem o comprimento centriolar por completo), o alongamento dos centríolos-filhos e por fim, a separação dos centrossomas (Figura 8) (Bettencourt-Dias and Glover 2007; Sluder and Khodjakov 2010). O afastamento dos centríolos é caracterizado pela perda da orientação ortogonal entre eles, ocorrendo no final da mitose e início da fase G1, precedendo a formação de novos centríolos. Assim, seguidamente, a duplicação começa no final da G1/S com a formação dos procentríolos nas extremidades proximais dos centríolos-mãe. Os procentríolos crescem progressivamente e adquirem o seu total comprimento na mitose. Cada par de centríolos, mãe e filho, constituem um novo centrossoma. Por fim, a maturação e separação dos dois centrossomas ocorre na transição da fase G2 para a mitose, sendo por isso necessário a aquisição de marcadores de maturação, o recrutamento do PCM e um aumento da atividade do MTOC. Por fim, os centrossomas separam-se fisicamente no começo da mitose, cada um migrando para os pólos do fuso mitótico (Lange 2002; Bettencourt-Dias and Glover 2007; Debec, Sullivan et al. 2010; Pelletier and Yamashita 2012)



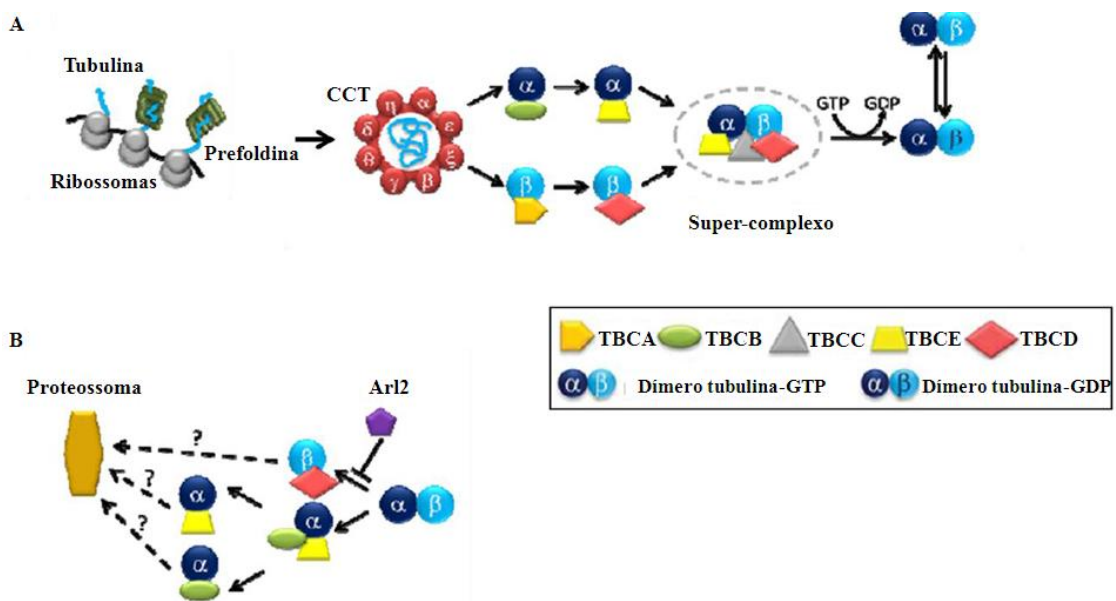
**Figura 8: Representação esquemática das várias fases do ciclo de duplicação dos centrosomas.** Este ciclo tem início na fase G1 com a perda da orientação ortogonal dos centríolos e consequente separação. Seguidamente, a duplicação começa no fim da fase G1, início da fase S com a formação dos procentríolos na extremidade dos centríolos existentes. Os procentríolos apresentam o seu tamanho final no da fase G2 ou no início fase G1 do próximo ciclo. Por fim, ocorre a maturação e separação dos dois centrosomas na transição da fase G2-Mitose, com aquisição de marcadores moleculares e recrutamento do PCM. Consequentemente, no final da fase G2 existem dois pares de centríolos, os quais assumem um papel na organização do fuso mitótico (adaptado de Bettencourt-Dias and Glover 2007).

### B.I.3 Via de *foldin* da tubulina

A síntese de microtúbulos é um processo dinâmico com vários passos, que começa com uma *pool* de subunidades de  $\alpha$ - e  $\beta$ - tubulina (Szolajska and Chroboczek 2011). A via de *foldin* da tubulina é um processo necessário para a montagem do heterodímero de  $\alpha$ - e  $\beta$ - tubulina, para que este possa adquirir a sua estrutura tridimensional funcional. Este requer a interação das tubulinas com vários chaperones moleculares e cofatores da tubulina (TBCA-E) (Gonçalves, Tavares et al. 2010b) (Figura 9).

O primeiro passo desta via é a captação de intermediários do *foldin* da tubulina pela chaperonina CCT (*cytosolic chaperonin-containing TCP1*), que podem ser captados diretamente dos ribossomas ou de um complexo hetero-hexamérico, a chaperone prefoldina (PFD) (Tian, Bhamidipati et al. 1999; Llorca, Martin-Benito et al. 2000). O CCT é um complexo hetero-oligomérico formado por dois anéis sobrepostos, cada um composto por oito subunidades distintas (CCT $\alpha$ -CCT $\zeta$ ). Está bem estabelecido que o complexo CCT medeia o *foldin* impulsionado pela ligação de ATP e sua

hidrólise, libertando a  $\alpha$ - e  $\beta$ - tubulina numa conformação *quasi-nativa* (Llorca, Martin-Benito et al. 2000; Lopez-Fanarraga, Avila et al. 2001). Após a interação com o complexo CCT, as subunidades de tubulina seguem duas vias de *foldings* diferentes. A  $\alpha$ -tubulina é capturada pelo cofator B (TBCB) enquanto a  $\beta$ -tubulina é capturada pelo cofator A (TBCA). Depois disto, o cofator E (TBCE) e o cofator D (TBCD) capturam, respetivamente a  $\alpha$ - e  $\beta$ -tubulina (Tian, Lewis et al. 1997). Seguidamente, as duas vias convergem, havendo a formação de um supercomplexo, composto pela  $\alpha$ - e  $\beta$ -tubulina, pelo TBCE e pelo TBCD, com o qual o cofator C (TBCC) interage, promovendo a hidrólise de GTP pela  $\beta$ -tubulina e a consequente libertação de heterodímeros  $\alpha/\beta$ -tubulina-GDP nativos (Tian, Lewis et al. 1997). No entanto, só após a troca de GDP por GTP é que estes heterodímeros se tornam competentes para polimerizarem os microtúbulos (Lopez-Fanarraga, Avila et al. 2001; Gonçalves, Tavares et al. 2010b).



**Figura 9: Representação esquemática da via de *foldings* da tubulina.** (A) O complexo prefoldina (PFD) interage com as cadeias recém-sintetizadas de  $\alpha$ - e  $\beta$ -tubulina, direcionando-as para a chaperonina citosólica (CCT). Após este complexo assistir no seu *foldings*, são libertados monómeros de  $\alpha$ - e  $\beta$ -tubulina numa forma *quasi-nativa*, que por sua vez vão seguir vias divergentes. Enquanto a  $\alpha$ -tubulina interage com o TBCB e o TBCE, a  $\beta$ -tubulina vai interagir com o TBCA e o TBCD. Através da interação do TBCC, estas vias voltam a convergir resultando na formação de um supercomplexo. A interação do TBCD com o TBCC promove a hidrólise de GTP da  $\beta$ -tubulina e

subsequente libertação do heterodímero de  $\alpha$ - e  $\beta$ -tubulina. Por fim, ocorre ainda a troca de GDP por GTP, no local de ligação ao GTP da  $\beta$ -tubulina, pois só assim os heterodímeros de  $\alpha$ - e  $\beta$ -tubulina estão aptos a polimerizar os microtúbulos (Gonçalves, Tavares et al. 2010b). **(B)** A via de *fold*ing da tubulina pode ainda ocorrer no sentido inverso. Aqui, os heterodímeros nativos de  $\alpha$ - e  $\beta$ -tubulina após a despolimerização dos microtúbulos podem ser dissociados em  $\alpha$ - e  $\beta$ -tubulina através de uma interação com o TBCE e TBCB ou pela interação com o TBCD, respetivamente (Kortazar, Fanarraga et al. 2007). A dissociação do heterodímero é regulada pela proteína Arl2, através formação de um complexo com o TBCD e consequente impedimento da ligação deste à  $\beta$ -tubulina (Bhamidipati, Lewis et al. 2000). Os monómeros de  $\alpha$ - e  $\beta$ -tubulina podem depois ser reciclados pelo super-complexo ou pelo contrário degradados pelo proteossoma, no caso de estarem danificados de (adaptado de Gonçalves, Tavares et al. 2010b).

Sabe-se ainda que a via de *fold*ing da tubulina pode ocorrer no sentido inverso. Isto significa que os cofatores B, E e D têm a capacidade de dissociar os dímeros de tubulina, causando a despolimerização dos microtúbulos, quando sobre-expressos em mamíferos (Gonçalves, Nolasco et al. 2010a; Gonçalves, Tavares et al. 2010b). Após a dissociação do heterodímero, o TBCB e o TBCE formam um complexo ternário capaz de capturar a  $\alpha$ -tubulina. Por sua vez a  $\beta$ -tubulina é recuperada pelo TBCE (Kortazar, Fanarraga et al. 2007). A interação da proteína Arl2 (*ADP ribosylation factor-like protein 2*), uma GTPase, com o TBCD impede que este se ligue à  $\beta$ -tubulina, o que vai prevenir a dissociação dos heterodímeros nativos de tubulina e por sua vez a despolimerização dos microtúbulos é também evitada (Bhamidipati, Lewis et al. 2000).

### **B.I.3.1 Cofatores da tubulina**

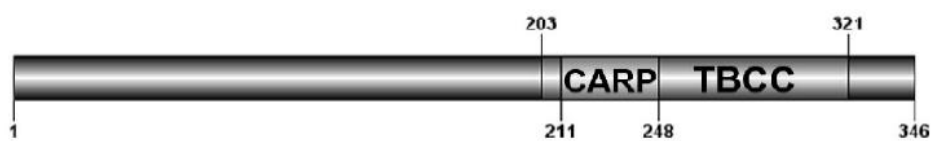
Os fatores que regulam a polimerização e dinâmica do citoesqueleto de microtúbulos assumem um papel crítico para a determinação do comportamento da célula e o seu destino. Neste contexto, os cofatores da tubulina (TBCE-E) assumem papéis de grande relevância na via de *fold*ing da tubulina e regulam a dinâmica de microtúbulos de diferentes formas (Szolajska and Chroboczek 2011). A participação dos cofatores da tubulina no controlo da nucleação/formação dos microtúbulos tem também sido foco de estudo. De fato, tanto o TBCB como o TBCD encontram-se localizados no centróssoma. Em humanos, o TBCD contém dois domínios

envolvidos nesta localização centrossomal, onde é necessário para o recrutamento do  $\gamma$ -TuRC e organização dos fusos mitóticos, e também um domínio de ligação aos microtúbulos. A sobre expressão de estes dois cofatores provoca excessivos MTOCs acentriolares contendo  $\gamma$ -tubulina, que no caso do TBCD é acompanhado por figuras mitóticas aberrantes (Cunningham and Kahn 2008; Fanarraga, Bellido et al. 2010). Em linhas celulares humanas o TBCA tem um papel importante na reciclagem de heterodímeros maduros de tubulina (Nolasco, Bellido et al. 2005).

O TBCB e o TBCA servem de reservatórios para excesso de  $\alpha$ - e  $\beta$ -tubulina, respetivamente. Como já referido anteriormente, o TBCD e o TBCE podem dissociar os dímeros de tubulina, o que induz à despolimerização dos microtúbulos, processo que é regulado por, pelo menos, uma das proteínas *ARF-like* (*ARF – ADP ribosylation factor*), o *Arl2 (ADP ribosylation factor-like protein 2)*.

Estes dados juntamente com o facto que domínios específicos dos TBCs foram identificados em proteínas funcionais, a *RP2* e *E-like*, relacionadas com o TBCC e TBCE, respetivamente, sugerem que estes podem ter evoluído para interagir com requisitos específicos da formação e dinâmica dos microtúbulos, possivelmente em tipos de células específicos.

#### B.I.3.1.1 Cofator C (TBCC)



**Figura 10: Proteína humana TBCC.** Representação esquemática dos seus domínios funcionais CARP e TBCC (adaptado de Gonçalves 2010).

Sabe-se que o cofator C, identificado por Cowan e seus colaboradores é uma cofator crítico para o correto *folding* da tubulina. Este apresenta atividade de GAP (*GTPase activating protein*) a qual é essencial para a hidrólise do GTP pela  $\beta$ -tubulina, passo

fundamental para a libertação do heterodímero de  $\alpha/\beta$ -tubulina a partir do supercomplexo, tal como referido anteriormente (Tian, Huang et al. 1996; Tian, Bhamidipati et al. 1999).

O TBCC apresenta dois domínios funcionais na sua região C-terminal, o domínio CARP e o domínio TBCC (Figura 10). Esta característica é partilhada pela proteína RP2, na sua região N-terminal, que será discutida em pormenor mais à frente.

Existem poucos estudos acerca da função deste cofator *in vivo*, no entanto ensaios de complementação em levedura, mostraram que a expressão do TBCC e da RP2 de mamífero consegue reverter o fenótipo da deleção do gene *cin2* (gene homólogo do TBCC em levedura *Saccharomyces cerevisiae*) o que confirma a conservação da função entre as duas proteínas e consequentemente uma atividade GAP para com a tubulina, derivada das propriedades funcionais do domínio TBCC. Além disso, uma mutação no resíduo de arginina conservado entre as proteínas TBCC e RP2, no domínio TBCC, suprime a sua atividade GAP para com a tubulina, não sendo por isso possível a reversão do fenótipo da deleção do CIN2. Sabe-se que este resíduo poderá corresponder a um “dedo de arginina” que é característico das proteínas GAP e essencial a esta atividade (Bartolini, Bhamidipati et al. 2002).

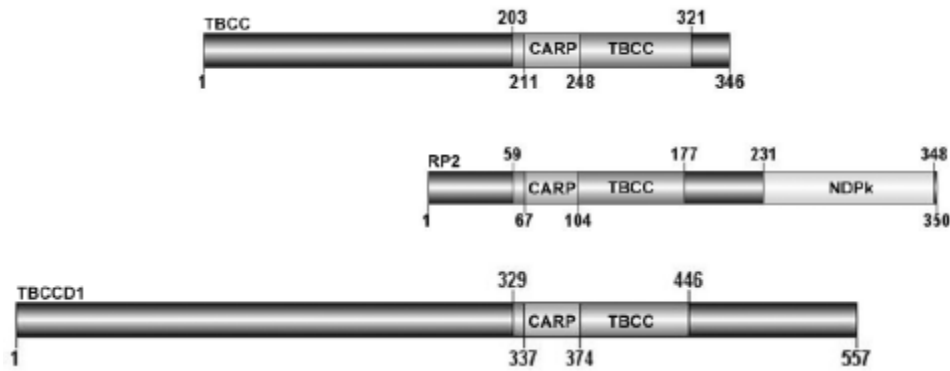
Estudos de sobre-expressão do TBCC na linha celular MCF-7 permitiram verificar que este influencia a formação da rede de microtúbulos, uma vez que nas células onde ocorreu um aumento do TBCC, aumentou também o conteúdo de heterodímeros de tubulina não polimerizados, diminuindo a dinâmica dos microtúbulos. A distribuição das células pelo ciclo celular também apresentou alterações, observando-se uma maior quantidade de células na fase G2-M e menor na fase S, juntamente com um atraso na passagem para a mitose (Hage-Sleiman, Herveau et al. 2010)

## **B.I.4 Proteínas relacionadas com os cofatores da tubulina**

### **B.I.4.1 Proteínas que contém o domínio TBCC: TBCC, RP2 e TBCCD1**

Os dados acima mencionados levaram investigadores do nosso grupo a procurarem na base de dados do genoma humano outras proteínas relacionadas com os cofatores da tubulina através de domínios conservados. Durante este processo foi identificada uma nova proteína conservada ao longo da árvore filogenética, a qual foi designada por TBCCD1, *TBCC-domain containing 1*. O TBCCD1 está relacionado com o TBCC, uma vez que partilha os domínios funcionais TBCC e CARP (Gonçalves, Nolasco et al. 2010a). Além do TBCCD1, também a proteína RP2 (*Retinitis pigmentosa protein 2*) está relacionada com o TBCC, visto que apresenta homologia nesses mesmos domínios (Schwahn, Lenzner et al. 1998). Este motivo CARP que integra o domínio TBCC está associado às proteínas CAPs (*cyclase-associated-proteins*). As CAPs são proteínas altamente conservadas que se ligam a monómeros de G-actina e estão presentes em todos os organismos eucariotas. Estudos têm sugerido um papel importante destas proteínas na sinalização celular e polimerização do citoesqueleto de actina, no transporte vesicular e no desenvolvimento (Hubberstey and Mottillo 2002).

No que respeita ao TBCCD1 e cofator C ambos os domínios, TBCC e CARP encontram-se na região C-terminal, ao passo que na RP2 encontram-se na sua região N-terminal (Figura 11).



**Figura 11: Representação esquemática das proteínas que estão relacionadas com o TBCC.** Domínios funcionais TBCC e CARP partilhados pelas proteínas humanas TBCC, RP2, e TBCCD1 (adaptado de Gonçalves 2010).

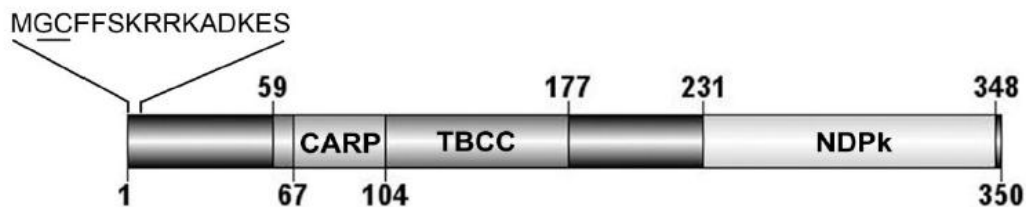
#### B.I.4.1.1 Retinitis pigmentosa protein 2 (RP2)

A retinite pigmentosa é uma doença geneticamente heterogénea caracterizada pela degeneração progressiva da retina, sendo a principal forma de cegueira hereditária. As mutações no gene que codifica para a proteína RP2 estão relacionadas com a segunda forma mais frequente de *X-linked retinis pigmentosa* (XLRP).

Relativamente à proteína RP2 constituída por 350 resíduos de aminoácidos sabe-se que é expressa ubiquamente, e através de estudos de imunofluorescência, em fusão com a proteína fluorescente verde (GFP), verificou-se que esta se localiza no citoplasma e associada à membrana plasmática. No entanto esta localização pode variar consoante a linha celular usada (Schwahn, Lenzner et al. 1998; Chapple, Hardcastle et al. 2002; Chapple, Grayson et al. 2003). Estudos de filogenia permitiram saber que a RP2 se encontra presente em organismos que possuem cílios/flagelos (Stephan, Vaughan et al. 2007).

Além do que já foi dito, substituições de aminoácidos na RP2 nos resíduos conservados com o TBCC que levaram à patogenicidade, sugerem também uma homologia funcional entre as duas proteínas, no entanto a sua função ainda não é bem conhecida (Chapple, Hardcastle et al. 2002). Como já referido anteriormente o

domínio N-terminal desta proteína partilha homologia com o TBCC e tal como este, a RP2 estimula *in vitro* a atividade GTPase da tubulina nativa na presença do TBCD. No entanto, não substitui o TBCC na via de *folding* da tubulina (Bartolini, Bhamidipati et al. 2002). Por sua vez a sua região C-terminal apresenta um domínio com homologia às NDPk (*nucleoside diphosphates kinases*) (Figura 12).



**Figura 12: Representação esquemática da proteína RP2.** Além dos domínios funcionais CARP, TBCC e NDK, está ainda representado a sequência de resíduos de aminoácidos onde ocorrem as modificações pós-traducionais desta proteína. Os resíduos G e C (sublinhados) são essenciais para a localização da RP2 na membrana plasmática (Chapple, Hardcastle et al. 2000) (adaptado de Gonçalves 2010).

A maioria das mutações encontradas na RP2 resulta em proteínas truncadas com perda do domínio C-terminal, onde se encontra o domínio NDPK, indicando uma função importante. As NDKs estão envolvidas na transferência de grupos fosfato do ATP e além da sua função como cinases existe uma grande evidência que algumas destas, em mamíferos, apresentam funções no processamento do DNA. Algumas delas, tais como a NDK1 e NDK2, *in vitro*, podem clivar sequências de DNA com estruturas invulgares. Quanto à proteína humana NDK1 além do seu papel na transferência de grupos trifosfato, tem sido descrita como tendo uma atividade de exonuclease de 3'-5'. Nesta sequência, estudos provaram que a RP2 é uma proteína de ligação ao DNA semelhante à NDK1 humana que exibe atividade de exonuclease. Sabe-se ainda que assume um papel como fator de resposta aos danos no DNA (Yoon, Qiu et al. 2006).

A localização da proteína RP2 na membrana plasmática é dependente de modificações pós-traducionais que ocorrem no motivo Met-Gly-Cys-Xaa-Phe-Ser-

Lys (onde Xaa, representa um aminoácido qualquer), presente na sua região N-terminal (Figura 12) (Chapple, Hardcastle et al. 2002; Chapple, Grayson et al. 2003).

Atá à data, a única proteína que parece interagir com a RP2 é a Arl3 (*ADP ribosylation factor-like protein 3*). A Arl3 pertence à família das pequenas GTPases Arl, funcionando como uma GAP para a RP2 (Grayson, Bartolini et al. 2002; Evans, Schwarz et al. 2010).

Estudos indicam que a proteína RP2 se localiza no cílio conector de células fotorreceptoras, mais concretamente no seu corpo basal e associação com os centríolos da base do cílio. Foi ainda encontrada na região peri-centriolar e no complexo de Golgi, o que sugere um papel importante da RP2 na regulação do transporte peri-centriolar de vesículas e no transporte intraflagelar. Mais ainda, tanto o silenciamento por técnicas de RNAi dos genes que codificam para a RP2 como para a Arl3, leva à fragmentação do complexo de Golgi, o que resulta na dispersão do mecanismo do transporte vesicular deste organito para o corpo basal do cílio (Evans, Schwarz et al. 2010).

#### B.I.4.1.2 TBCC-domain containing protein 1 (TBCCD1)

Como já referido anteriormente, recentemente foi identificada uma terceira proteína com um domínio TBCC e um domínio CARP, o TBCCD1, sendo por isso relacionada com as proteínas TBCC e RP2. O TBCCD1 é uma proteína com 557 aminoácidos, uma massa molecular prevista de ~ 64 kDa e um pI previsto de 8.8688. Por sua vez, o gene humano que codifica esta proteína, *tbccd1*, está localizado no cromossoma 3 (3q27.3) e possui 8 exões (Figura 13) (Gonçalves 2010).

O TBCCD1 além de descrita e caracterizada em células humanas pelos investigadores do nosso grupo está também caracterizada na alga verde *Chlamydomonas reinhardtii* (Feldman and Marshall 2009; Gonçalves, Nolasco et al. 2010a). Através de um estudo filogenético, sabe-se ainda que esta proteína é conservada na maioria dos organismos analisados, estando ausentes nas leveduras *S. cerevisiae* e

*Shizosaccharomyce pombe* e ainda na *Hydra magnipapillata* e *D. melanogaster* (Gonçalves 2010).



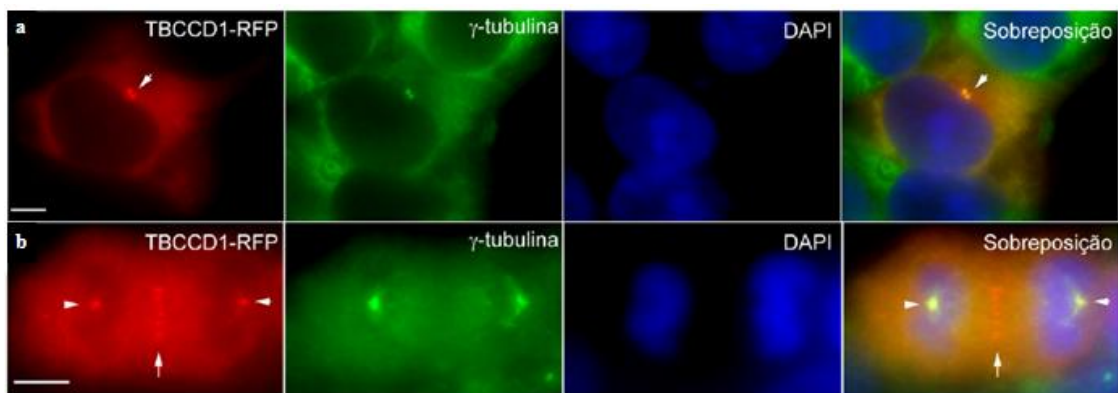
**Figura 13: Proteína TBCCD1 humana.** Representação esquemática dos seus domínios funcionais CARP e TBCC (adaptado de Gonçalves 2010).

O facto de o TBCCD1 apresentar os dois domínios homólogos ao TBCC e RP2 sugeriu a hipótese de partilhar funções com estas proteínas, inclusivamente a ativação da atividade GTPase da  $\beta$ -tubulina e um consequente possível papel na via de *fold*ing da tubulina. No entanto, sabe-se que o TBCCD1 humano não possui um resíduo de arginina, que como referido anteriormente, é crucial para a atividade GAP destas proteínas e que constitui um “dedo de arginina” característico das proteínas com função GAP. Apesar disto, o TBCCD1 humano possui um resíduo de arginina próxima da posição conservada no TBCC e na RP2 e por isso propôs-se que poderia desempenhar uma função idêntica (Feldman and Marshall 2009; Gonçalves, Nolasco et al. 2010a). Nesta sequência realizou-se um ensaio de complementação em *S.cerevisiae*, no qual se observou que o TBCCD1 é incapaz de reverter o fenótipo da deleção do gene *cin2* (gene homólogo do TBCC em levedura) o que revelou à sugestão de que o TBCCD1 não possui a função de ativar GTPases, tal como acontece com o TBCC e a RP2 (Gonçalves, Nolasco et al. 2010a).

Estudos de microscopia de imunofluorescência, usando a expressão de proteínas de fusão com proteínas fluorescentes e com um anticorpo específico para o TBCCD1 permitiram saber que esta proteína se localiza no citoplasma, na matriz pericentriolar do centrossoma, na zona mediana do fuso mitótico na anafase, no corpo médio do fuso e no corpo basal de cílios primários e motores em diferentes linhas celulares humanas (Figura 14) (Gonçalves, Nolasco et al. 2010a). Estudos em *C. reinhardtii* permitiram também observar a localização centrossomal do TBCCD1,

especificamente nos centríolos dos seus dois flagelos e numa estrutura que liga os dois centríolos ao núcleo (designado por rizoplasto) (Feldman and Marshall 2009).

Assim sendo, seria de esperar um papel importante do TBCCD1 ao nível do centrossoma. No entanto, a planta *Arabidopsis thaliana*, que apresenta duas proteínas ortólogas ao TBCCD1 é um organismo que não possui centrossoma. Isto sugere que esta proteína possa desempenhar funções que não estão diretamente relacionadas com a presença de centrossomas e centríolos nem com a presença cílios ou flagelos (Gonçalves 2010).

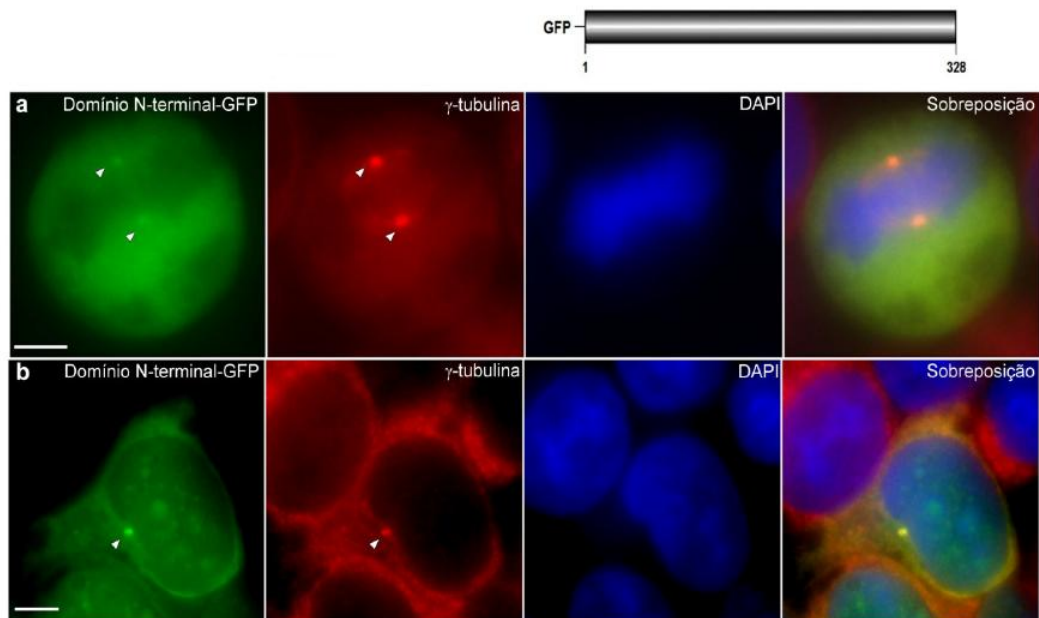


**Figura 14: Determinação da localização celular do TBCCD1 em células humanas.** As células HEK 293 foram transfetadas com o plasmídeo pIC112-TBCCD1 para se expressar a fusão TBCCD1-RFP e seguidamente processadas para imunofluorescência com um anticorpo contra a  $\gamma$ -tubulina e o DNA foi marcado com DAPI. Verificou-se que esta proteína se localiza no centrossoma, tanto em (a) como em (b), células em interfase e em mitose, respetivamente; as cabeças das setas apontam para o TBCCD1-RFP nos centrossomas (a e b); (b) as setas apontam para o TBCCD1-RFP na zona mediana do fuso, de células em anafase. As barras de escala correspondem a 5  $\mu$ m (Gonçalves, Nolasco et al. 2010a);(Gonçalves 2010).

Após a despolimerização da rede de microtúbulos com nocodazole, verificou-se que não houve alterações ao nível da localização do TBCCD1, o que significa que este é um elemento integral do centrossoma, visto que a sua localização é independente dos microtúbulos. Sabe-se ainda que o domínio que determina esta localização centrossomal está presente na sua região N-terminal. Ambas as regiões, C-terminal e N-terminal do TBCCD1 foram expressas em fusão com proteínas fluorescentes e

apenas a proteína de fusão que incluía a região N-terminal se localiza no centrôssoma (Figura 15) (Gonçalves 2010).

Na sequência destes resultados, estudos posteriores determinaram qual o domínio mínimo crítico na região N-terminal para a sua localização. Verificou-se, através de deleções nesta extremidade que nenhuma das proteínas construídas se localizava no centrôssoma, sendo a sequência correspondente aos primeiros 20 resíduos de aminoácidos essencial para a localização centrôssomal do TBCCD1. Ainda por microscopia de imunofluorescência observou-se que este domínio sozinho em fusão com a GFP localiza-se no centrôssoma, sendo por isso o domínio crítico que determina a localização desta proteína no centrôssoma (Pereira 2011).

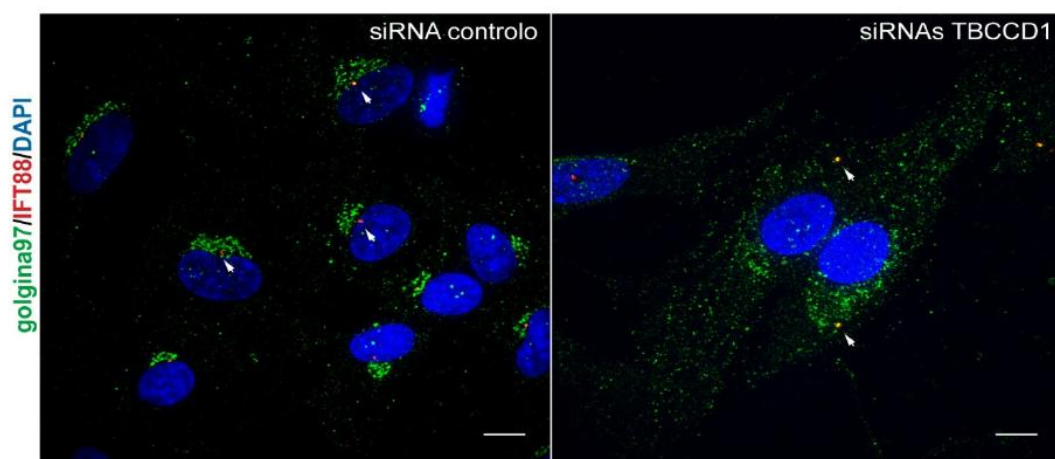


**Figura 15: A região N-terminal do TBCCD1 é responsável pela sua localização centrôssomal.** As células transfetadas com apenas a região N-terminal do TBCCD1 em fusão com a GFP indicaram a localização centrôssomal desta proteína truncada, tanto em mitose como em interfase, (a) e (b), respectivamente. As células HEK 293 foram processadas para análise de imunofluorescência indireta com um anticorpo contra a  $\gamma$ -tubulina e o DNA foi marcado com DAPI. As barras de escala correspondem a 5 $\mu$ m (adaptado de Gonçalves 2010; Gonçalves, Nolasco et al. 2010a).

Estudos desenvolvidos com esta proteína mostraram ainda por técnicas de *western blot* que o TBCCD1 é expresso ubiquamente nas diferentes linhas celulares humanas

apresentando níveis de expressão que variam com a linha. Por exemplo, nas células hTERT-RPE-1, linha com fenótipo normal, os níveis de TBCCD1 são muito menores que em linhas cancerígenas. O TBCCD1 é ainda expresso em vários órgãos de murganho, com principal destaque para os testículos e cérebro. Adicionalmente, efetuou-se a técnica de RT-PCR, de forma a analisar a expressão do gene *tbccd1* nos vários tecidos de murganho e linhas celulares. Comprovou-se que este é transcrito em todas as linhas celulares analisadas e em todos os tecidos de murganho, com uma maior prevalência do seu mRNA nos testículos (Gonçalves, Nolasco et al. 2010a).

Com o intuito de se compreender a função desempenhada pelo TBCCD1, procedeu-se ao silenciamento do gene *tbccd1*, em células hTERT-RPE-1, pela utilização da técnica de RNAi. Com este ensaio, observou-se que a diminuição da proteína TBCCD1 leva a um aumento de tamanho das células transfectadas e por isso estas não atingiam a confluência de igual forma às células controlo, mas também um atraso no ciclo celular na fase G1. Verificou-se ainda que o centrossoma encontrava-se deslocado do centro da célula, aumentando a sua distância ao núcleo, apresentando-se frequentemente na periferia da célula (Figura 16). Estes dados mostram que o TBCCD1 é essencial ao correto posicionamento do centrossoma no centro da célula em estreita associação com o núcleo (Gonçalves, Nolasco et al. 2010a).



**Figura 16:** O silenciamento do gene *tbccd1* causa desorganização do complexo de Golgi e um afastamento do centrossoma relativamente ao núcleo. Células RPE-1 transfectadas com o siRNA

controle ou com a mistura de siRNAs contra o *tbccd1* foram processadas para análise de imunofluorescência indireta com anticorpos contra a proteína golgina 97, marcador do complexo de Golgi e contra a proteína IFT88/polaris, marcador do centrossoma. O DNA foi marcado com DAPI. As barras de escala correspondem a 10  $\mu\text{m}$  (adaptado de Gonçalves 2010; Gonçalves, Nolasco et al. 2010a).

Ainda na sequência desta experiência foi possível encontrar outro fenótipo. Observou-se a desorganização do complexo de Golgi, apresentando-se muitas vezes fragmentado e espalhado por toda a célula ou parecendo seguir o centrossoma (Figura 16). Relacionado com este fenótipo poderá estar uma diminuição da produção de cílios primários, visto que complexo de Golgi é um organito essencial na organização destes. Nas células em que o gene *tbccd1* foi silenciado verificaram-se também problemas, tanto na migração celular, como na migração celular direcionada, tornando-a mais lenta que nas células controle (Gonçalves, Nolasco et al. 2010a).

## Parte C – Objetivos

O grupo de investigação, no qual o trabalho aqui apresentado se inclui tem como principais objetivos o estudo do papel dos cofatores da tubulina na regulação da biossíntese e dinâmica do citoesqueleto de microtúbulos, bem como a influência das concentrações reguladoras de  $H_2O_2$  nestes fenómenos.

O TBCCD1 foi uma proteína recentemente descoberta por investigadores do nosso grupo. Sabe-se que está relacionada com as proteínas TBCC e RP2, uma vez que as três partilham domínios funcionais, sendo eles o domínio TBCC e o domínio CARP (Gonçalves, Nolasco et al. 2010a).

Embora a principal função desta proteína ainda não seja conhecida estudos mostraram que se localiza no centrossoma sendo um potencial regulador do seu posicionamento e do seu ancoramento ao núcleo. Desta forma parece estar implicada na organização do citoesqueleto de microtúbulos e por sua vez na organização espacial de organitos, tal como o complexo de Golgi.

Consequentemente, o principal objetivo do trabalho aqui apresentado foi o estudo da função biológica da proteína TBCCD1, principalmente o seu possível envolvimento em processos dependentes dos microtúbulos e ancoramento do centrossoma ao núcleo.

No mesmo trabalho descobriram que existe uma maior quantidade de proteína TBCCD1 em linhas celulares tumorais, HeLa e HEK 293T, do que em linhas fenotipicamente normais, como a hTERT-RPE-1. Assim o facto das células cancerígenas serem células pouco aderentes e com uma capacidade de divisão e migração mais rápidas juntamente com algumas das características do TBCCD1 citadas na introdução, nomeadamente quando o gene que codifica para esta proteína é silenciado verificaram-se problemas tanto na migração celular como na migração celular direcionada, tornando-a mais lenta que nas células controlo, surge o interesse de estudar padrões de expressão deste gene em linhas cancerígenas.

De forma a estudar a função biológica da proteína TBCCD1 em maior pormenor e sabendo antecipadamente que o gene humano que a codifica está localizado no cromossoma 3, possui 8 exões e portanto apresenta a sua região codificante interrompida por vários intrões sugeriu a possibilidade de ocorrer *splicing* alternativo. Estas observações corroboram com o facto de que a maioria dos genes eucarióticos que codificam para proteínas são interrompidos por múltiplos intrões, o que parece ser um principal requisito para a ocorrência de *splicing* alternativo (Irimia, Rukov et al. 2007b).

Especificamente, de uma forma resumida, neste trabalho pretendeu-se:

Estudar a possibilidade de existência de transcritos alternativos da proteína TBCCD1 e conseqüente estudo da sua função, através da caracterização dos fenótipos da linha celular cancerígena HEK 293T expressando os possíveis isótipos alternativos em fusão com a proteína GFP, por técnicas de imunofluorescência indireta, analisando por exemplo:

- i) Localização intracelular;
- ii) Impacto na ligação do centrossoma ao núcleo.

## Capítulo II - Material e Métodos

### II.1 Reagentes

Os meios de cultura Dulbecco's Modified Eagles's medium DMEM 1x com Glutamax e DMEM/F12 (Ham) (1:1) 1x com Glutamax, os aminoácidos não essenciais, a solução 7,5% de bicarbonato de sódio, a tripsina (0,05%) 1x com EDTA e a solução tampão fosfato salino (PBS) 10x para cultura de células de mamífero foram adquiridos da GIBCO™, Invitrogen, tal como a Lipofectamine 2000 e o meio Optimem.

O metanol, o etanol, o Tween 20 e o cloreto de Sódio (NaCl) foram obtidos da Merck. O dimetilsulfóxido (DMSO), o Tris-Base, a solução de azul de tripano 0,4%, a Poli-L-Lisina e o 4',6-diamidino-2-fenilindole (DAPI) foram adquiridos da Sigma. O glicerol adveio da Fluka. A albumina do soro bovino para imunofluorescência proveio da Calbiochem® e o meio de montagem Dapi Fluoromount G da SouthernBiotech. A Bacto peptona, o Bacto agar e o extrato de levedura foram obtidos da Difco.

Todos os primers utilizados foram sintetizados pela Sigma. O DNase I foi obtido da MBI Fermentas GMBH assim como o enzima ligase de DNA do fago T4. O GreenSafe, o marcador de massas moleculares de DNA NZYLadder III, o enzima polimerase de DNA, NZY Premium DNA polimerase e os kits NZYMiniprep e NZYMidiprep foram adquiridos da Nzytech. O QIAquick Gel extraction, o QIAprep Spin Miniprep e o Plasmid Midi Kit foram adquiridos da QIAGEN.

As enzimas de restrição, BamHI e NotI foram adquiridas da New England BioLabs, quanto a NotI também foi usada uma obtida na Fermentas.

O anticorpo Anti- $\gamma$ -tubulina foi adquirido da Sigma, clone T6557, e o anticorpo secundário Anti-murganho Alexa 594 da Molecular Probes.

O oligo(dT) proveio da Amersham. Os enzimas transcriptase reverso SuperScript II e Platinum® *Pfx* DNA *Polymerase* foram obtidos da Invitrogen. O E.Z.N.A.® Kit *Micro-elute* Total RNA Omega proveio da Bioteck e o pGEM®-T Easy Vector da Promega.

## II.2 Equipamento

As linhas celulares utilizadas neste estudo foram manuseadas numa câmara de fluxo laminar vertical Danlaf VFRS 1206 e mantidas numa estufa Snijders CO190TC-20, com uma atmosfera de 90% (v/v) de humidade e com 5% (v/v) de CO<sub>2</sub>. De forma a visualizar e contar as células recorreu-se a um microscópio invertido de contraste de fase Olympus CK40.

Foram utilizadas várias centrífugas, designadamente a Centrifuge 5415 D da eppendorf, a Sigma 4K10 e a Sigma 202 MK.

Após o processamento das células para imunofluorescência, estas foram observadas num microscópio de fluorescência Leica DMRA2.

As reações de PCR foram realizadas nos termocicladores Mastercycler Personal (Eppendorf) e Peltier Thermal Cycler – 100 (MJ Research). A análise dos produtos de PCR foi conseguida através de um sistema de electroforese de DNA da BIORAD, usando uma fonte de tensão Bio Rad PAC200. As respetivas bandas de DNA foram visualizadas e fotografadas num transiluminador (UV) Kodak EDAS 290.

Para se quantificar as amostras de RNA total e DNA foi utilizado um espectrofotómetro que permite analisar volumes pequenos de soluções de ácidos nucleicos, na ordem dos µls, o Nanodrop (ND-1000 spectrophotometer, *software* ND-1000 V3.1.2, Nanodrop Technologies).

## **II.3 Linhas celulares e cDNAs**

Para a realização deste trabalho foram utilizadas quatro linhas celulares humanas, nomeadamente HEK 293T, HeLa, MCF-7 e h-TERT-RPE-1. A linha celular HEK 293T provém do epitélio de rim embrionário humano, HeLa é uma linha celular epitelial de adenocarcinoma cervical humano e MCF-7 é uma linha celular que provém de adenocarcinoma mamário humano. Por fim, a h-TERT-RPE-1 é uma linha celular do epitélio pigmentado imortalizada pela expressão da proteína telomerase humana. Todas elas apresentam uma morfologia epitelial e um crescimento aderente.

A linha celular HEK 293T apresenta uma grande eficiência de transfeção com plasmídeos e por isso o interesse em usá-las neste estudo.

Foram ainda usados cDNAs correspondentes às linhas celulares HTB-20 e MDA-231, as quais provém do cancro mamário humano. Os cDNAs foram gentilmente cedidos pela Doutora Carla Real, Investigador do Centro de Química e Bioquímica, Faculdade de Ciências e Tecnologia da Universidade de Lisboa.

Neste trabalho, para além das linhas celulares foram também usados cDNAs de órgãos humanos, nomeadamente de pulmão, fígado, timo, músculo liso, testículos, fígado de feto, coração, próstata, medula espinhal, placenta e baço, os quais existiam no grupo provenientes do trabalho do Doutor João Gonçalves.

### **II.3.1 Condições de cultura celular**

As linhas celulares HEK 293T, HeLa e MCF-7 foram cultivadas em meio de cultura DMEM com Glutamax, suplementado com 10% (v/v) de soro fetal bovino (FBS) e aminoácidos não essenciais.

A linha celular RPE-1 foi cultivada em meio de cultura DMEM/F-12 com Glutamax, suplementado com 10% (v/v) de FBS, aminoácidos não essenciais e 0,348% (m/v) bicarbonato de sódio. Todas as células foram mantidas numa estufa a 37°C, com 5% CO<sub>2</sub> e atmosfera húmida.

De forma a manter as células em fase de crescimento exponencial e em monocamadas sub-confluentes, estas foram passadas a cada 2-3 dias dependendo da sua confluência. No caso da confluência das células se apresentar abaixo dos 60-70%, apenas se mudava o meio de cultura por meio fresco e pré-aquecido a 37°C. Como as linhas celulares usadas neste estudo crescem aderentes na superfície da placa de cultura, durante a passagem de células foi necessário tratá-las com tripsina 0,05% durante cerca de 3 minutos a 37°C. Este passo é designado por tripsinização e permite que as células aderentes sejam removidas e fiquem em suspensão, possibilitando a sua passagem e contagem, se necessário. A reação foi parada por adição do meio cultura e as células foram plaqueadas numa nova caixa.

#### II.3.1.1 Contagem de células

Para os diferentes trabalhos experimentais é fundamental ter um número preciso de células, por isso após lavagem com PBS 1x e tratamento com tripsina, é necessário contá-las, recorrendo a um hemocitómetro. Para a contagem de células vivas foi usado o marcador vital azul de tripano, que cora de azul tecidos e células mortas. Assim e de acordo com a experiência a realizar as células vivas foram contadas seletivamente e distribuídas em placas de cultura.

### **II.4 Extração de RNA total de linhas celulares**

Para a extração dos RNAs das linhas celulares HeLa, MCF-7 e h-TERT-RPE-1 foi usado o E.Z.N.A.® *kit Micro-elute Total RNA* da OMEGA bio-tek, que permite a extração de RNA a partir de um número reduzido de células utilizando mini-colunas de alta velocidade com matriz reversível. As células foram destacadas das placas de cultura ( $\pm$  1,5 milhões de células) com o auxílio de raspadores após terem sido lavadas com PBS (NaCl 137mM; KCl 2,7 mM; Na<sub>3</sub>PO<sub>4</sub> 10mM; KH<sub>2</sub>PO<sub>4</sub> 1,8 mM) 1x concentrado. Estas são recolhidas em 3 mL de PBS e centrifugadas a 300 g durante 1 minuto. O sedimento de células obtido foi ressuspensionado em 300µl de solução tampão lise presente no kit. Centrifugou-se a 10 000g durante 2 minutos e adicionou-

se ao lisado igual volume de etanol a 70%, seguidamente aplicou-se a amostra numa coluna *MicroElute*<sup>™</sup> RNA. Esta foi centrifugada a 10 000g durante 15 segundos. Após passar a coluna para um novo microtubo coletor, adicionou-se 500 µl da solução tampão RWF e centrifugou-se a 10 000g durante 30 segundos. Por último, a coluna foi lavada através da adição da solução tampão de lavagem e centrifugada nas mesmas condições. O RNA total é eluído e conseqüentemente obtido no sobrenadante.

## **II.5 Síntese de cDNA**

Para se efetuar a síntese de cDNA foram usados 8µl de cada amostra de RNA total obtido como descrito na secção II.4. Antes da síntese do cDNA as amostras de RNA total foram tratadas com DNase I de acordo com as instruções do fornecedor, para eliminar a hipótese de contaminação e só depois este foi convertido em cDNA através da utilização da enzima transcriptase reversa, SuperScript II e do “primer” oligo(dT)<sub>12-18</sub>.

## **II.6 Pesquisa de possíveis transcritos alternativos do gene *tbccd1* por amplificação específica de cDNAs por PCR**

No geral, foi usada sempre a mesma estratégia para a pesquisa de cDNAs correspondentes a transcritos alternativos do gene *tbccd1*.

Assim a presença destes foi investigada por amplificação, através da técnica de PCR, a partir de amostras de cDNAs sintetizadas a partir de RNA total extraído de linhas celulares e tecidos. Para a amplificação das sequências com interesse foram desenhados *primers* capazes de emparelhar especificamente com sequências nucleotídicas dos transcritos do gene *tbccd1*.

Inicialmente, a análise detalhada da sequência genómica do gene *tbccd1* permitiu localizar um possível codão de iniciação (ATG) alternativo ao previamente caracterizado (Gonçalves, Nolasco et al. 2010a). Deste modo, desenhou-se um *primer* cuja sequência continha este codão de iniciação putativamente alternativo.

Após a averiguação de seqüências nucleotídicas indicadoras de possíveis locais de *splice* foram desenhados *primers* em dois exões diferentes, logo a seguir ao exão onde se localiza o ATG previsto. Este de forma a tornar a sua compreensão mais acessível foi denominado de exão 1 e os seguintes, onde se localizam os *primers*, designados de exão 2 e exão 3 (Figura 17) (Tabela 1).

```

1 GTTCTCCTCTCACCCCACTAGCCTGGCCATACTA000CTT0CT00G0C0A0000G0C0A0C 60
61 CCACAGCGATGCCCTCTCCACTCCGTCCAGTCTGGCAATGCCACTTCC00CACC000CTCC 120
121 CAC0000CAGCAAAGGCTGACCCACAGGCC000GTTGTCTGCGAGTGTCTGTTCTCACC0C 180
181 GGCAATCACCGCCCGATGAGGTATAGCAGCAACAGGCCAAAAGGTGCTCAGGTTCTGCGG 240
241 AGCCATGGTTCCTCTGTCCCGGGTCTGTTCACACACTCCTCACAGCCTCCTCCG000C 300
301 CGCGTCGGCTCGCCAGCCAC00GG00CTGTGAGGG00G0CTCACACC0GGGGGG0G0G0G 360
361 GGTCTCCTTGGTCCGAGAGACAGCTAGCTAGCC00CTCTTCTG00GGTCCACTTCC0GG 420
421 GAGTCCCGAGTCCCGGTGAGAGAGCC00CGGTAGGCGAGAGCCAAATCGTTGCC0GGGAC 480
481 ATCACACGCGGCGCG00CTGGAGCCAAATCAGCGCAGTACACTTCACCTCCGAGGCGGC 540
541 ACGGGCAACCAGTTAGGCTTCAGAATGCACTCGCAGCTCTTGACCAATTAGAAGCGGAG 600
601 ACGTCAGAGGGT0GCG00GCTGCAGGAAC00GCTCTGCCTGAGCTAC00GAGT0GTAGA 660
661 AAATTACGTAATGATAGGGGGCCGGTCAGAGTGAGGGCA0CTGAGGTACACGAAATCCCT 720
721 CCGGCTTGCCCACTTCTCGGACTTTCACCAATCAGGAGCTTCTCC00CAGCACCA0C 780
781 AATTGTGGGAGAGAAGTTTCTCTGTCTACAAGCC00GGAATTCATTATAC00CAGCTA 840
841 CGCTCCGAATGAGGGAGCAGAGGTACAGCGCC00GAATGAG00GGAA0CAGTTATTCACGA 900
901 CAGGTGTACATCAGACGCATTCCTTAG00CTCTTATTTAG0TTCTCCAAGCAACTCCAT 960
961 AAGACAACGTCTCGGAAAGGCTACAGTGATGACAGACCGACTTAAGTT0GGATT0GGCT 1020
1021 TCCACCGCTTCTTAGTGTGAGACCC00CAC000CTCC00CTCCCAAGTCTGAGCCTAA 1080
...
3781 GCGGGCGCGCGAGGAGGGTGAGGGGGGGCC00CAGCGAG00GACGG00GG00GG00GG0C 3840
3841 GCGGAGAAGTGCGGGCGAGCGCGCC00TGCATTAGCAGGTAACCGGCACTGGGTGCGCCG 3900
...
6721 CATAAATGAGACTTGTITCAAATCAGTITATGAAGAACTTGTITCCTAGSTATGCAA 6780
6781 GAAGCTTTTCA000TGATGT0CTTACAGATAATATGATCAGTCCAGATTCT0CTCTG 6840
6841 GGTGAAGCAGAA000TTATAGTGGGTGCCTTGAGGT00000CTCCATCCAAGTTTAG 6900
6901 TCTTCACTATCTCAGGAAGATATCCACCTATGTGCAAAATCCGGG00CACAGAAGGAGCTTA 6960
6961 C000CG00CTCTACTGGTCTACATGGAGGCACATCGCTTGTGGGAAGCTCCAGTTGCCAA 7020
7021 GGACCTGG0GTGCCTTACTT0GAAATATTTGATAGTCTTCAATGAAGCAC0CTGAGGA 7080
7081 GCG0CTGGAATGGTCTGAGGTCTGTCCA0CTGCATGTCTGAGGAGGAAGTTGAAAAGCA 7140
7141 GAGAAATCAGGTACGGATTGATATGTGTATACTTAATTACTTTTGGAAAGAA0CAAT 7200
...
12541 TCTAATCATCTTTTATTTCTTCTTGGACAGCTTTCAGTGGACA0GCTACAGTTTCTGCT 12600
12601 CTTCTTATACATCAACAGTTGACAAAGGTCTCC0TAAGGACATCTTTGATTGG0GAAGA 12660
12661 GTGGCC0CAGTCC0CAGAA0AAATCTCAGTCTCCTGAC0CTGACTGAAAATCTAATTTGTC 12720
12721 TAATAAGGTACTCCTTATATCTTGATCTCATTCTCAAAAAGGACAATTTGTGTGGGAGA 12780

```

**Figura 17: Sequência parcial do cromossoma 3 humano, onde está localizado o gene *tbccd1*. As seqüências com a letra vermelho escuro correspondem aos exões do TBCCD1. Por sua vez, as seqüências sublinhadas correspondem aos *primers* que foram desenhados para determinar a possível**

existência de um transcrito alternativo. O codão de iniciação (ATG) previsto encontra-se destacado a verde, localizando-se no exão 1, enquanto o codão de iniciação já descrito em estudos anteriores encontra-se destacado a azul, localizando-se no Exão 2 (Gonçalves, Nolasco et al. 2010a). Após uma análise detalhada da estrutura do gene, e de forma a pesquisar a existências de outros possíveis intrões identificou-se os dinucleótidos (destacados em figuras retangulares) invariáveis que eventualmente poderiam flanquear esses mesmos intrões. Esta sequência foi obtida recorrendo à ferramenta informática Ensembl Genome Browser, no seguinte site [www.ensembl.org](http://www.ensembl.org) (Gene: TBCCD1 ENSG00000113838).

**Tabela 1:** Sequências dos *primers* desenhados para a pesquisa de possíveis transcritos alternativos do gene *tbccd1*.

<i>Primer</i>	<i>Sequência</i>
TBCCD1 <sub>ATG alternativo</sub> <i>Forward</i>	5'-TGTCTCGGAAAGGCTACAGTGATG-3'
TBCCD1 <sub>exão 1</sub> <i>Reverse</i>	5'-ACAAGCGATGTGCCTCCATGTA-3'
TBCCD1 <sub>exão 2</sub> <i>Reverse</i>	5'-GTCAGGTCAGGAGACTGAGATTTG-3'

**Tabela 2:** Condições das reações de amplificação por PCR, usadas na pesquisa de transcritos alternativos do gene *tbccd1*.

<i>Iniciação</i>	<i>Desnaturação</i>	<i>Annealing</i>	<i>Elongação</i>	<i>Elongação final</i>
95°C	94°C	55°C	72°C	72°C
5 min	30 seg	30 seg	30 seg	10 min
30 ciclos				

Em suma, após o desenho dos *primers* descritos na tabela 1, estes foram utilizados com o objetivo de pesquisar se correspondiam a possíveis transcritos alternativos do gene *tbccd1*. Esta pesquisa foi feita recorrendo-se aos vários cDNAs das linhas celulares e de tecidos humanos referidos na secção II.3. As reações de amplificação por PCR ocorreram, na sua maioria, segundo as condições descritas na tabela 2.

De modo a determinar as sequências nucleotídicas dos produtos amplificados foi necessário analisá-los por sequenciação automática (método explicado em maior detalhe posteriormente). Para isso, após a amplificação as reações de amplificação foram analisadas em gel de agarose (1-1,5%) (m/v) e a banda correspondente ao produto pretendido foi cortada e o DNA extraído e purificado utilizando o kit QIAquick Gel extraction de acordo com o protocolo do fabricante. Seguiu-se a ligação (cuja descrição se encontra detalhada na secção II.8.4) do produto de DNA purificado ao vector pGEM®-T Easy Vector, durante a noite a 4°C, também de acordo com as instruções do fabricante. Posteriormente, as reações de ligação foram usadas para transformar bactérias competentes (preparadas como mais à frente se descreve) de *Escherichia coli* da estirpe JM109. As bactérias transformadas foram semeadas em meio LB sólido com o antibiótico para o qual o vector possui um marcador de resistência (ampicilina), de forma a serem selecionados os clones recombinantes. Após a obtenção de colónias de bactérias recombinantes, geralmente eram selecionadas 2 ou 3 colónias, as quais eram inoculadas em meio LB líquido (3 a 5 mL), suplementado com ampicilina (100 µg/mL). Estas cresceram durante a noite, com agitação contínua, a 37°C. No dia seguinte, a extração do DNA plasmídico, foi feita usando o kit NZYMiniprep ou QIAprep Spin Miniprep. Por último, o DNA extraído foi precipitado e analisado por sequenciação automática.

As sequências nucleotídicas resultantes do processo de sequenciação foram analisadas com a ajuda do *software* Chromas Lite.

## **II.7 Estirpes bacterianas e condições de cultura**

Neste estudo, para a amplificação de DNA plasmídico foi utilizada estirpe de *E. coli* JM109 (genótipo: endA1 glnV44 thi-1 relA1 gyrA96 recA1 mcrB+ Δ(lac-proAB) e14- [F' traD36 proAB+ lacIq lacZΔM15] hsdR17(rK-mK+)).

A estirpe foi cultivada em meio de cultura LB líquido (bactopeptona 10 g/l; extracto de levedura 5 g/l; NaCl 5 g/l) ou sólido [com 1,5% (m/v) de agar] suplementado com o antibiótico adequado (ampicilina 100 µg/mL), ambos a 37°C e no caso das culturas em meio líquido com agitação. De forma a conservar as células durante grandes

períodos de tempo estas foram congeladas quando se encontravam na sua fase exponencial, a  $-80^{\circ}\text{C}$ , em meio LB suplementado com 15% glicerol (v/v).

### **II.7.1 Preparação de células de *E. coli* competentes**

Com o intuito de tornar as células de *E. coli* competentes para poderem ser transformadas com DNA plasmídico foi usado protocolo que a seguir se descreve. Colocou-se uma pré-cultura a crescer em meio líquido durante a noite a  $37^{\circ}\text{C}$  e com agitação contínua. Na manhã seguinte, preparou-se uma cultura em meio líquido a partir da pré-cultura, tendo esta sido semeada a 0,075 OD a uma absorvância a 600 nm ( $A_{600}$ ) e incubada a  $37^{\circ}\text{C}$  com agitação permanente. Após atingir o início da fase exponencial (0,3OD a  $A_{600}$ ) as células foram recolhidas por centrifugação a 7000 rpm numa centrífuga *ependorf* durante 2 minutos a  $4^{\circ}\text{C}$ . As etapas seguintes ocorreram sempre à temperatura de  $4^{\circ}\text{C}$  e o material/reagentes usados foram também previamente refrigerados a  $4^{\circ}\text{C}$ . O sedimento celular formado foi ressuscitado em  $\frac{1}{2}$  do volume inicial da cultura de células numa solução de  $\text{MgCl}_2$  0,1M e esta suspensão foi novamente centrifugada a 7000 rpm numa centrífuga *ependorf* durante 2 minutos. Por sua vez, o sedimento de células obtido foi ressuscitado em  $\frac{1}{2}$  do volume inicial de uma solução de  $\text{CaCl}_2$  0,1M, e deixado a repousar no gelo durante 20 minutos. Seguidamente realizou-se uma nova centrifugação das células nas mesmas condições, sendo o sedimento de células ressuscitado em  $\frac{1}{15}$  do volume inicial, numa solução de  $\text{CaCl}_2$  0,1M com glicerol a 15% (v/v).

As células *E. coli* tornadas competentes por este processo foram rapidamente congeladas e guardadas em azoto líquido, a  $-80^{\circ}\text{C}$ , até serem precisas para serem transformadas com de DNA plasmídico.

## **II.8 Clonagem de fragmentos de DNA em vectores de expressão**

### **II.8.1 Amplificação de sequências DNA por PCR**

De forma a clonar fragmentos de DNA em vectores de expressão em células de mamífero recorreu-se inicialmente à técnica de PCR para assim os amplificar. Com o objetivo de diminuir os erros durante a amplificação dos fragmentos contendo as sequências codificantes dos genes a clonar, recorreu-se ao uso de polimerases de DNA com atividade de revisão de provas (*proofreading*) (Platinum® *Pfx* DNA Polymerase) em reações de PCR com um volume final de 25 ou 50 µL [30 ng/µl de cada *primer*, tampão de PCR 1x, 2mM dNTPs] a partir de amostras de cDNA (h-TERT-RPE-1 e MCF-7). Sempre que necessário, as sequências dos *primers* utilizados nas reações de amplificação foram desenhadas de forma a incluir locais de reconhecimento de enzimas de restrição. Estes locais de reconhecimento permitem a hidrólise dos fragmentos amplificados, de forma a gerarem extremidades adequadas à sua inserção e ligação aos DNAs dos vectores de clonagem, que foram previamente hidrolisados com as mesmas enzimas.

### **II.8.2 Análise de DNA por electroforese em gel de agarose**

As electroforeses onde se analisaram os fragmentos de DNA amplificados por PCR foram realizadas em géis de agarose de 1 ou 2% (m/v), de acordo com o tamanho dos fragmentos a observar, preparados em tampão de electroforese TAE 1X (Tris 40mM, EDTA 1mM pH 8,3) com GreenSafe a uma concentração de 0,0025%. Antes da aplicação das amostras no gel foi adicionado tampão de amostra (azul de bromofenol 0,2% (m/v) em TAE 1X). A migração realizou-se em tampão de electroforese 1X TAE e geralmente a voltagem utilizada foi entre 80-100V. Foi usado, sempre que necessário, um marcador de massa molecular de DNA, o NZYLadder III, de forma a confirmar se o tamanho dos fragmentos analisados eram o esperado. Por fim, as bandas que correspondiam às massas moleculares esperadas para os fragmentos de DNA de interesse foram cortadas a partir do gel de agarose e de seguida purificadas,

usando um kit que permite a extração de DNA do gel, *QIAquick Gel extraction kit*. Seguiu-se o protocolo de acordo com as instruções do fornecedor.

### **II.8.3 Hidrólise dos produtos de PCR com enzimas de restrição**

Na reação de hidrólise, os produtos resultantes da reação PCR foram hidrolisados com os enzimas de restrição *BamHI* e *NotI*, de forma a estes poderem ser inseridos num vector de expressão e portanto o vector de expressão foi também hidrolisado com estes mesmos enzimas de restrição. O volume final da reação era geralmente 30µL, contendo 1µL de cada enzima (1U) e tampão de reação 1x, recomendado pelo fabricante para os enzimas em causa. Na reação de hidrólise dos produtos de PCR, a quantidade de DNA usada era variável, pois fazia-se uma estimativa, por excesso, para que na reação de ligação se conseguisse usar uma quantidade igual a 120ng. Quanto à hidrólise do vector, a quantidade de DNA usada também eram variável, desde que no fim ficasse a uma concentração de 40ng/µl. As digestões foram efetuadas a 37°C, durante a noite.

### **II.8.4 Ligação do fragmento de DNA ao vector de expressão**

Para ligar os fragmentos de DNA de interesse ao vector, por norma usou-se uma proporção de 1:3, ou seja, utilizaram-se 40ng de vector e 120ng de fragmento, tendo em conta um tamanho médio dos produtos de PCR. As reações ocorreram num volume total de 10 µL e foram geralmente incubadas a 4°C durante a noite. A ligação dos fragmentos ao vector foi feita com o enzima ligase de DNA do fago T4, de acordo com as instruções do fabricante.

### **II.8.5 Transformação de células competentes**

Para cada reação de transformação foram usados 100µl de células competentes, às quais se juntou as reações de ligação (10µl) ou o DNA plasmídico (~100ng) e de seguida estas foram incubadas no gelo durante 20 minutos. Após estes 20 minutos as

células foram submetidas a um choque térmico a 42°C durante 90 segundos e seguido de 2 minutos no gelo. No final deste período adicionaram-se 600µl de meio líquido LB e incubou-se durante 1 hora a 37 °C com agitação. Terminada a incubação as células foram centrifugadas durante 2 minutos a 7000 rpm numa centrífuga *ependorf* e desprezou-se o sobrenadante. Seguidamente, ressuspendeu-se o sedimento de células no meio restante e posteriormente plaqueados em meio LB sólido suplementado com o antibiótico, neste caso ampicilina (100 µg/mL), para o qual o vector possui um marcador de resistência, de forma a seleccionar os recombinantes e em seguida foram deixadas a crescer durante a noite numa estufa a 37°C.

### **II.8.6 Protocolo de *Cracking***

Após o crescimento de bactérias recombinantes, algumas colónias foram escolhidas de forma a verificar se eram de facto positivas, ou seja, se estavam transformadas com vectores recombinantes, contendo inserido o fragmento de interesse. Para este fim, utilizou-se o método de *Cracking* consiste em lisar um pouco de uma colónia em 20µl da solução que é composta por NaOH 50mM, SDS 0,5% (m/v), EDTA 5mM. Seguidamente os lisados foram incubados durante 30 minutos a 55°C e terminado este tempo, foram levados ao vortéx à velocidade máxima durante um minuto para se quebrar o DNA genómico. Por último, a cada lisado foi adicionado um volume apropriado de tampão de amostra e estes foram analisados em gel de agarose 1% (m/v), tal como o vector vazio que serve de controlo. Por diferenças na migração relativamente ao vector vazio, este método permite estimar quais os clones positivos, pois a inserção de um fragmento no vector aumenta a sua massa molecular o que se traduz num atraso da migração face ao vector vazio.

### **II.8.7 Extração e purificação de DNA plasmídico recombinante em pequena escala**

De forma a obter DNA plasmídico recombinante em pequena escala selecionaram-se as colónias que correspondiam aos clones positivos (anteriormente identificados pelo método de *cracking*). Para isso, lançaram-se pequenas culturas, sendo que cada uma resulta da inoculação de uma única colónia, em meio LB líquido (3 a 5 mL), suplementado com ampicilina (100 µg/mL). Estas cresceram durante a noite em agitação, a 37°C. No dia seguinte as culturas foram centrifugadas numa centrífuga *ependorf* e os sedimentos de células foram recolhidos. De seguida, usou-se o kit de extração de DNA NZYMiniprep ou QIAprep Spin Miniprep, para a extração do DNA plasmídico, conforme as instruções do fabricante.

Realizou-se a hidrólise do DNA plasmídico com os enzimas de restrição adequados (capazes de remover o fragmento clonado), sempre que necessário, de forma a analisar a qualidade do DNA purificado e visualizar o fragmento inserido através de uma análise em electroforese em gel de agarose.

### **II.8.8 Sequenciação automática de DNA**

Após a identificação dos clones positivos e de forma a determinar as suas sequências nucleotídicas, estes foram analisados por sequenciação automática com o intuito de garantir que as sequências clonadas foram inseridas corretamente e que não tinham ocorrido erros no processo de amplificação.

Em cada reação de sequenciação usou-se cerca de 500ng do DNA molde e 30 *pmol* do *primer* específico. As reações de sequenciação foram feitas por PCR, no total de 25 ciclos compostos por 10 segundos a 96°C, 5 segundos a 50°C e 4 minutos a 60°C. Seguidamente, precipitou-se o DNA com 1/5 do volume inicial de acetato de sódio (CH<sub>3</sub>COONa) 3M e 5 volumes de etanol 95%, incubando 30 minutos à temperatura ambiente. Os produtos precipitados foram centrifugados numa centrífuga *ependorf* durante 30 minutos a 4°C à velocidade máxima (14000 rpm). O sobrenadante foi desprezado e ao sedimento de DNA adicionou-se etanol 70%, sendo novamente centrifugado a 14000 rpm numa centrífuga *ependorf* durante 15 minutos a 4°C. No

fim, desprezou-se o sobrenadante e o sedimento de DNA foi deixado a secar ao ar. A análise dos sedimentos de DNA foi realizada pelo serviço de sequenciação do Instituto Gulbenkian de Ciência.

Como já referido anteriormente as cromatogramas e as sequências nucleotídicas foram analisadas com a ajuda do *software Chromas Lite*.

### **II.8.9 Produção de DNA plasmídico em grande escala**

A sequenciação permitiu comprovar a qualidade da sequência nucleotídica dos plasmídeos recombinantes obtidos e por isso após esta confirmação, o seu DNA foi preparado em grande escala. Assim, o DNA plasmídico recombinante foi usado para transformar células competentes de forma a ser amplificado. As células transformadas foram inoculadas em meio LB líquido (100ml) suplementado com ampicilina. As culturas cresceram durante a noite a 37°C com agitação e por fim usadas para extração do DNA plasmídico com os kits Plasmid Midi Kit (QIAGEN) e NZYMidiprep.

### **II.9 Clonagem dos cDNAs alternativos do gene *tbccd1* em vectores de expressão para células de mamíferos**

Após a pesquisa e posterior identificação dos transcritos alternativos do gene *tbccd1*, estes foram produzidos em células de mamífero. Desta forma, os transcritos alternativos do gene *tbccd1* foram amplificados a partir de uma amostra de cDNA da linha celular RPE-1 e MCF-7. Para este fim, foram desenhados *primers*, que não só permitissem a amplificação dos transcritos alternativos do gene *tbccd1*, mas também a sua clonagem em fusão com o *tag* GFP. Como explicado anteriormente, em cada *primer* foi introduzido um local de reconhecimento de um enzima de restrição para se poderem hidrolisar as extremidades dos produtos de amplificação obtidos e se efetuar a clonagem dirigida dos mesmos nos vectores de clonagem (Tabela 3). As reações de amplificação por PCR foram realizadas consoante as condições descritas na tabela 4. A clonagem das sequências alternativas da proteína TBCCD1 em fusão

com a proteína GFP foi feita no vector de expressão pIC111 (Cheeseman and Desai 2005) nos locais de restrição BamHI e NotI do mesmo. Este permite a fusão do “LAP tag” 6xHis-PreScission-GFP à região C-terminal da proteína em estudo.

**Tabela 3:** Sequências dos *primers* utilizados na clonagem dos transcritos alternativos do gene *tbccd1*, sendo que as sequências sublinhadas correspondem aos locais de reconhecimento dos enzimas de restrição utilizados na estratégia de clonagem.

<i>Primer</i>	<i>Sequência</i>
<i>Forward (BamHI)</i>	5' <u>CGCGGATCCT</u> GTCTCGGAAAGGCTACAGTGATG 3'
<i>TBCCD1 Reverse (NotI)</i>	5' AGGAAAAA <u>AGCGGCCG</u> CAGCTGCTTGTGGAGC 3'

**Tabela 4:** Condições das reações de amplificação por PCR, usadas na estratégia de clonagem.

<b>Iniciação</b>	<b>Desnaturação</b>	<b>Annealing</b>	<b>Elongação</b>	<b>Elongação final</b>
<b>94°C</b>	94°C	55°C	68°C	72°C
<b>5 min</b>	15 seg	30 seg	1 min/Kb	10 min
35 ciclos				

## **II.10 Transfecção de células de mamífero com vetores de expressão contendo transcritos alternativos do gene que codifica para a proteína TBCCD1**

Para as transfeções as células HEK 293T foram sempre semeadas em placas de 24 poços (1x10<sup>6</sup> células por poço), 18h a 24h antes da transfecção. Como as células se destinavam ao processamento para visualização por microscopia de imunofluorescência estas foram semeadas sobre lamelas de vidro revestidas com Poli-L-lisina para melhor adesão. Para a transfecção recorreu-se ao reagente de lipofecção (Lipofectamine 2000), sendo que este foi usado de acordo com o protocolo do fabricante. Para cada poço, a quantidade de Lipofectamine 2000 usada foi de 2µl e a quantidade de DNA foi de 1µg. De forma breve, a Lipofectamine 2000

foi diluída em Optimem e incubada durante cinco minutos à temperatura ambiente. Em simultâneo, o DNA plasmídico foi também diluído em Optimem. De seguida, misturou-se a Lipofectamine 2000 ao DNA plasmídico e estes foram incubados em conjunto durante 20 minutos à temperatura ambiente de modo a formarem os complexos DNA:agente lipídico. Neste espaço de tempo o meio de cultura foi substituído por meio de cultura sem soro, visto que o soro inibe o processo de lipofecção. Por último, a mistura Lipofectamine 2000:DNA foi adicionada às células. Estas foram incubadas durante cerca de 4/6 horas e seguidamente o meio de cultura foi removido e substituído por meio de cultura suplementado com soro para parar a transfecção. Após a paragem da transfecção, as células foram incubadas em meio completo durante cerca de 18h a 24h para permitir a expressão do plasmídeo. No final deste tempo as células foram processadas para imunofluorescência.

## **II.11 Microscopia de Imunofluorescência**

No final do período de transfecção, para o processamento das células HEK 293T para imunofluorescência, o meio foi aspirado e as células aderentes às lamelas lavadas com PBS. De seguida as células foram fixadas e permeabilizadas com metanol 100% e gelado, durante 10 minutos a -20°C. Finda a fixação, aspirou-se o agente fixador e lavaram-se as células duas vezes com PBS (500µl), durante 5 minutos e uma lavagem rápida com PBS-Tween 0.1% (v/v). Seguiu-se o bloqueio das células com uma solução de 0,3 % BSA (albumina de soro bovino) (m/v) em PBS, durante 30 minutos à temperatura ambiente. De seguida fez-se a incubação com o anticorpo primário, anti- $\gamma$ -tubulina (tabela 5), diluído na solução de bloqueio (40µl) durante 1 hora à temperatura ambiente. Terminada a incubação as células foram lavadas duas vezes com PBS (500µl), durante 5 minutos e uma lavagem rápida com PBS-Tween 0.1% (v/v). Seguiu-se, uma vez mais, uma hora de incubação com o anticorpo secundário, anti-murganho, diluído na solução de bloqueio (40µl). Após a incubação as células foram novamente lavadas como descrito anteriormente. Finalmente, as lamelas foram montadas numa lâmina de vidro sobre uma gota de meio de montagem Dapi Fluoromount-G, o qual também permite fazer a marcação do DNA, pois apresenta na sua constituição a solução de DAPI.

Após a marcação, as células foram observadas num microscópio de fluorescência Leica DMRA2. Todas as imagens resultam de projeções em Z, de forma a obter *stacks* que contemplassem toda a célula, para não falhar o centróssoma.

O processamento de imagens obtidas por microscopia de fluorescência foi conseguido recorrendo ao uso do *software Image J*.

**Tabela 5:** Anticorpos primários e secundários usados na análise de microscopia de imunofluorescência e respetivas diluições.

	<b>Anticorpo Primário</b>	<b>Anticorpo Secundário</b>
	Anti- $\gamma$ -tubulina (Sigma, clone T6557)	Anti-murganho Alexa 594 (Molecular Probes)
<b>Diluição</b>	1:200	1:500
<b>Animal em que foi produzido</b>	Murganho	Cabra

## Capítulo III – Resultados e Discussão

### III.1 O gene que codifica para a proteína TBCCD1 apresenta um putativo codão de iniciação alternativo

Os estudos em torno da função biológica do TBCCD1 ainda se encontram numa fase muito precoce, nomeadamente o seu possível envolvimento em processos dependentes dos microtúbulos e ancoramento do centrossoma ao núcleo.

O TBCCD1 encontra-se localizado no centrossoma, no fuso mitótico, no corpo médio e no corpo basal de cílios primários e motores, parecendo assumir um papel essencial para o correto posicionamento do centrossoma no centro da célula em estreita associação com o núcleo, e portanto na organização interna da célula (Gonçalves, Nolasco et al. 2010a).

Sabe-se que o centrossoma é o principal centro organizador de microtúbulos nas células animais, regulando a organização espacial e nucleação dos microtúbulos, ocupando um volume pequeno e encontrando-se próximo do centro da célula em estreita relação com o núcleo (Bettencourt-Dias and Glover 2007; Lawo, Hasegan et al. 2012). O centrossoma está ainda envolvido na regulação da mobilidade celular, a adesão e polaridade na interfase e facilita a formação dos fusos mitóticos durante a mitose (Andersen, Wilkinson et al. 2003; Chae, Yun et al. 2005; Bettencourt-Dias and Glover 2007).

Resultados não publicados por investigadores do nosso grupo, nomeadamente experiências por técnicas de *western blot*, mostravam por vezes a existência de uma proteína com uma massa molecular um pouco mais pequena que a prevista para a proteína TBCCD1. Embora esta banda pudesse surgir devido a inespecificidade das reações anticorpo-antigénio, seria possível que esta pudesse resultar da existência de um transcrito mais pequeno do gene que codifica para o TBCCD1 resultante de *splicing* alternativo. Deste modo, e para aprofundar o estudo da função biológica do

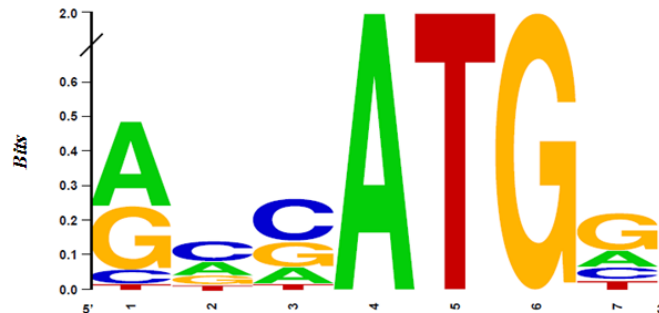
TBCCD1 decidiu-se analisar a relação entre a estrutura do gene *tbccd1* e a sua regulação, especificamente investigando a ocorrência de transcritos alternativos que pudessem originar variantes proteicas do TBCCD1, tendo sido esta busca um dos principais objetivos deste trabalho.

O gene humano que codifica para a proteína TBCCD1 está localizado no cromossoma 3, possui 8 exões e portanto apresenta a sua região codificante interrompida por vários intrões, o que consolida a possibilidade de ocorrer *splicing* alternativo neste gene. De facto, este processo de regulação da expressão dos genes eucarióticos tem maior probabilidade de ocorrer em genes que codificam para proteínas cujos genes que as codificam apresentam a sua região codificante interrompidas por vários intrões (Irimia, Rukov et al. 2007b). Além disso o *splicing* alternativo é um processo que ocorre em mais de 90% dos genes humanos (Luco, Allo et al. 2011).

Para dar início a este estudo recorreu-se a uma análise bastante exaustiva da sequência deste gene (ver figura 17, secção II.6, Capítulo II). Nesta pesquisa verificou-se que além do codão de iniciação já identificado ( $AUG_{TBCCD1}$ ), localizado no exão 2 (assim designado para facilitar a compreensão do texto) identificou-se um possível AUG alternativo ( $AUG_{previsto}$ ) localizado num exão anterior (exão 1). Esta constatação foi tentadora e levou-nos a pensar que eventualmente poderia haver um putativo local de *splice* 5' alternativo, o que resultaria na sua inclusão ou exclusão no mRNA maduro regulando desta forma a expressão de um isótipo mais longo ou mais curto do TBCCD1.

No entanto, para que um AUG possa de facto ser um AUG iniciador este deve apresentar determinadas características para que possa ser reconhecido/escolhido como tal pelos ribossomas. A eficiência do reconhecimento do AUG pelo ribossoma é determinada por vários parâmetros, onde se incluem os nucleótidos que flanqueiam o AUG (e definem o contexto em que este encontra), a proximidade deste ao local *cap* e a presença de estruturas secundárias no mRNA (Morris and Geballe 2000). Quanto aos nucleótidos que flanqueiam o AUG, é de referir que existe um contexto

considerado ótimo para o reconhecimento deste codão pelas subunidades 40S do ribossoma, GCC(A/G)CCAUGG, o qual foi estabelecido por Kozak (Kozak, 1987), sendo as posições destacadas a negrito, -3 e +4, as mais importantes. Um contexto considerado adequado é AnnAUGn ou GnnAUGG, e um contexto mais fraco é (C/T)nnAUG(A/C/T), sendo n um nucleótido qualquer (Figura 18) (Kozak 1987; Kochetov 2008).



**Figura 18: Representação gráfica da sequência conservada dos nucleótidos**, calculada para os sete nucleótidos que rodeiam o tripleto AUG (de iniciação da tradução), desde a posição -3 (ilustrada a na figura pelo número 1) à posição +4 (ilustrada a na figura pelo número 7) para os AUG, em mRNAs humanos (adaptado de Iacono, Mignone et al. 2005).

O AUG<sub>TBCCD1</sub> apresenta um contexto favorável visto que quando comparado à sequência considerada ótima os dois nucleótidos mais importantes estão conservados. Embora o AUG<sub>previsto</sub> apenas apresente um dos nucleótidos mais importantes conservado, e portanto não exhibe um contexto tão favorável, pode-se afirmar que apresenta um contexto considerado adequado (Tabela 6).

**Tabela 6: Representação do contexto nucleotídico considerado ótimo para um AUG iniciador (sublinhado) e do contexto nucleotídico que rodeia o AUG descrito para o *tbcd1* e o AUG putativo analisado neste estudo.** As posições -3 e +4 (negrito) são consideradas as mais importantes.

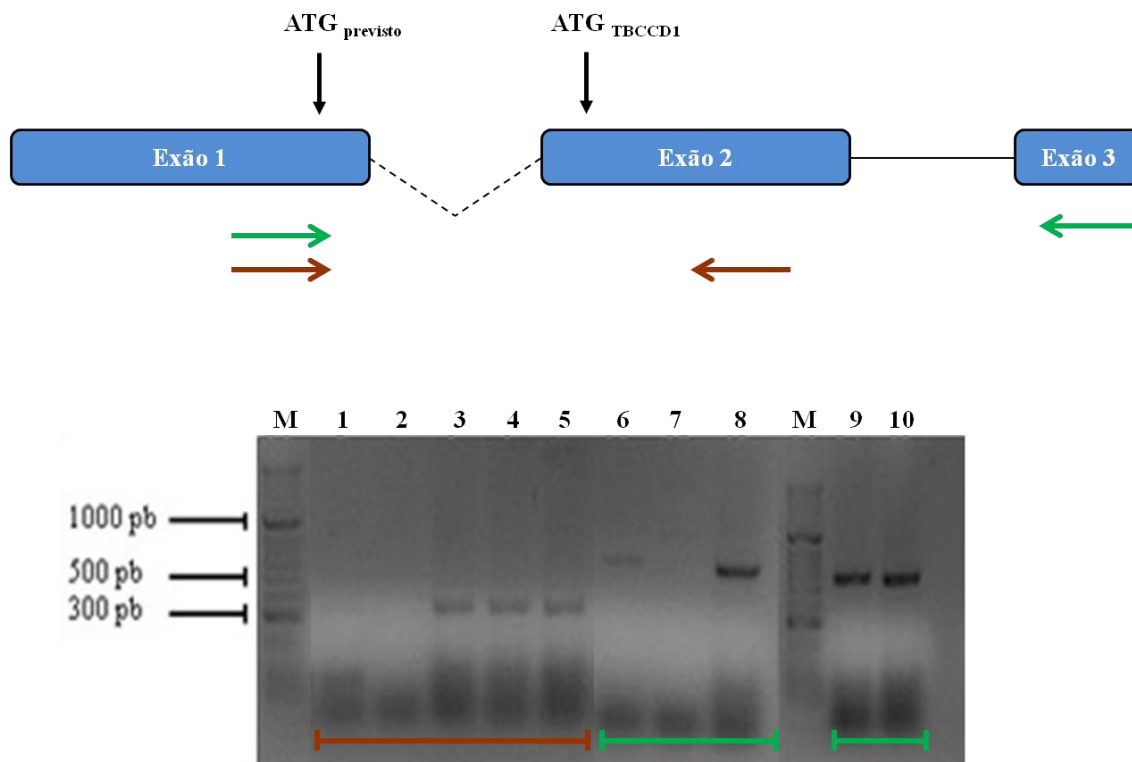
Sequência Kozak	(GCC) GCC (A/G) CCA <u>UGG</u>
AUG <sub>TBCCD1</sub>	AGA GAT <b>A</b> ATA <u>UGG</u> A
AUG <sub>previsto</sub>	GCT ACA <b>G</b> TGA <u>UGAC</u>

Tal como foi descrito detalhadamente nos métodos, o primeiro passo para verificar a existência de um transcrito alternativo funcional cuja região codificante se iniciava no AUG<sub>previsto</sub> foi o desenho de *primers*, nomeadamente um *primer forward* no exão 1, o qual se usou com diferentes *primers reverse* (um no exão 2 e o outro no exão 3). Assim, inicialmente realizou-se uma reação de PCR, na qual se utilizou como DNA molde um cDNA produzido a partir de RNA extraído de várias linhas celulares humanas, designadamente HeLa, RPE-1, HEK 293T, MCF-7, MDA-231 e HTB-20. Esperava-se pois que os produtos de PCR a obter resultariam em diferentes bandas com distintas massas moleculares, correspondente eventualmente a/uma variante/s do TBCCD1.

O objetivo de realizar esta reação de PCR seria a de comparar tamanhos entres as bandas que suportassem a hipótese de haver um local 5' alternativo e por outro lado verificar a ocorrência de um *splicing* alternativo.

Desta forma, o tamanho esperado para o produto de PCR a obter quando os *primers* ilustrados com as setas a castanho na Figura 19 (ver Tabela 1, Capítulo II, Secção II.6, para a sequência nucleotídica) era de 255 pb, ao passo que o tamanho esperado para o produto de PCR a obter quando os *primers* ilustrados pelas setas a verde na Figura 19 era de 200 pb (ver Tabela 1, C, secção II.6, para a sequência nucleotídica).

Estes produtos esperados, com as massas referidas, resultariam das seguintes hipóteses: (1) o transcrito amplificado pelos pares de *primers* ilustrados a castanho deveria conter ambos os AUGs, ou seja o AUG<sub>TBCCD1</sub> e o AUG<sub>previsto</sub>; (2) alternativamente apenas estaria presente o AUG<sub>previsto</sub> o que implicaria que no exão 2, onde se encontra o AUG<sub>TBCCD1</sub>, ocorreria um *splicing* alternativo, como consequência de um salto deste exão (Figura 19).



**Figura 19: Análise de transcritos alternativos do gene *tbccd1* por PCR em cDNAs produzidos a partir de RNA proveniente de várias linhas celulares humanas.** Os *primers* utilizados que se encontram localizados no exão 1 e 2, estão ilustrados pelas setas a castanho (amostras 1-5) e os *primers* localizados no exão 1 e 3, estão ilustrados pelas setas a verde (amostras 6-10). (M) Marcador de massas moleculares; (1 e 6) RPE-1; (2 e 7) HEK 293T; (3 e 8) MDA-231; (4 e 9) HTB-20; (5 e 10) MCF-7. O tamanho esperado para os produtos amplificados utilizando os *primers* ilustrados com as setas a castanho era de 255 pb, ao passo que o tamanho esperado para produtos amplificados utilizando os *primers* ilustrados pelas setas a verde era de 200 pb. A análise da separação electroforética dos produtos de PCR obtidos mostra que não foram estes os produtos de PCR obtidos. De facto, por comparação ao marcador de massas moleculares, os *primers* utilizados amplificam um produto aproximado de 300 pb e 600 pb, respetivamente. As linhas coloridas na parte inferior do gel indicam os pares de *primers* usados como indicado no esquema.

Após a realização deste PCR, verificou-se que as respetivas massas moleculares dos produtos amplificados com os *primers* localizados no exão 1 e exão 2 (a castanho) e no exão 1 e 3 (a verde) eram aproximadamente de 300 pb e 600 pb, respetivamente. Isto levou-nos a concluir que ambos os AUGs iniciadores estavam presentes no transcrito identificado. Desta forma, pensou-se que o transcrito que incluía o

AUG<sub>previsto</sub> poderia codificar para uma variante do TBCCD1, originando uma proteína de maiores dimensões, ou por outro lado ter um papel regulador sobre o AUG<sub>TBCCD1</sub>.

Paralelamente a este PCR efetuou-se um outro que serviu como controlo positivo de forma a verificar a qualidade do cDNA produzido. Para este efeito utilizaram-se *primers* que amplificam parte do gene HPRT (*hypoxanthine phosphoribosyltransferase 1*), um gene que é expresso de uma forma constitutiva, normalmente referido como *housekeeping*. A análise em gel de agarose dos produtos de amplificação obtidos a partir deste PCR, permitiu constatar que a maioria dos cDNAs das linhas celulares encontravam-se em boas condições, visto que no controlo positivo se obteve bandas no tamanho esperado para todas as amostras, exceto para o cDNA produzido a partir da linha celular HEK 293T.

É ainda de referir que os RNAs extraídos das várias linhas celulares foram submetidos a um tratamento com o enzima DNase I e posteriormente testados para uma eventual contaminação com DNA genómico.

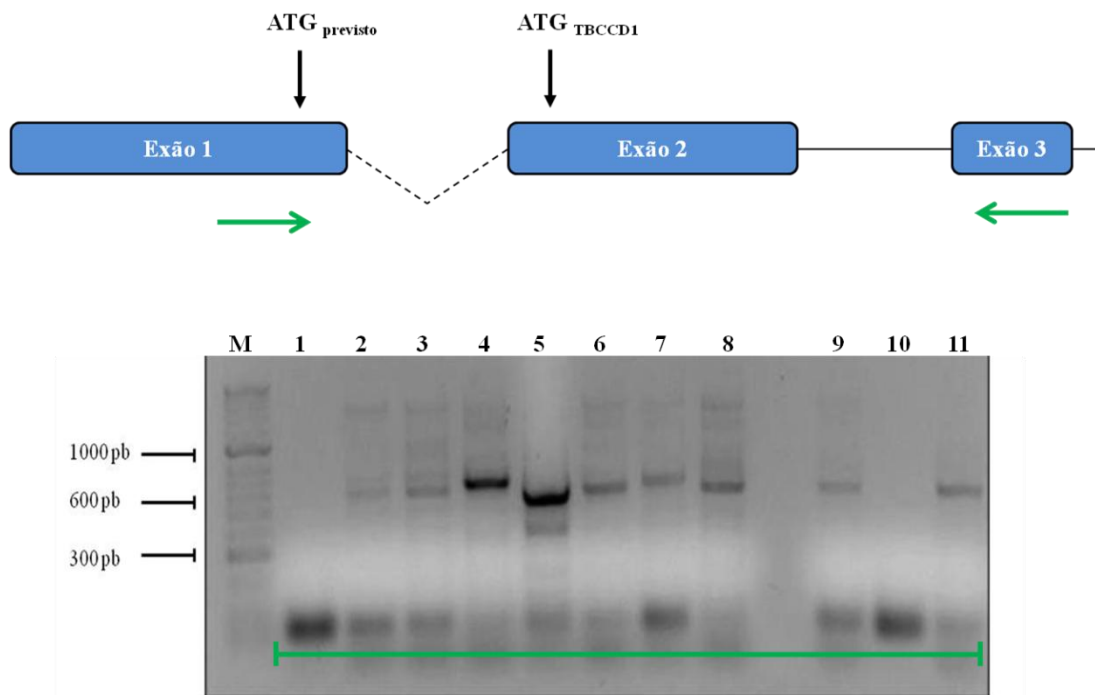
A análise dos resultados obtidos para a amostra 1, correspondente à amplificação que teve como molde o cDNA sintetizado a partir de RNA extraído das células RPE-1, mostra a existência de uma banda muito ténue, esta razão pode ser explicada pelas condições do cDNA. Embora no controlo positivo o cDNA desta linha seja amplificado, sabe-se que quando um *primer* está desenhado na 5' (N-terminal), tal como é o caso, parece ser crítico para um cDNA de má qualidade, uma vez que frequentemente a sua síntese não é terminada corretamente ficando estes truncados, especialmente nos casos em que os transcritos são relativamente grandes como é o caso dos do *tbccd1*. Em conjunto estes factos poderão contribuir de forma negativa para a reação de amplificação por PCR.

Poder-se-ia também colocar a hipótese de que o transcrito contendo o AUG<sub>previsto</sub> não é produzido nesta linha celular, até porque esta difere das restantes não sendo cancerígena, estando imortalizada por expressar a telomerase. No entanto, o resultado correspondente à amostra 6 vai ao encontro do acima mencionado, pois

quando comparada às restantes amostras, em que se usou o mesmo par de *primers*, a banda surge mais fraca. Porém, também não se pode excluir a hipótese de a abundância dos transcritos analisados poder ser muito inferior nesta linha celular. Note-se que a quantidade de cDNA usado na reação de amplificação por PCR foi igual para todas as amostras.

Observa-se ainda que as reações de PCR efetuadas utilizando os pares de *primers* representados pelas setas a verde resultam melhor, o que em certa parte também se pode dever às condições de otimização do PCR, nomeadamente o  $T_m$  (ver Tabela 2, Capítulo II, secção II.6), que favoreceram este conjunto de *primers*. Os resultados obtidos para as amostras referentes ao número 2 e 7 podem, em princípio, dever-se à qualidade do cDNA produzido, visto que no controlo positivo também não se obteve bandas a partir desta linha celular.

Estes resultados promissores relativamente à existência de transcritos alternativos suscitaram-nos o interesse de efetuar um novo PCR mas desta vez utilizando como molde cDNAs produzidos a partir de diferentes tecidos humanos. O objetivo de realizar esta experiência era o de comparar resultados e verificar se parte do exão 1, onde se localiza o AUG<sub>previsto</sub> era produzido de forma semelhante ao encontrado nas diferentes linhas celulares analisadas. No entanto, neste PCR apenas se usaram os *primers* que se encontram localizados no exão 1 e 3 visto que, tal como já foi referido, os resultados terem parecido mais esclarecedores com o uso destes *primers* (Figura 20).



**Figura 20: Análise de transcritos alternativos do gene *tbccd1* por PCR em cDNAs produzidos a partir de RNA proveniente de diferentes tecidos humanos.** Os *primers* utilizados correspondem aos que se encontram localizados no exão 1 e 3 (ilustrados pelas setas a verde). (M) Marcador de massas moleculares; (1) Pulmão; (2) Fígado; (3) Timo; (4) Músculo liso; (5) Testículo; (6) Fígado fetal; (7) Coração; (8) Próstata; (9) Medula espinhal; (10) Placenta; (11) Baço.

A análise em gel de agarose dos produtos de PCR apresentada na Figura 20 permitiu tirar algumas conclusões mas também levantou algumas dúvidas/questões. Em primeiro lugar, as bandas correspondentes aos produtos de amplificação não apresentam todas a mesma massa molecular nos diferentes tecidos. Mais interessante ainda, é que essa diferença é clara quando se compara as amostras 4 e 5 com as restantes amostras. Estas duas amostras correspondem a amplificações utilizando os cDNAs produzidos a partir de RNAs extraídos de músculo liso e de testículos, respectivamente. Estes são dois tecidos especiais, um rico em actina e outro rico em flagelos e células com citoesqueleto de microtúbulos especializado. É bastante interessante o resultado da amostra de testículos pois apesar de estudos anteriores verificarem que o gene *tbccd1* era expresso de forma ubíqua, o seu mRNA era mais abundante nos testículos (Gonçalves 2010). O que já era curioso, pois nos testículos o citoesqueleto de microtúbulos assume uma função decisiva durante, por exemplo, a

diferenciação dos espermatozoides. De facto, a formação do flagelo dos espermatozoides, uma estrutura de microtúbulos altamente especializada, requiere profundas remodelações ao nível do citoesqueleto de microtúbulos. É ainda curioso a intensidade da amostra de músculo liso, pois embora não existam dados sobre uma eventual função do TBCCD1 no citoesqueleto de actina, sabe-se que este possui o domínio CARP, que está presente nas proteínas CAP que se ligam à G-actina e regulam a sua polimerização.

Para determinar o tamanho dos produtos amplificados nas amostras 4 e 5, recorreu-se ao cálculo das massas moleculares usando como referência uma reta padrão construída com base nos marcadores de massa molecular. Quanto à amostra 4 o resultado obtido foi de 702 pb. Por sua vez o cálculo das massas moleculares correspondentes às 2 bandas da amostra 5 foi de 601 pb para a banda de maior massa molecular e de 416 pb para a banda de menor massa molecular.

No sentido de interpretar os resultados obtidos procedemos novamente à análise da estrutura do gene que codifica para o TBCCD1, a qual sugeriu que a possibilidade de o produto obtido pela amplificação com os *primers* ilustrados pelas setas a castanho ser de 298 pb e pelas setas a verde ser de 579 pb, respetivamente. Estes tamanhos seriam justificados pelo facto de que os dois pares de *primers* originam produtos de amplificação em que ambos os codões de iniciação estão presentes.

A discrepância entre os valores esperados e os encontrados deveria residir no facto dos calculados serem valores aproximados afetados pela resolução e migração no gel de agarose. Como estes tamanhos são previstos e os outros calculados, os produtos de maior massa molecular nas amostras 4 e 5 poderiam corresponder de facto ao transcrito de 579 pb. Em suma, os resultados obtidos nesta reação de PCR foram ap encontro dos analisados para as linhas celulares (Figura 19).

No entanto é de notar que o tamanho de 298 pb não vai de encontro aos 255 pb esperados nas interpretações iniciais dos possíveis *splicings* (em que ambos os AUGs estariam presentes), pois nesta nova análise da estrutura do gene *tbccd1* observou-se que no exão 2 existia uma sequência nucleotídica, com 43 nucleótidos, precedente ao

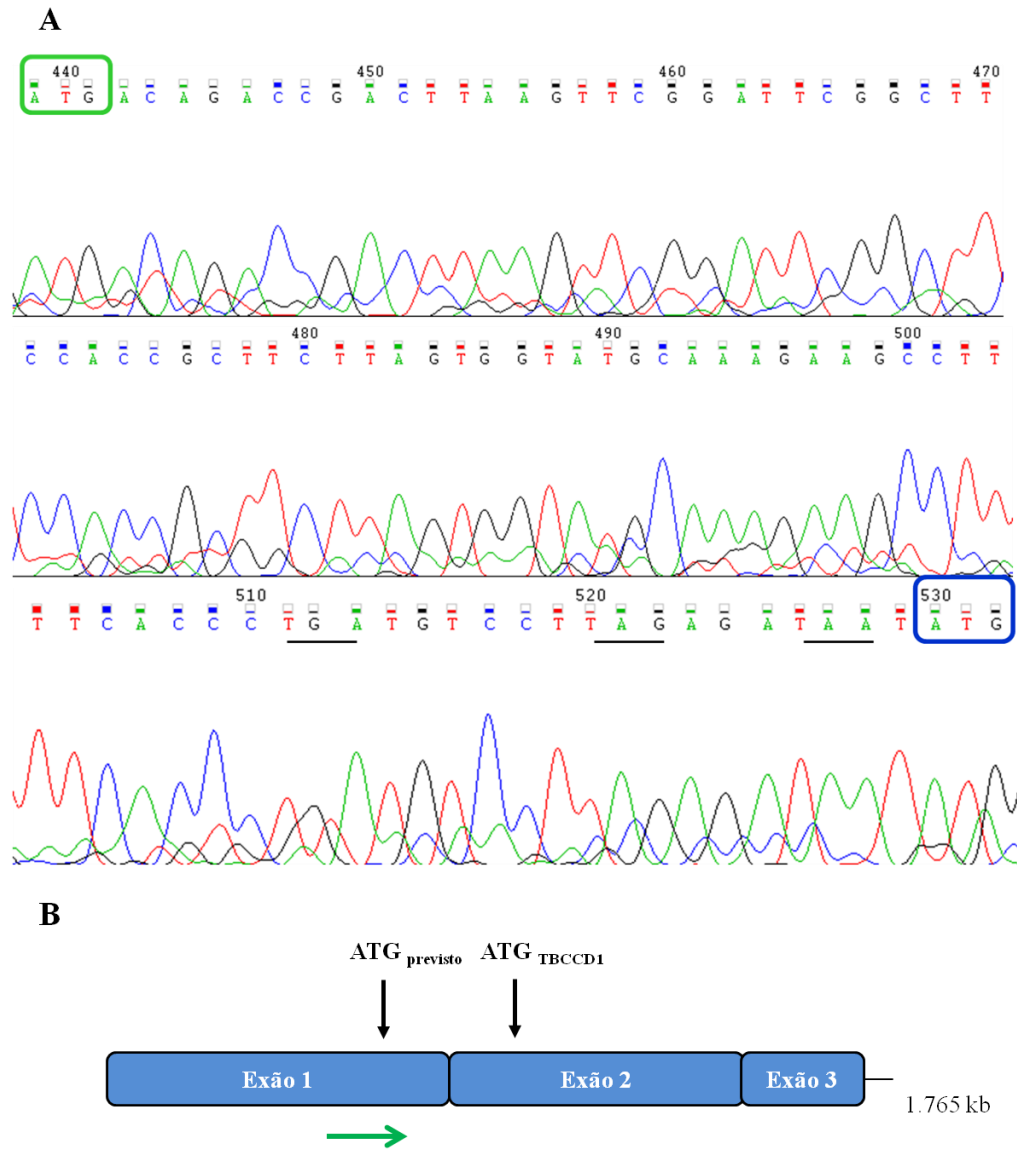
codão de iniciação ( $AUG_{TBCCD1}$ ), que inicialmente não foi tida em conta por não codificar para uma proteína funcional. Levantou-se então a hipótese de esta também estar presente nos produtos obtidos pela amplificação por PCR pois o tamanho previsto tornar-se-ia muito semelhante àquele calculado a partir do gel de agarose, para ambos os *primers*.

Assim, as massas moleculares obtidas não estavam de acordo com as interpretações iniciais de como poderiam ser originados transcritos alternativos a partir do gene *tbccd1* contendo o  $AUG_{previsto}$  e por isso, parecia longe de estar claro se realmente haveriam transcritos alternativos do gene *tbccd1* e/ou como é que estes eram gerados.

Para clarificar os resultados obtidos e saber exatamente a sequência nucleotídica do transcrito amplificado a partir do cDNA sintetizado com base nos RNAs de testículos procedeu-se à sua clonagem e posteriormente à sua sequenciação. Note-se que para este efeito foi necessário repetir a reação de PCR para a amostra de testículos, sendo que a quantidade que restava dela já era muito reduzida (a sua origem era humana o que torna a amostra limitante), o que pode explicar o fato de nesta segunda reação, na análise em gel de agarose só ter sido visível a banda de maior massa molecular e portanto a única que foi sequenciada. Por outro lado, a análise da estrutura do gene também não nos permitiu propor uma hipótese explicativa para a estrutura deste eventual transcrito de menores massas moleculares presente no cDNA de testículos.

A análise da sequência nucleotídica obtida veio comprovar que ambos os AUG estavam presentes no transcrito amplificado a partir de cDNA de testículos (Figura 21A). Ou seja, para além do  $AUG_{TBCCD1}$  que codifica para a proteína funcional TBCCD1, encontrava-se também presente o  $AUG_{previsto}$ . Verificou-se ainda que o número de nucleótidos, que separam estes dois codões de iniciação eram 88 (esta contagem não inclui os nucleótidos que integram os codões de iniciação; para análise detalhada ver anexo 1) (Figura 21A), o que justificava as massas moleculares dos produtos de PCR calculadas a partir da análise em gel de agarose.

Na figura 21B está representado um esquema ilustrativo do transcrito do *tbcd1* que é amplificado a partir dos cDNAs obtidos a partir dos testículos humanos e linhas celulares.



**Figura 21: Transcrito alternativo de maiores dimensões do gene *tbcd1*. (A) Electroferograma da sequenciação automática que comprova a existência do transcrito maior do TBCCD1 presente em cDNAs de diferentes linhas e tecidos humanos. Os códons de iniciação encontram-se salientados em figuras retangulares, sendo que a verde destaca-se o AUG<sub>previsto</sub> e a azul o AUG<sub>TBCCD1</sub>. Sublinhados estão os códons stop encontrados na mesma ORF que o AUG<sub>previsto</sub>. (B) Representação esquemática parcial do transcrito presente no cDNA das diferentes linhas e tecidos humanos. Este tem início no fim do exão 1 (seta verde), onde se localiza o ATG<sub>previsto</sub>, portanto os 1,765 kb representados no esquema incluem tanto o ATG<sub>previsto</sub> como o codão STOP.**

Os novos dados obtidos levantaram novas questões uma vez que a análise da sequência nucleotídica revelou que no seguimento do AUG<sub>previsto</sub> não se encontra nenhuma grelha de leitura aberta (existem 3 codões stop nesta grelha de leitura como é possível ver na Figura 21A), e portanto não é possível que este transcrito codificasse para uma proteína funcional que se iniciasse neste AUG.

Por consequência, a primeira hipótese que surgiu foi a de que o AUG<sub>previsto</sub> se encontraria localizado numa região 5' não traduzida (5'UTR), também denominada por sequência líder, do transcrito do *tbcd1* canónico (cuja região codificante se inicia no AUG<sub>TBCCD1</sub>) não sendo deste modo funcional. Porém é de referir que a existência deste transcrito requiere a remoção de um intrão de aproximadamente 5734 nucleótidos (Anexo1), presente entre os dois codões de iniciação identificados. Este facto levou-nos a propor duas hipóteses alternativas. No primeiro caso podemos assumir que este transcrito poderia corresponder a um transcrito precursor não maduro, onde os codões stop encontrados fizessem, na verdade, ainda parte do intrão que neste caso conteria parte da sequência do exão 2 incluindo o AUG<sub>TBCCD1</sub>. Esta hipótese irá ser explorada na próxima secção deste capítulo pelo que a deixaremos por agora em aberto. Uma segunda hipótese é a de uma pequena ORF (grelha de leitura aberta, de *Open reading frame*) e o próprio AUG<sub>previsto</sub> terem um papel regulador da tradução do mRNA canónico que origina o TBCCD1 a partir do AUG<sub>TBCCD1</sub>.

De facto, a regulação dos genes eucariotas inclui vários passos, entre eles a transcrição, o processamento pós-transcricional, a exportação nuclear e localização e a tradução do mRNA maduro. A própria tradução é regulada por um diverso grupo de mecanismos que atuam ao longo da iniciação, alongação e terminação (Morris and Geballe 2000). Além da sequência nucleotídica que flanqueia o codão de iniciação da tradução, também a sequência 5' UTR de um mRNA desempenha um papel fundamental na determinação da eficiência da tradução, o que é suportado pelo aumento de exemplos de identificação de elementos reguladores presentes nas UTRs (Nagao and Obokata 2006).

Existem algumas características estruturais da 5'UTR que podem influenciar a eficiência da tradução, em particular o comprimento da 5'UTR, a presença de estruturas secundárias e a presença de tripletos AUG a montante do verdadeiro início da tradução no mRNA. Estes AUGs são conhecidos como AUGs a montante (de *upstream* uAUGs) do codão de iniciação da grelha de leitura (ORF) principal (Churbanov, Rogozin et al. 2005). Entenda-se por uAUG, qualquer codão localizado numa região 5' líder de um mRNA, a montante do codão de iniciação usado pelos ribossomas para dar início à síntese proteica e produzir o principal produto proteico. A grelha de leitura a jusante do uAUG corresponde à uORF, a ele associada. Tendo em conta esta informação a pequena ORF presente entre o AUG<sub>previsto</sub> e o AUG<sub>TBCCD1</sub> poderia ser considerada uma uORF e o AUG<sub>previsto</sub> um uAUG.

Sabe-se ainda que os codões AUG a montante da principal ORF podem induzir a formação de ribossomas competentes para a tradução, que por sua vez podem traduzir e (1) terminar e reiniciar, (2) terminar e deixar o mRNA, resultando numa sob-regulação da tradução da principal ORF ou (3) sintetizar uma proteína maior ao nível da N-terminal. A frequência relativa da presença de um uAUG dentro de um contexto bom ou adequado sugere que a maquinaria traducional pode lidar com estas sequências na 5' UTR e por isso os AUGs podem estar envolvidos no controlo traducional dos mRNAs. Assim, a presença de uAUGs na 5' UTR geralmente reduz a eficiência da iniciação da tradução no AUG da ORF principal. É de notar que nem todas as proteínas devem ser sintetizadas em taxas elevadas e portanto, as ORF principais nem sempre são traduzidas eficientemente (Meijer and Thomas 2002). Neste contexto não podemos excluir que a pequena ORF presente entre o AUG<sub>previsto</sub> e o AUG<sub>TBCCD1</sub> desempenhe um papel regulador na tradução do mRNA codificante para o TBCCD1 (Ver tabela 7, para a sequência de resíduos de aminoácidos gerados por esta ORF).

Por outro lado, sabe-se que o uso de promotores alternativos e/ou o *splicing* alternativo podem remover as uORF, que estão presentes em alguns mRNAs (Sachs and Geballe 2006). No caso do TBCCD1 o mecanismo de *splicing* estaria envolvido em “trazer” uma ORF para a *leader* do mRNA codificante para o TBCCD1. Poderíamos assim imaginar que em certas condições seria essencialmente produzido

o transcrito típico/canónico do TBCCD1 que possui uma *leader* que não contém o AUG<sub>previsto</sub>. Porém, noutras situações o gene passaria a ser transcrito a partir de outro local de início da transcrição o que implicaria a remoção de um grande intrão (5734 nucleótidos) na sua extremidade 5' e a criação do transcrito com a estrutura descrita na Figura 19 e 20, o qual conteria agora uma uORF na 5'-UTR. Esta última poderia assim estar envolvida na regulação da taxa de tradução deste RNA. Estas hipóteses requerem, no entanto comprovação experimental.

**Tabela 7: Sequência de resíduos de aminoácidos que constituem a ORF gerada entre os dois AUGs presentes na 5' UTR.** Os codões STOP estão representados pelos símbolos asterisco.

#### ORF gerada entre os dois AUGs presentes na 5' UTR

MTDRLKFGFGFHRFLVVCKEAFSP\*CP\*R\*

### III.2 O gene *tbccd1* apresenta um padrão de *splicing* alternativo que codifica para uma proteína diferente do TBCCD1 canónico

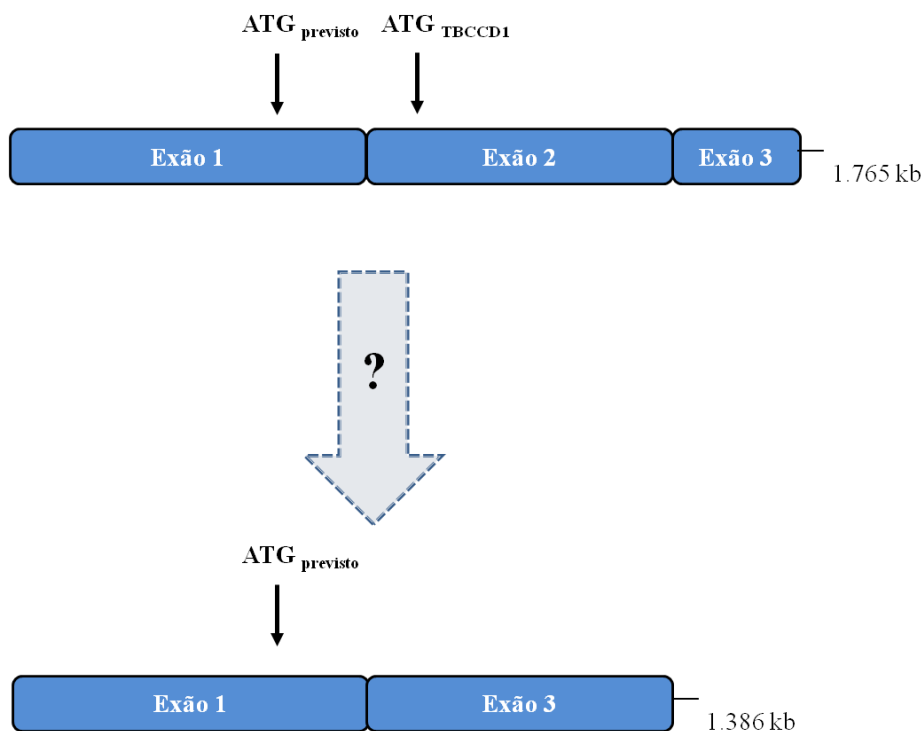
Os resultados acima mencionados levaram-nos, pela segunda vez, a fazer uma reanálise bastante crítica da estrutura do gene *tbccd1*. De facto, e tal como referido anteriormente, o transcrito encontrado com a estrutura descrita na figura 21B poderia corresponder a um transcrito precursor não maduro, onde os codões stop encontrados fizessem na verdade ainda parte de um intrão, que neste caso conteria parte da sequência do exão 2 incluindo o AUG<sub>TBCCD1</sub>. Se esta hipótese fosse verdadeira seria possível após a excisão deste hipotético intrão originar um transcrito capaz de codificar para uma variante TBCCD1 com massa molecular mais pequena.

Na nova abordagem realizou-se novamente uma busca exaustiva na base de dados provenientes da sequenciação de genomas do NCBI, através da ferramenta

bioinformática *protein blast* (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). Nesta pesquisa usaram-se as sequências de resíduos de aminoácidos previstas se uma região codificante tivesse como codão iniciador o AUG<sub>previsto</sub> e fim no exão onde ele se localiza (exão 1), as quais correspondem a 16 resíduos de aminoácidos.

Esta estratégia permitiu identificar um cDNA previamente descrito, mas não acessível nas primeiras abordagens de análises de sequência anteriormente realizadas, capaz de codificar para um TBCCD1 cujo AUG iniciador era de facto o AUG<sub>previsto</sub> e que possuía uma sequência de resíduos de aminoácidos diferente do TBCCD1 canónico. Curiosamente, este cDNA foi inicialmente identificado num estudo em que se determinou a sequência completa de possíveis cDNAs humanos completos, servindo de base para o transcriptoma humano e genómica funcional. Neste mesmo estudo, além de se obter a informação das sequências de cDNA o objetivo era também proporcionar uma fonte física de clones de cDNA. Assim, apresentaram a primeira caracterização de 21,243 clones (Ota, Suzuki et al. 2004).

Em conclusão, a grande diferença entre esta proteína alternativa do TBCCD1 e o TBCCD1 canónico reside no facto de apresentar um AUG de iniciação diferente, encontrando-se este no AUG localizado num exão 1. Além disso, o exão onde se localiza o AUG<sub>TBCCD1</sub>, não está presente nesta proteína, levantando a hipótese que pode ocorrer um salto do exão (Figura 22). Mais ainda, a identificação deste transcrito suporta a ideia de que o transcrito anteriormente identificado neste trabalho cuja estrutura se encontra presente na Figura 21A e B poderia ser de facto um precursor deste mRNA.



**Figura 22: Representação esquemática do possível padrão de *splicing* alternativo que pode ocorrer no gene *tbccd1*, originando por isso uma proteína alternativa ao TBCCD1 canônico.**

O novo isótipo identificado para o TBCCD1 foi a partir daqui e ao longo deste trabalho designada por variante do TBCCD1. A sequência de resíduos de aminoácidos prevista para esta variante é de 461 aminoácidos, com uma massa molecular prevista de ~ 52 kDa e um pI previsto de 9.19 (estes valores foram obtidos recorrendo à ferramenta presente no site [http://web.expasy.org/compute\\_pi/](http://web.expasy.org/compute_pi/)). Por sua vez, o TBCCD1 canônico é composto por 557 aminoácidos, uma massa molecular prevista de ~ 64 kDa e um pI previsto de 8.8688 (Gonçalves 2010). As duas proteínas apresentam uma identidade de 96,96 % entre as suas sequências de resíduos de aminoácidos previstas (Figura 23). A nova variante TBCCD1 apresenta os domínios funcionais TBCC e CARP, característica que é partilhada pelas proteínas TBCC, RP2 e TBCCD1 canônico. Esta comparação mostra claramente que a grande diferença entre as duas proteínas reside essencialmente na extremidade N-terminal do TBCCD1 (Figura 23).

CLUSTAL 2.1 multiple sequence alignment

```

TBCCD1_Variante      -----
TBCCD1_Canónico     MDQSRVLLWVKAEPFIVGALQVPPPSKFSLHYLRKISTYVQIRATEGAYPRLYWSTWRHI 60

TBCCD1_Variante      -----MTDRLKFGFGFHRFLV-----LSVDTLQF 24
TBCCD1_Canónico     ACGKQLAKDLAWLYFEIFDSLMSKTPPEERLEWSEVLSNCMSEEEVEKQRNQLSVDTLQF 120
                        :*::. . : . :
                        *:*:. . : . :
                        *:*:. . : . :

TBCCD1_Variante      LLFLYIQQLNKVSLRSLIGEEWSPRNKSQSPDLTEKSNCHNKNWNDYSHQAFVYDHL 84
TBCCD1_Canónico     LLFLYIQQLNKVSLRSLIGEEWSPRNKSQSPDLTEKSNCHNKNWNDYSHQAFVYDHL 180
                        *****

TBCCD1_Variante      DLLELLDPKQLTASFHSTHSSLSVSREAVVALSFLIEGTISRARKIYPLHELALWQPLHA 144
TBCCD1_Canónico     DLLELLDPKQLTASFHSTHSSLSVSREAVVALSFLIEGTISRARKIYPLHELALWQPLHA 240
                        *****

TBCCD1_Variante      DSGFSKISKTFsfyKLETWLRscltGNPFGTSACLKSGKKLAWAHQVEGTTKRAKIACNT 204
TBCCD1_Canónico     DSGFSKISKTFsfyKLETWLRscltGNPFGTSACLKSGKKLAWAHQVEGTTKRAKIACNT 300
                        *****

TBCCD1_Variante      HVAPRMHRLVMSQVYKQTLAKSSDTLAGAHVKIHRcNESFIYLLSPLRSVTIEKCRNSI 264
TBCCD1_Canónico     HVAPRMHRLVMSQVYKQTLAKSSDTLAGAHVKIHRcNESFIYLLSPLRSVTIEKCRNSI 360
                        *****

TBCCD1_Variante      FVLGPGVGTTLHLHSCDNVKVIAVCHRLSISSTTGCI FHVLTPTRP LILSGNQTVTFAPFH 324
TBCCD1_Canónico     FVLGPGVGTTLHLHSCDNVKVIAVCHRLSISSTTGCI FHVLTPTRP LILSGNQTVTFAPFH 420
                        *****

TBCCD1_Variante      THYPML EDHMARTGLATVPNYWDNPMVVCRENSDTRVFQLLPPEFYVFIIPFEMEGD TT 384
TBCCD1_Canónico     THYPML EDHMARTGLATVPNYWDNPMVVCRENSDTRVFQLLPPEFYVFIIPFEMEGD TT 480
                        *****

TBCCD1_Variante      EIPGGLPSVYQKALGQREQKIQIWQKTVKEAHLTKDQRKQFQVLVENKfyEWLINTGHRQ 444
TBCCD1_Canónico     EIPGGLPSVYQKALGQREQKIQIWQKTVKEAHLTKDQRKQFQVLVENKfyEWLINTGHRQ 540
                        *****

TBCCD1_Variante      QLDSLVPAAAGSKQAAG 461
TBCCD1_Canónico     QLDSLVPAAAGSKQAAG 557
                        *****

```

**Figura 23: Alinhamento das sequências de resíduos de aminoácidos da proteína TBCCD1 canónico e das sequências de resíduos de aminoácidos previstas para a variante TBCCD1.** As sequências de resíduos de aminoácidos das proteínas TBCCD1 foram alinhadas no programa ClustalW (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>), o qual também nos permitiu verificar que as duas proteínas apresentam uma identidade de 96,96%. Observa-se que a grande diferença entre as duas proteínas reside principalmente na região N-terminal de ambas.

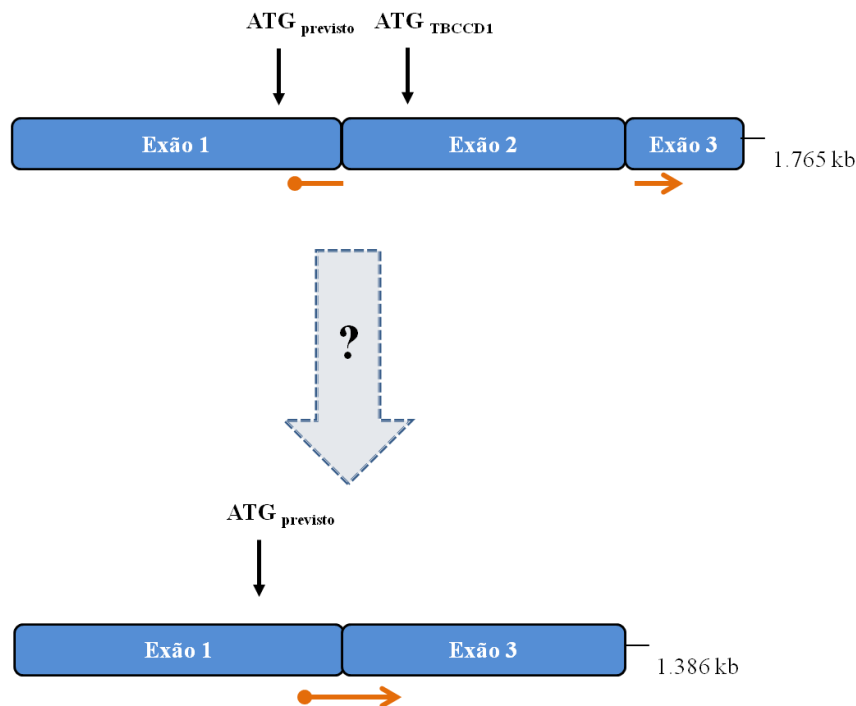
### **III.3 Pesquisa do transcrito alternativo do gene *tbccd1* que codifica para a variante identificada na base de dados por amplificação específica de cDNAs**

Os resultados acima mencionados foram bastante atrativos e levaram-nos de imediato a equacionar várias questões. Inicialmente pensou-se que este transcrito não seria expresso de igual forma ao TBCCD1 canónico, visto que até à data nunca tinha sido identificado, a não ser no estudo acima referido em que se analisou a sequência completa de vários cDNAs humanos e portanto à partida o seu padrão de expressão, ou seja, a sua presença ou abundância numa célula ou tecido em particular seria diferente.

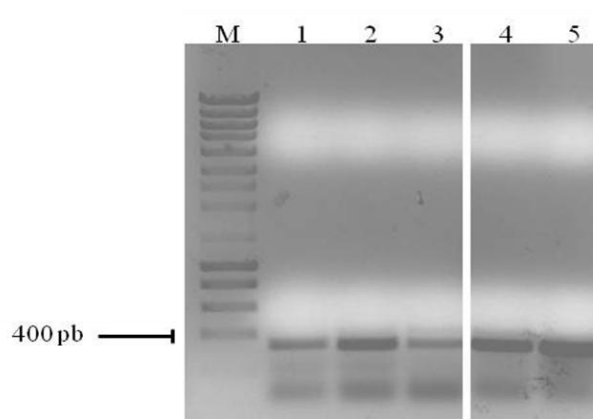
Outra questão que se levantou diz respeito à sua localização. Sabe-se que o TBCCD1 canónico, como referido anteriormente, localiza-se no centróssoma durante a interfase e as diferentes fases da mitose, sendo que durante anafase parece ser também ser recrutado para a zona mediana do fuso mitótico e posteriormente para o corpo médio na citocinese (Gonçalves 2010). Além disto, sabia-se ainda que é o domínio na N-terminal, mais propriamente os primeiros 20 resíduos de aminoácidos, do TBCCD1 que determina esta localização centróssomal. Esta foi uma questão interessante visto que na proteína alternativa estes 20 resíduos de aminoácidos não se encontram presentes e portanto poderia de alguma forma afetar a localização desta nova proteína TBCCD1.

Imediatamente surgiu o interesse de provar que o cDNA descrito na literatura e no banco de dados existia realmente. Assim, de forma a confirmar esta existência e consequentemente a expressão do RNA que codifica para proteína variante do TBCCD1 canónico, a primeira estratégia definida consistiu em desenhar um conjunto de *primers* que numa reação de amplificação por PCR permitissem amplificar unicamente o cDNA codificante para esta proteína. Para isso foi desenhado um *primer forward* cuja sequência emparelhasse repartidamente com as sequências alvo situadas no exão 1 e 3 (Figura 24). Isto significa que na totalidade o *primer* era

composto por 21 nucleótidos, sendo que os 11 primeiros correspondem aos últimos 11 nucleótidos presentes no exão 1 e os restantes nucleótidos do *primer* correspondem aos 10 primeiros nucleótidos do exão 3, e portanto ao originarem um produto de amplificação este deveria corresponder necessariamente à junção destes dois exões. Quanto ao *primer* reverse, não foi tido em consideração nenhum aspeto específico, apenas que a amplificação por PCR originasse uma banda de massa molecular relativamente visível pela análise em gel de agarose. Desta forma o *primer* foi desenhado sobre o exão 4 (não se encontra esquematizado na Figura 24), o que pela análise da estrutura do gene que codifica para o *tbccd1* resultaria em produtos de PCR correspondentes a bandas de massa molecular de 358 pb.



**Figura 24:** Representação esquemática do possível padrão de *splicing* alternativo que pode ocorrer no gene *tbccd1* e dos passos que integraram o desenho do *primer forward* (laranja). Este *primer* foi utilizado na pesquisa da proteína que codifica para o transcrito alternativo do gene *tbccd1* por amplificação específica por PCR de cDNAs. O produto esperado seria de 358 pb.



**Figura 25: Amplificação de cDNA em diferentes tecidos e linhas celulares humanas.** (M) Marcador de massas moleculares; (1) Pulmão; (2) Coração; (3) Músculo Liso; (4) HeLa; (5) MCF-7. Os *primers* usados nesta reação de amplificação, estão ilustrados pelas setas a laranja na figura 23. É de referir que o gel se encontra cortado, pois continha amostras que não eram relevantes para o presente estudo.

Na figura 25 são apresentados os resultados da amplificação obtida por PCR, com o *primer* ilustrado a laranja (Figura 24) definido para a estratégia apresentada.

Note-se que a escolha dos cDNAs analisados deve-se ao facto destes terem sido sintetizados a partir de RNAs extraídos de diferentes tecidos humanos cuja escolha se prendeu com os dados referentes à localização da proteína TBCCD1 canónica. De facto, sabe-se que o TBCCD1 endógeno se localiza no corpo basal dos cílios motores como observado em culturas primárias do cerebelo de murganho, em que se diferenciam células multiciliadas que apresentam cílios motores (Gonçalves 2010). Os cílios motores conferem mobilidade à célula, como no caso do flagelo dos espermatozoides ou por outro lado promovem o movimento de fluidos através do seu batimento como acontece nas células multiciliadas do epitélio respiratório, onde parecem assumir um papel importante para o *clearance* respiratório e prevenção da colonização bacteriana (Satir and Christensen 2007). Estes factos justificam a escolha dos cDNAs de pulmão para pesquisa da variante do TBCCD1.

Por outro lado, a proteína TBCCD1 canónica foi também localizada na zona de transição dos cílios primários na linha celular RPE-1. Os cílios primários ao contrário dos motores são imóveis sendo essencialmente estruturas sensoras. Estes cílios têm chamado a atenção dos especialistas devido à sua distribuição generalizada nas

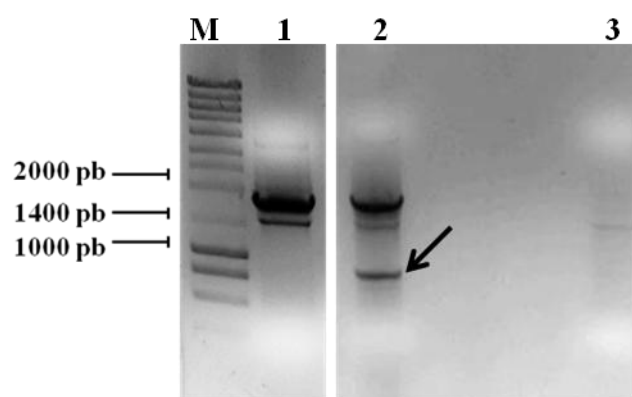
células do corpo, onde se incluem células não epiteliais como os fibroblastos (representam a maioria das células no coração “normal”) e as células do músculo liso (Satir and Christensen 2007).

Como já referido anteriormente, não existam dados sobre uma eventual função do TBCCD1 no citoesqueleto de actina, no entanto sabe-se que este possui o domínio CARP, o qual está presente nas proteínas CAP que se ligam à G-actina e regulam a sua polimerização. Por sua vez, encontra-se explicada a razão da escolha dos restantes tecidos humanos.

Os resultados obtidos e apresentados na Figura 25 mostram que o cDNA correspondente à variante do TBCCD1 está presente em abundância semelhante nas diferentes linhas celulares e tecidos humanos analisados.

Embora estes resultados fossem promissores mostrando experimentalmente a ocorrência do cDNA correspondente à variante do TBCCD1 decidimos pesquisar se este transcrito alternativo do TBCCD1 poderia ser amplificado na sua totalidade a partir cDNAs produzidos a partir de RNAs provenientes de linhas celulares humanas, nomeadamente MCF-7 e HeLa.

A obtenção destes produtos de PCR poderia permitir eventualmente a posterior clonagem do cDNA correspondente à variante do TBCCD1. Para isso, decidiu-se amplificar, por reação de PCR, amostras de cDNA, em que se recorreu ao uso dos *primers* desenhados sobre a N-terminal e C-terminal desta proteína (Figura 26). O objetivo era verificar pela análise em gel de agarose, bandas correspondentes a fragmentos com uma massa molecular esperada de 1765 pb e 1386 pb. Na análise em gel de agarose observou-se, de facto que estes pares de *primers* amplificaram um produto de PCR com massas moleculares de aproximadamente 1900 pb e 1400 pb e ainda na amostra correspondente ao cDNA de MCF-7 observou-se um produto de massa molecular aproximada de 1000 pb. Os dois primeiros transcritos podem, portanto, resultar do facto da existência de dois transcritos alternativos do gene *tbcd1*.



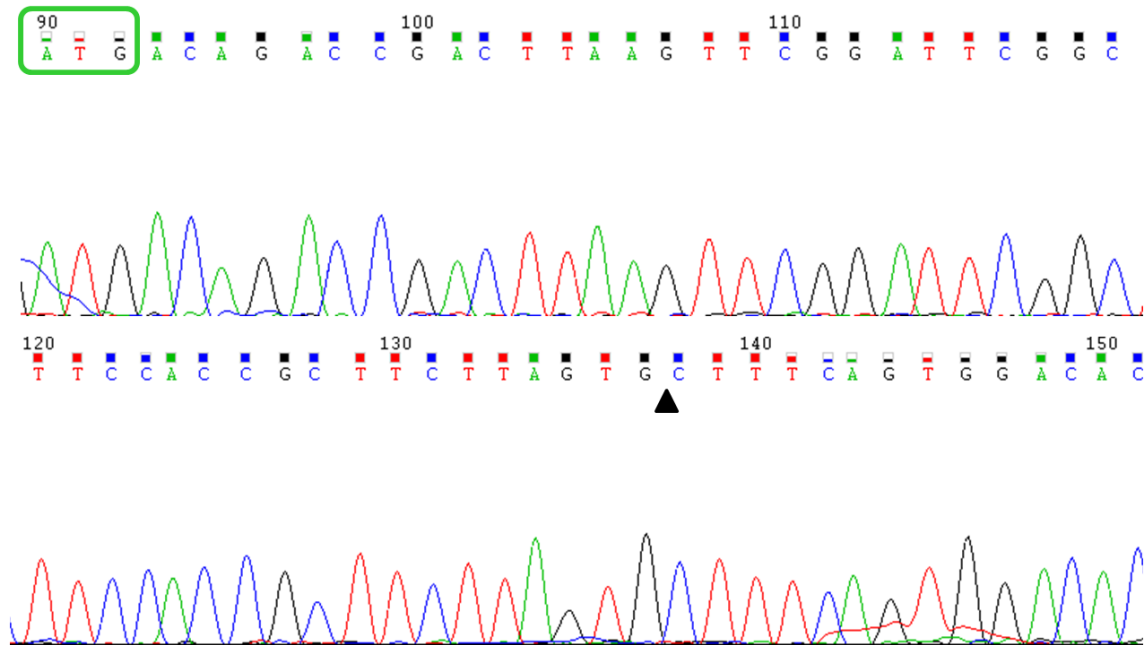
**Figura 26: Amplificação de cDNA produzido a partir de RNAs extraídos das linhas celulares HeLa e MCF7. (M) Marcador de massas moleculares; (1) HeLa; (2) MCF-7; (3) Controlo negativo.** Verifica-se que o tamanho das bandas é de aproximadamente 1400 pb e 1900 pb, sendo que na amostra 2 existe uma com um tamanho aproximado de 1000 pb (seta). Note-se que o gel se encontra cortado, pois continha amostras que não eram relevantes para o presente estudo.

O transcrito correspondente à banda de massa molecular de aproximadamente 1400 pb foi clonado e sequenciado para determinar com exatidão a sua estrutura e sequência. Os resultados desta sequenciação confirmaram a estrutura do transcrito como sendo uma variante capaz de codificar um TBCCD1 de menor massa molecular devido à diferença encontrada na região codificante para os primeiros aminoácidos do domínio N-terminal do TBCCD1 comprovando a existência do cDNA descrito anteriormente por (Ota, Suzuki et al. 2004) (Figura 27). De facto, esta variante do TBCCD1 apresenta a sequência de resíduos de aminoácidos prevista apresentada na figura 23.

Por sua vez, a sequenciação do produto de maior massa molecular (~1900 pb) provou mais uma vez a existência de um transcrito que além do AUG<sub>previsto</sub>, apresenta o AUG<sub>TBCCD1</sub>.

Tentou-se ainda sequenciar o transcrito ilustrado na Figura 26 (ver seta) correspondente à amostra 2, MCF-7, e com uma massa molecular de aproximadamente 1000 pb, no entanto não nos foi possível clonar este transcrito o que impediu a determinação da sua sequência nucleotídica e por consequência

investigar a sua estrutura. Por outro lado, a análise da estrutura do gene também não nos permitiu propor uma hipótese explicativa para a estrutura deste eventual transcrito, o que sugere que a referida banda poderá corresponder a um produto de PCR inespecífico.



**Figura 27: Electroferograma da sequenciação automática que comprova a existência do transcrito previsto e alternativo ao TBCCD1 canónico.** O ATG<sub>previsto</sub> encontra-se destacado pela figura retangular a verde. A seta preta indica o local de junção do exão 1 ao exão 3, do gene que codifica para a proteína TBCCD1.

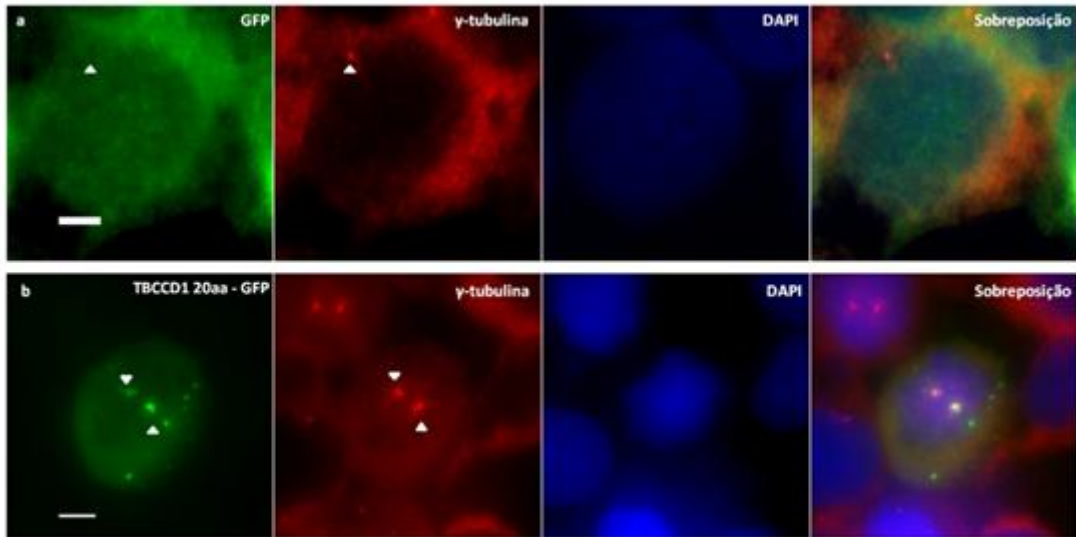
### III.4 Análise da sequência de resíduos de aminoácidos da região N-terminal da variante do TBCCD1

Estudos anteriores mostraram que o domínio constituído pelos 20 primeiros resíduos de aminoácidos da proteína TBCCD1 são críticos para a localização centrossomal da proteína. Verificou-se que, através da expressão deste péptido em fusão com a proteína GFP, este domínio por si só é capaz de se localizar no centrossoma, quer em células em interfase, quer em células em mitose. No mesmo estudo verificou-se que

dentro da região N-terminal, os primeiros 20 resíduos de aminoácidos correspondiam a um domínio hidrófobo (Pereira 2011). Por exemplo, este domínio poderia promover uma ligação direta ou indireta desta proteína às membranas. É ainda de referir, que esta característica foi tida em consideração visto que o TBCCD1 está envolvido na ligação do núcleo ao centróssoma por um mecanismo que ainda não é conhecido e desta forma esta ligação poderia envolver interações com o invólucro nuclear. Ou por sua vez, estar relacionado com o complexo de Golgi, embora o TBCCD1 não co-localize com este organito, quando o gene *tbccd1* foi silenciado observou-se a desorganização do complexo de Golgi. Provavelmente o fenótipo observado seja causado pela localização anormal do centróssoma, no entanto não está excluída a hipótese de um envolvimento direto ou indireto da proteína TBCCD1 na organização do Golgi.

O alinhamento das sequências de resíduos de aminoácidos das proteínas TBCCD1 ao nível da região N-terminal permitiu verificar que o domínio responsável pela localização da proteína no centróssoma é altamente conservado nos vertebrados (Pereira 2011).

Neste mesmo estudo foi previsto um local de clivagem para uma protease (peptidase sinal) entre as posições 19 e 20 da proteína TBCCD1, podendo este facto também indicar a existência de duas formas hipotéticas da proteína. Estes resultados em conjunto permitiram prever a função importante dos 20 primeiros resíduos de aminoácidos deste domínio, a qual está diretamente relacionada com a localização centróssomal da proteína (Pereira 2011) (Figura 28).

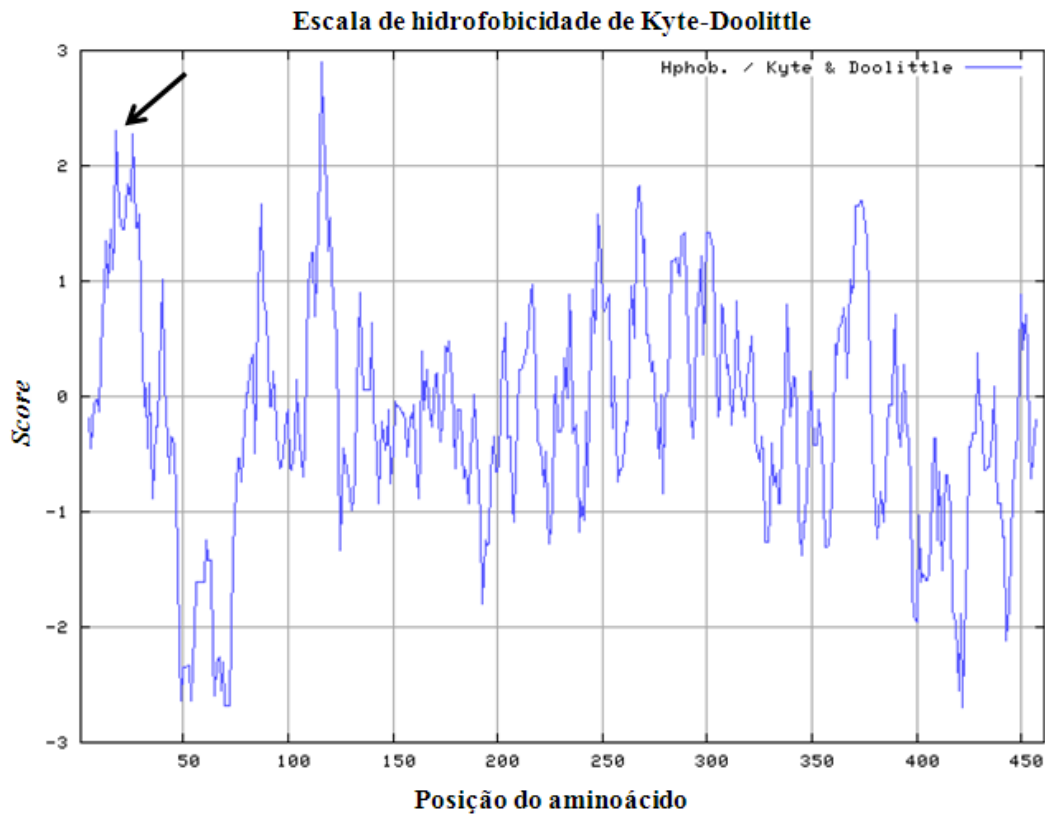


**Figura 28: O domínio mínimo da proteína TBCCD1 humana, composto pelos 20 primeiros resíduos de aminoácidos é responsável pela sua localização centrossomal.** As células HEK 293T foram transfectadas com a proteína GFP (**a**) e com a proteína TBCCD1 20aa em fusão com a proteína GFP (**b**) e processadas para análise de imunofluorescência indireta com um anticorpo contra a  $\gamma$ -tubulina. O DNA foi marcado com DAPI. Em **a** as cabeças das setas em **a** apontam para a  $\gamma$ -tubulina nos centrossomas e para a ausência de GFP nesta estrutura e em **b** apontam para o TBCCD1 20aa e para a  $\gamma$ -tubulina nos centrossomas. As barras de escala correspondem a 5  $\mu$ m (adaptado de Pereira 2011).

De facto, estes resultados foram decisivos no estudo da N-terminal do TBCCD1 e consequente determinação do domínio mínimo para a sua localização no centrossoma.

Desta forma, surgiu o interesse de se analisar a sequência de resíduos de aminoácidos da proteína variante do TBCCD1 de modo a poder identificar locais que possam ter um potencial interesse na sua função. Note-se que a proteína variante do TBCCD1, não apresenta na sua sequência os mesmos 20 resíduos de aminoácidos do TBCCD1 canónico. No entanto, verificou-se que a variante apresenta um domínio equivalente composto por 16 resíduos de aminoácidos (estes são codificados pela sequência do exão 1, que por *splicing* alternativo se irá juntar ao exão 3). Para investigar a natureza destes 16 resíduos de aminoácidos e comparar com o correspondente domínio no TBCCD1 canónico, procedeu-se numa primeira fase a uma análise do perfil de hidrofobicidade da sequência de resíduos de aminoácidos da proteína

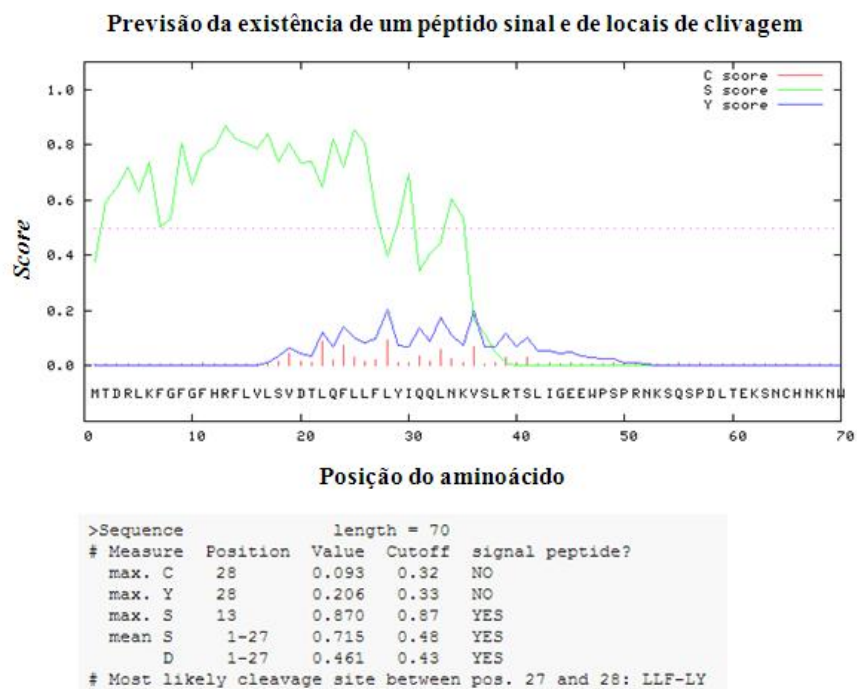
variante (Figura 29). Este estudo mostrou que tal como para o TBCCD1 canónico os primeiros resíduos da região N-terminal da variante correspondem a uma região hidrofóbica sendo o restante perfil de hidropatia também semelhante ao do TBCCD1 canónico.



**Figura 29: Perfil de hidrofobicidade da variante do TBCCD1.** O perfil de hidrofobicidade do TBCCD1 foi conseguido recorrendo à ferramenta bioinformática no site <http://web.expasy.org/protscale/>, onde se seleccionou o algoritmo de hidrofobicidade de Kyte-Doolittle, o qual estabelece que regiões acima de 0 têm um perfil hidrófobo. Os valores aqui mostrados preveem uma região hidrófoba nos primeiros 16 resíduos de aminoácidos do TBCCD1 alternativo, indicados na figura pela seta.

Com o objetivo de averiguar a existência de um eventual péptido sinal e de locais de clivagem do TBCCD1 alternativo, seguiu-se uma análise da sequência primária do mesmo. Em geral um péptido sinal é uma sequência localizada na N-terminal das proteínas de secreção e membranares, recém-sintetizadas, e normalmente compreende 16 a 30 resíduos de aminoácidos. Nos eucariotas, o péptido sinal

direciona a inserção de proteínas na membrana do retículo endoplasmático e são normalmente removidos por clivagem proteolítica por uma peptidase sinal, ou então pode manter-se na proteína, como parte integrante desta (Bendtsen, Nielsen et al. 2004; Neto Ade, Alvarenga et al. 2012). Uma vez que o TBCCD1 canônico possuía um local de clivagem proteolítica putativo entre as posições 19 e 20, investigou-se também se existia um local de clivagem proteolítica putativo entre as posições 15 e 16 da variante TBCCD1. Embora não se tenha verificado a existência de um local de clivagem proteolítica putativo entre estas posições a análise dos resultados sugere a existência do mesmo entre as posições 27 e 28 da variante TBCCD1 (Figura 30).



**Figura 30: Previsão da existência de um péptido sinal na proteína alternativa do TBCCD1.** A pesquisa da previsão do péptido sinal foi obtida recorrendo à ferramenta bioinformática Signal IP 3.0 Server através do site <http://www.cbs.dtu.dk/services/SignalIP/>. Os valores apresentados são resultado da probabilidade de existência de um péptido sinal e da previsão de existência de locais de clivagem. Os *scores* C, S e Y indicam o local de clivagem, o péptido sinal e previsões combinadas de locais de clivagem, respetivamente. Em suma, o *score* Y é o valor que representa a melhor previsão da posição de um local de clivagem (Bendtsen, Nielsen et al. 2004).

Em suma, a análise dos 20 primeiros resíduos de aminoácidos do TBCDD1 canónico sugeriam de facto uma função importante com a localização centrossomal da proteína visto que, além do perfil de hidrófobo desses mesmos resíduos também se previu um putativo local de clivagem entre as posições 19 e 20. Por sua vez, a variante TBCCD1 apresenta também um perfil hidrófobo para os primeiros 16 resíduos aminoácidos, tal como um local de clivagem, embora não tão “forte”, entre as posições 27 e 28. De certa forma, este domínio poderia promover uma ligação direta ou indireta desta proteína às membranas, tal como se referiu anteriormente para o TBCDD1 canónico.

No seu conjunto os resultados aqui apresentados mostram claramente a existência de pelo menos três transcritos obtidos a partir do gene que codifica para o TBCCD1: (1) o transcrito canónico até agora descrito e caracterizado contendo o codão iniciador AUG<sub>TBCCD1</sub> (Gonçalves et al, 2010); (2) um transcrito de maiores dimensões contendo dois possíveis codões iniciadores designados ao longo deste trabalho por AUG<sub>previsto</sub> e o AUG<sub>TBCCD1</sub> e separados por uma uORF e (3) um transcrito que apenas possui o codão de iniciação AUG<sub>previsto</sub>, que codifica para uma variante TBCCD1 de massa molecular mais pequena que o TBCCD1 canónico e codificando para um domínio N-terminal distinto do até agora descrito para o TBCCD1. Por consequência os resultados obtidos apontam claramente para a existência de um *splicing* alternativo para o gene que codifica para a proteína TBCCD1.

Porém algumas questões permanecem em aberto como por exemplo como são gerados estes transcritos? Como já referido no capítulo da introdução, a produção de diferentes isótipos a partir de um mesmo transcrito por *splicing* alternativo é um processo bastante comum e pode ocorrer por diversos padrões. Apesar de menos frequente, um dos padrões de *splicing* que se pode sugerir no presente estudo como ocorrendo para o gene *tbccd1* é o uso de promotores alternativos, originando um isótipo diferente. A iniciação da transcrição em diferentes promotores origina exões com uma extremidade terminal 5' alternativa, que podem ser ligados a jusante a um exão 3' comum. Sabe-se ainda que este tipo de *splicing* pode introduzir alterações

subtis nas sequências codificantes, como por exemplo apenas a alteração de um único codão (Li, Lee et al. 2007; Keren, Lev-Maor et al. 2010).

Este tipo de padrão de *splicing* poderia explicar a presença do transcrito codificante para a variante TBCCD1, porém a existência do transcrito de maiores dimensões contendo o AUG<sub>previsto</sub> e o AUG<sub>TBCCD1</sub> sugere que os padrões de *splicing* do gene TBCCD1 deverão ser mais complexos. Como referido anteriormente este transcrito poderá ter um papel regulador da tradução do TBCCD1 ou alternativamente ser um precursor do transcrito codificante para a variante TBCCD1. Neste último caso não só seria necessário a ocorrência de promotores alternativos, mas também seria necessário que ocorresse um mecanismo de *splicing* alternativo mutuamente exclusivo. Este seria por sua vez um caso especial de exões *cassette* duplos, em que um ou outro exão são incluídos no produto final mas não ambos, pois a inclusão de ambos leva a alterações da grelha de leitura. Estudos permitiram identificar uma variedade de genes humanos e animais que continham pares duplicados de exões, que na sua maioria eram mutuamente exclusivos (Letunic, Copley et al. 2002; Li, Lee et al. 2007).

Poderíamos ainda assumir que o transcrito de maiores dimensões contendo o AUG<sub>previsto</sub> e o AUG<sub>TBCCD1</sub> é simultaneamente capaz de regular a produção de TBCCD1 canónico através da sua uORF e de ser um precursor da variante TBCCD1. Estas hipóteses requerem experiências que as comprovem.

Como já referido no capítulo da introdução a maioria dos organismos apresenta uma proteína ortóloga ao TBCCD1 humano, sendo uma proteína muito conservada nos organismos vertebrados. Porém, quanto à variante TBCCD1 não se pode dizer o mesmo, visto que no presente estudo não se encontraram proteínas homólogas da mesma, neste grupo filogenético. Isto significa que poderá ser uma particularidade do transcriptoma humano ou por sua vez é uma proteína que ainda não foi descrita.

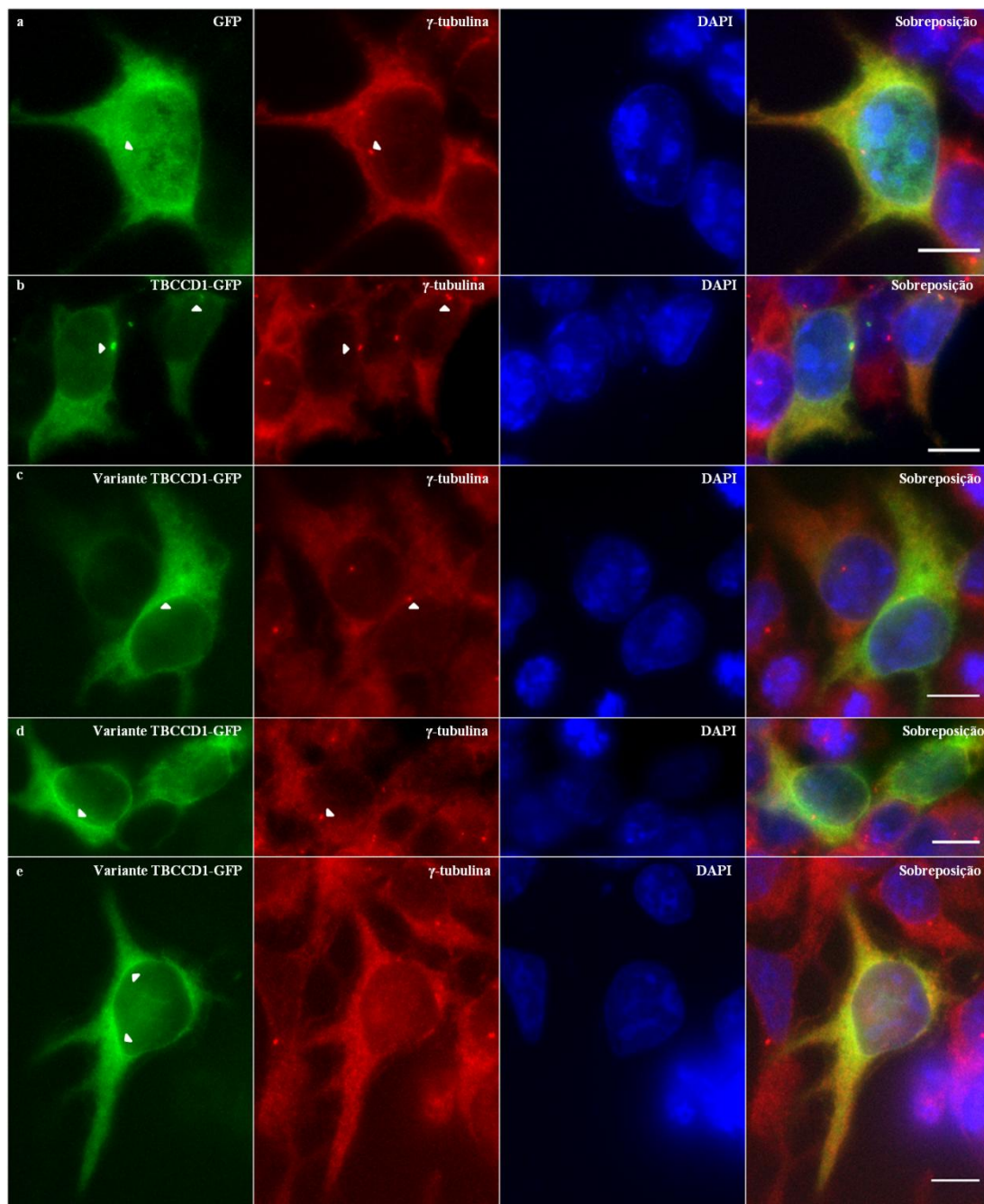
Finalmente seria interessante fazer um estudo da expressão destes transcritos de forma a verificar a sua abundância relativa em diferentes tecidos humanos. Esta análise, além de permitir verificar a ubiquidade ou não destes transcritos poderia de certa forma dar pistas para a função dos dois isótipos do TBCCD1 nos humanos.

### **III.5 Estudo da localização celular da proteína que codifica para a variante do gene *tbccd1* por imunofluorescência nas células humanas HEK 293T**

Como foi referido ao longo deste capítulo, o TBCDD1 canónico apresenta na sua região N-terminal uma sequência de 20 resíduos de aminoácidos responsável pela sua localização no centróssoma. Por sua vez, quando se compara a variante TBCCD1 ao TBCCD1 canónico, uma das grandes diferenças é a região N-terminal destas proteínas e consequentemente a ausência deste domínio na variante TBCCD1. Foi ainda referido que a variante apresenta um domínio equivalente composto por 16 resíduos de aminoácidos de natureza essencialmente hidrofóbica. Desta forma, estudar a sua localização em comparação ao TBCCD1 canónico foi outro dos objetivos do presente trabalho.

Para se estudar a localização da variante TBCCD1 que é codificada pelo transcrito alternativo do gene *tbccd1* que só contém o AUG<sub>previsto</sub>, a sua sequência codificante foi clonada num vetor de expressão em fusão com a proteína fluorescente GFP na extremidade C-terminal da proteína variante. As células HEK 293T foram transfectadas com este plasmídeo recombinante, com o objetivo de expressarem esta proteína de fusão e consequentemente analisar a sua localização na célula e eventuais efeitos no centróssoma (Figura 31).

Como controlo, as células HEK 293T foram transfectadas com um plasmídeo recombinante a expressar o TBCCD1 canónico em fusão com a GFP na C-terminal (controlo positivo) e um plasmídeo recombinante a expressar unicamente a proteína fluorescente GFP (controlo negativo). Estas células foram em seguida processadas para serem analisadas por microscopia de imunofluorescência, tendo sido incubadas com um anticorpo monoclonal para a  $\gamma$ -tubulina, que neste caso foi usada como um marcador do centróssoma.



**Figura 31: Análise de imunofluorescência indireta da proteína variante TBCCD1 humana em células de mamífero através da expressão de proteínas de fusão com proteínas fluorescentes.**

Células HEK 293T a expressar a proteína fluorescente GFP **(a)**, a proteína de fusão TBCCD1-GFP **(b)** e a proteína variante TBCCD1 em fusão com GFP **(c, d, e)** foram processadas para análise de imunofluorescência indireta com o anticorpo contra  $\gamma$ -tubulina. **(a)** As cabeças das setas apontam para a  $\gamma$ -tubulina nos centrossomas e para a ausência da proteína GFP nesta estrutura. **(b)** As cabeças das setas apontam para a  $\gamma$ -tubulina e para a proteína de fusão TBCCD1-GFP nos centrossomas. **(c e d)** As cabeças das setas apontam para a  $\gamma$ -tubulina nos centrossomas e para a ausência da proteína variante TBCCDD1 nesta estrutura. **(e)** As cabeças das setas mostram a acumulação da variante do TBCCD1 em

torno ou no invólucro nuclear. Os resultados apresentados são representativos de um total de duas experiências independentes. O DNA foi marcado com DAPI. As barras de escala correspondem a 10  $\mu$ m.

Este estudo permitiu observar que a variante TBCCD1-GFP apresentava-se difusamente no citoplasma havendo uma acumulação em torno do núcleo (Figura 31; c, d, e) mas ao contrário do TBCCD1-GFP canónico, não se localizava no centrossoma em células em interfase como se pode observar pela não localização com a  $\gamma$ -tubulina (Figura 31; c, d).

Estes resultados mostram que, ao contrário dos primeiros 20 resíduos de aminoácidos essenciais à localização do TBCCD1 canónico (Figura 31; b), os novos 16 resíduos de aminoácidos da variante TBCCD1 constituem um domínio N-terminal que não é competente para localizar a proteína no centrossoma, levando-a a espalhar-se pelo citoplasma onde apresenta uma clara acumulação preferencial em torno do invólucro nuclear (Figura 31; c, d, e). Deste modo, a variante TBCCD1 poderá desempenhar uma função distinta da do TBCCD1 canónico ou alternativamente apresentar uma função sinérgica à do TBCCD1 canónico na manutenção da ligação do centrossoma ao núcleo.

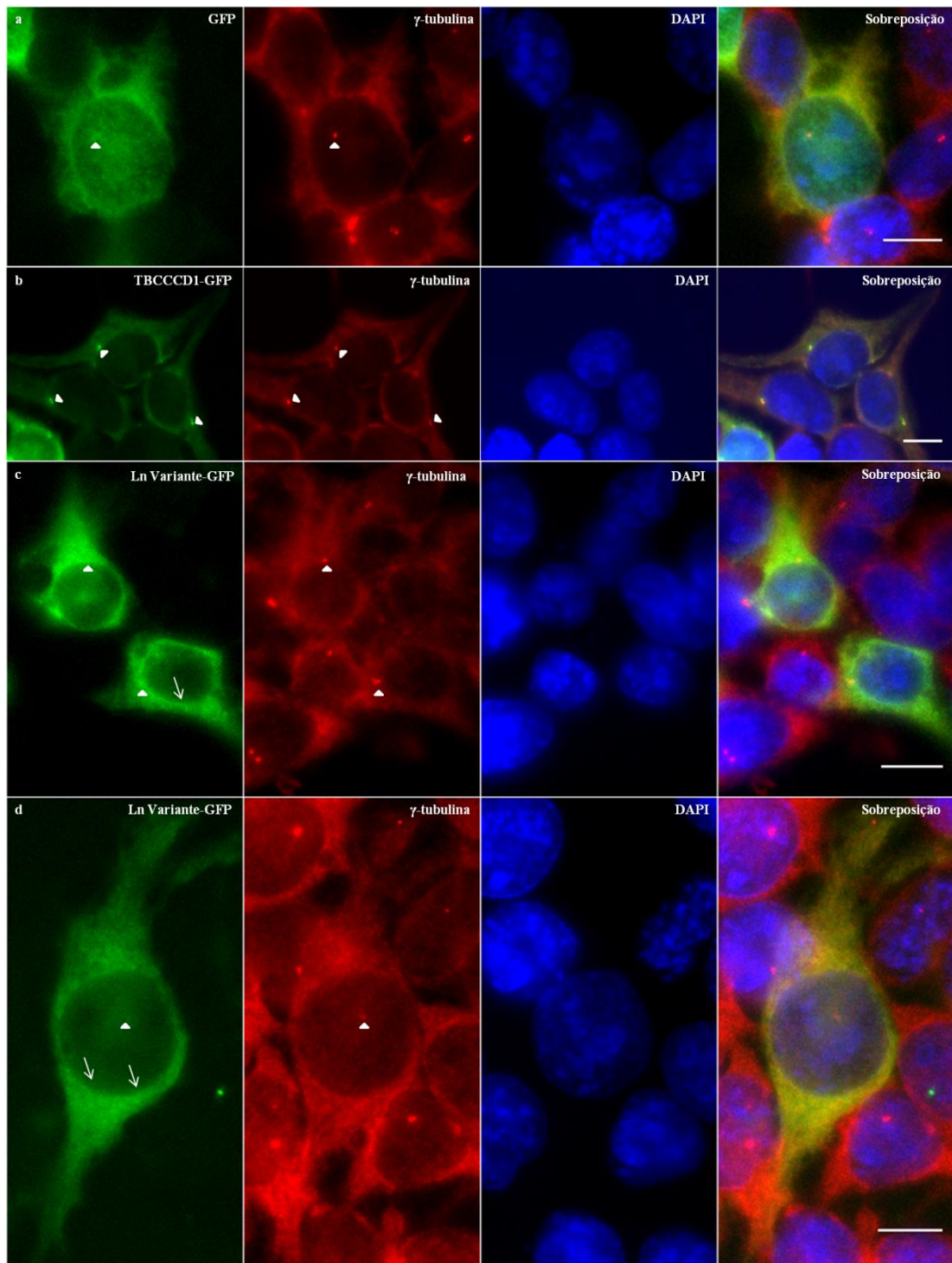
Estes resultados não são inesperados porque encontram-se disponíveis na literatura numerosos exemplos em que os isótipos de proteínas produzidos a partir de um *splicing* alternativo podem diferir na estrutura, função, localização e outras propriedades (Wang, Sandberg et al. 2008).

### **III.6 Estudo da localização celular do transcrito alternativo de maiores dimensões do gene *tbccd1* por imunofluorescência nas células humanas HEK 293T**

Com os resultados obtidos anteriormente procurou-se clarificar algumas das suposições propostas para o RNA do transcrito alternativo ser expresso nas diferentes linhas e tecidos humanos. As hipóteses iniciais para a existência deste transcrito apoiavam-se no facto de este ser um precursor não maduro, onde os codões stop

encontrados fizessem parte do intrão que neste caso conteria parte da sequência do exão 2 incluindo o AUG<sub>TBCCD1</sub> ou por outro lado, a sequência nucleotídica entre os dois uAUG ser uma pequena ORF e juntamente com o AUG<sub>previsto</sub> terem um papel regulador da tradução do mRNA canônico que origina o TBCCD1 a partir do AUG<sub>TBCCD1</sub>. Assim, efetuou-se o estudo da localização do transcrito alternativo de maiores dimensões do gene *tbccd1*.

Para se estudar se era produzida alguma proteína a partir transcrito alternativo do gene *tbccd1* contendo o AUG<sub>previsto</sub> e o AUG<sub>TBCCD1</sub>, a sua sequência nucleotídica foi clonada num vetor de expressão em fusão com a proteína fluorescente GFP na sua extremidade C-terminal. Esta construção permitiria ainda, se houvesse a produção de uma proteína, determinar a sua localização celular. Deste modo o plasmídeo recombinante obtido foi em seguida utilizado para transfetar células HEK 293T, com o objetivo de expressarem esta proteína de fusão (Figura 32). Como controlo, as células HEK 293T foram transfectadas com um plasmídeo recombinante a expressar o TBCCD1 canônico em fusão com a GFP na C-terminal (controlo positivo) e um plasmídeo recombinante a expressar unicamente a proteína fluorescente GFP (controlo negativo). Estas células foram em seguida processadas para serem analisadas por microscopia de imunofluorescência indireta, tendo sido incubadas com um anticorpo monoclonal para a  $\gamma$ -tubulina que neste caso foi usada como um marcador do centrossoma.



**Figura 32: Análise de imunofluorescência indireta do transcrito alternativo de maiores dimensões do TBCCD1 em células de mamífero através da expressão de proteínas de fusão com proteínas fluorescentes.** Células HEK 293T a expressar a proteína fluorescente GFP (a), a proteína de fusão TBCCD1-GFP (b) e o transcrito alternativo de maiores dimensões do TBCCD1 em fusão com a GFP [descrito na imagem como Ln (*long*) Variante-GFP] (c, d) foram processadas para análise de imunofluorescência indireta com o anticorpo contra  $\gamma$ -tubulina. (a) As cabeças das setas apontam

para a  $\gamma$ -tubulina nos centrossomas e para a ausência da proteína GFP nesta estrutura. **(b)** As cabeças das setas apontam para a  $\gamma$ -tubulina e para a proteína de fusão TBCCD1-GFP nos centrossomas. **(c e d)** As cabeças das setas apontam para a  $\gamma$ -tubulina nos centrossomas e para a ausência do transcrito alternativo de maiores dimensões do TBCCD1 nesta estrutura. As setas mostram a acumulação do transcrito alternativo de maiores dimensões do TBCCD1 em torno ou no invólucro nuclear. Os resultados apresentados são representativos de um total de duas experiências independentes. O DNA foi marcado com DAPI. As barras de escala correspondem a 10  $\mu$ m.

É importante referir que devido a problemas ocorridos durante a sua amplificação este clone apresenta uma mutação (devido a um possível erro cometido pela *taq* polimerase) num resíduo de aminoácido na posição 384 do TBCCD1 canónico, existindo a troca de um aminóácido de cisteína por tirosina. Note-se que esta mutação está localizada na C-terminal da proteína e portanto à partida não teria implicações no estudo da localização deste transcrito. Tal como foi descrito detalhadamente ao longo do presente estudo, é a região N-terminal da proteína canónica a responsável pela sua localização no centrossoma. Esta mutação poderia, no entanto, ter influência no caso do estudo da função deste transcrito

Após a análise por imunofluorescência indireta da localização da proteína TBCCD1 de maiores dimensões, uma das primeiras observações é a de que, embora a sequência nucleotídica do transcrito que a codifica apresente uma pequena uORF a montante do uAUG do TBCCD1 canónico, os resultados sugerem que a proteína é produzida. Outra observação importante é que tal como a variante do TBCCD1 e ao contrário do TBCCD1 canónico (Figura 32; b), esta proteína não se acumula no centrossoma (Figura 31; c, d). Além disso, a proteína produzida a partir da expressão deste transcrito de maiores dimensões parece acumular-se preferencialmente em torno do núcleo, parecendo concentrar-se na zona perinuclear (Figura 32; c, d). Os resultados aqui apresentados sugerem que a proteína produzida partir da expressão deste transcrito provavelmente é a variante do TBCCD1 apoiando fortemente a ideia inicial de este transcrito ser um precursor não maduro, onde os codões stop descritos fazem parte de um intrão. Este intrão, por sua vez contém a totalidade do exão 2 onde se localiza o AUG<sub>TBCCD1</sub>. Neste caso a formação deste transcrito requeria primeiramente a remoção do grande intrão de 5734 pb, remoção esta que gera a

uORF, a qual poderá subsequentemente ser removida sobre a forma de um intrão gerando a nova variante TBCCD1. No entanto, não é possível ainda afastar a ideia de que o grande transcrito, antes da remoção da uORF, possa ter um papel regulador da expressão do TBCCD1 canônico, tendo deste modo dupla função. Seria assim extremamente interessante investigar os níveis de TBCCD1 canônico em células a sobre-expressar este grande transcrito.

## Capítulo IV – Conclusões e Perspetivas Futuras

O TBCCD1 é uma proteína humana que está relacionada em termos de sequência com o cofactor C da tubulina e com a proteína RP2, uma vez que as três partilham domínios funcionais, sendo eles o domínio TBCC e o domínio CARP (Gonçalves, Nolasco et al. 2010a).

Tal como foi detalhadamente descrito ao longo deste trabalho, verificou-se que nas células humanas, o TBCCD1 é um componente do centrossoma, sendo a sua região N-terminal, nomeadamente os primeiros 20 resíduos de aminoácidos, responsável por esta localização (Gonçalves 2010).

O principal objetivo do trabalho aqui apresentado foi contribuir para o estudo da função biológica da proteína TBCCD1, nomeadamente estudar a regulação da expressão do gene *tbccd1* investigando a possibilidade de ocorrência de *splicing* alternativo. Sabe-se que este mecanismo é um passo essencial para a expressão de genes permitindo o aumento da capacidade de codificação de um único gene. Desta forma proporciona a síntese de vários isótipos de proteínas estruturalmente e funcionalmente diferentes (Caceres and Kornblihtt 2002). Sendo uma das principais fontes para aumentar a diversidade proteómica e o número de produtos codificados por um genoma.

De um modo geral, podemos dizer que o objetivo a que nos propusemos alcançar com a realização deste trabalho foi conseguido. Verificou-se que o gene que codifica para a proteína TBCCD1 origina três transcritos diferentes pelo processo de *splicing* alternativo. O transcrito inicialmente descrito por Gonçalves et al (2010) e que codifica para o TBCCD1 dito canónico, um transcrito de maiores dimensões quando comparado ao que codifica o TBCCD1 canónico e um de menores dimensões. Estes transcritos resultam da existência de um exão alternativo, que origina duas proteínas putativas diferindo do TBCCD1 canónico nos primeiros resíduos de aminoácidos da sua N-terminal. Apesar de se ter verificado que a produção de estes transcritos é

devida ao mecanismo de *splicing* alternativo não é, no entanto, claro de que forma este processo ocorre. Sugerimos que estes transcritos surgem (1) de um uso de promotores alternativos, originando um isótipo diferente do TBCDD1 canónico ou por outro lado (2) o transcrito de maiores dimensões contendo o AUG<sub>previsto</sub> e o AUG<sub>TBCDD1</sub> é simultaneamente capaz de regular a produção de TBCDD1 canónico através da sua uORF e de ser um precursor da variante TBCDD1.

Em suma, o estudo da localização dos transcritos originados pelo gene *tbccd1* permitiram verificar que a variante TBCDD1 não se localiza no centróssoma, o que nos permite concluir que a sua região N-terminal não é competente para localizá-la no centróssoma. Por sua vez, esta parece ter uma acumulação preferencial em torno do invólucro nuclear. Por outro lado, quanto ao transcrito de maiores dimensões os resultados sugerem que a proteína produzida a partir da sua sobre-expressão provavelmente é a variante do TBCDD1, pois a análise por imunofluorescência indireta sugere uma localização idêntica, apoiando desta forma a ideia de este transcrito ser um precursor não maduro da variante TBCDD1.

De forma a completar o estudo iniciado e apresentado neste trabalho sobre a proteína TBCDD1 humana e a sua variante seria extremamente importante a realização de várias experiências no futuro que visam continuar a caracterização desta proteína, bem como dos transcritos alternativos, nomeadamente:

- Aprofundar e comparar o estudo da expressão dos dois transcritos (canónico e variante) em diferentes tecidos humanos tentando estabelecer o seu padrão de expressão e abundância relativa, o que poderia, eventualmente, dar pistas sobre as funções específicas de cada proteína;
- Aprofundar o estudo da localização intracelular da variante TBCDD1 usando, por exemplo, anticorpos que marquem o complexo de Golgi ou o retículo endoplasmático, ou ainda marcadores do invólucro nuclear de forma a apurar com maior detalhe a sua localização e estudar a localização da variante nas

diferentes etapas da mitose e células com cílios primários e comparar com o TBCCD1 canónico;

- Investigar os níveis de TBCCD1 canónico em células a sobre-expressar o transcrito de maiores dimensões, de forma a verificar uma eventual regulação por parte deste transcrito sobre a proteína canónica;
- De forma a investigar o papel do transcrito maior poder-se-ia fazer *constructs* com a ORF presente neste colocando-a, por exemplo, na 5'UTR de outros transcritos que se saiba serem estáveis e codifiquem para proteínas bem conhecidas e deste modo verificar a sua estabilidade ou taxa traducional.
- De forma a distinguir as funções da variante TBCCD1 e do TBCCD1 canónico poder-se-ia recorrer à técnica de RNAi para deletar os dois transcritos. No entanto, seria uma estratégia complicada visto que as sequências dos transcritos são muito semelhantes. A única hipótese seria usando um siRNA que emparelha-se especificamente com a região N-terminal da variante TBCCD1. Devido a esta mesma razão também seria difícil produzir anticorpos capazes de distinguir entre as duas sequências sendo que a opção seria a produção de um anticorpo contra um péptido específico na N-terminal da variante TBCCD1.

Em suma, este trabalho completou os estudos iniciais em torno da proteína TBCCD1, permitindo uma nova abordagem sobre alguns mecanismos que regulam a expressão do gene que a codifica, mostrando que existem diferentes isótipos da proteína produzidos a partir de um *splicing* alternativo, diferendo do TBCCD1 no que diz respeito à localização.

## Referências bibliográficas:

Amos, L. A. and D. Schlieper (2005). "Microtubules and maps." Adv Protein Chem **71**: 257-298.

Andersen, J. S., C. J. Wilkinson, et al. (2003). "Proteomic characterization of the human centrosome by protein correlation profiling." Nature **426**(6966): 570-574.

Bartolini, F., A. Bhamidipati, et al. (2002). "Functional overlap between retinitis pigmentosa 2 protein and the tubulin-specific chaperone cofactor C." J Biol Chem **277**(17): 14629-14634.

Bendtsen, J. D., H. Nielsen, et al. (2004). "Improved prediction of signal peptides: SignalP 3.0." J Mol Biol **340**(4): 783-795.

Bettencourt-Dias, M. and D. M. Glover (2007). "Centrosome biogenesis and function: centrosomics brings new understanding." Nat Rev Mol Cell Biol **8**(6): 451-463.

Bhamidipati, A., S. A. Lewis, et al. (2000). "ADP ribosylation factor-like protein 2 (Arl2) regulates the interaction of tubulin-folding cofactor D with native tubulin." J Cell Biol **149**(5): 1087-1096.

Black, D. L. (2003). "Mechanisms of alternative pre-messenger RNA splicing." Annu Rev Biochem **72**: 291-336.

Blencowe, B. J. (2000). "Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases." Trends Biochem Sci **25**(3): 106-110.

Blencowe, B. J. (2006). "Alternative splicing: new insights from global analyses." Cell **126**(1): 37-47.

Bouissou, A., C. Verollet, et al. (2009). " $\gamma$ -Tubulin ring complexes regulate microtubule plus end dynamics." J Cell Biol **187**(3): 327-334.

Buratti, E. and F. E. Baralle (2004). "Influence of RNA secondary structure on the pre-mRNA splicing process." Mol Cell Biol **24**(24): 10505-10514.

Buvoli, M., A. Buvoli, et al. (2007). "Interplay between exonic splicing enhancers, mRNA processing, and mRNA surveillance in the dystrophic Mdx mouse." PLoS One **2**(5): e427.

Caceres, J. F. and A. R. Kornblihtt (2002). "Alternative splicing: multiple control mechanisms and involvement in human disease." Trends Genet **18**(4): 186-193.

Cartegni, L., S. L. Chew, et al. (2002). "Listening to silence and understanding nonsense: exonic mutations that affect splicing." Nat Rev Genet **3**(4): 285-298.

Catania, F. and M. Lynch (2013). "A simple model to explain evolutionary trends of eukaryotic gene architecture and expression: how competition between splicing and cleavage/polyadenylation factors may affect gene expression and splice-site recognition in eukaryotes." Bioessays **35**(6): 561-570.

Chabin-Brion, K., J. Marceiller, et al. (2001). "The Golgi complex is a microtubule-organizing organelle." Mol Biol Cell **12**(7): 2047-2060.

Chae, S., C. Yun, et al. (2005). "Centrosome amplification and multinuclear phenotypes are induced by hydrogen peroxide." Exp Mol Med **37**(5): 482-487.

Chapple, J. P., C. Grayson, et al. (2003). "Organization on the plasma membrane of the retinitis pigmentosa protein RP2: investigation of association with detergent-resistant membranes and polarized sorting." Biochem J **372**(Pt 2): 427-433.

Chapple, J. P., A. J. Hardcastle, et al. (2000). "Mutations in the N-terminus of the X-linked retinitis pigmentosa protein RP2 interfere with the normal targeting of the protein to the plasma membrane." Hum Mol Genet **9**(13): 1919-1926.

Chapple, J. P., A. J. Hardcastle, et al. (2002). "Delineation of the plasma membrane targeting domain of the X-linked retinitis pigmentosa protein RP2." Invest Ophthalmol Vis Sci **43**(6): 2015-2020.

Chasin, L. A. (2007). "Searching for splicing motifs." Adv Exp Med Biol **623**: 85-106.

Cheeseman, I. M. and A. Desai (2005). "A combined approach for the localization and tandem affinity purification of protein complexes from metazoans." Sci STKE **2005**(266): p11.

Chen, M. and J. L. Manley (2009). "Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches." Nat Rev Mol Cell Biol **10**(11): 741-754.

Churbanov, A., I. B. Rogozin, et al. (2005). "Evolutionary conservation suggests a regulatory function of AUG triplets in 5'-UTRs of eukaryotic genes." Nucleic Acids Res **33**(17): 5512-5520.

Conde, C. and A. Caceres (2009). "Microtubule assembly, organization and dynamics in axons and dendrites." Nat Rev Neurosci **10**(5): 319-332.

Crabb, T. L., B. J. Lam, et al. (2010). "Retention of spliceosomal components along ligated exons ensures efficient removal of multiple introns." RNA **16**(9): 1786-1796.

Cunningham, L. A. and R. A. Kahn (2008). "Cofactor D functions as a centrosomal protein and is required for the recruitment of the gamma-tubulin ring complex at centrosomes and organization of the mitotic spindle." J Biol Chem **283**(11): 7155-7165.

De Conti, L., M. Baralle, et al. (2013). "Exon and intron definition in pre-mRNA splicing." Wiley Interdiscip Rev RNA **4**(1): 49-60.

Debec, A., W. Sullivan, et al. (2010). "Centrioles: active players or passengers during mitosis?" Cell Mol Life Sci **67**(13): 2173-2194.

Delaval, B. and S. J. Doxsey (2010). "Pericentrin in cellular function and disease." J Cell Biol **188**(2): 181-190.

Desai, A. and T. J. Mitchison (1997). "Microtubule polymerization dynamics." Annu Rev Cell Dev Biol **13**: 83-117.

Donahue, C. P., C. Muratore, et al. (2006). "Stabilization of the tau exon 10 stem loop alters pre-mRNA splicing." J Biol Chem **281**(33): 23302-23306.

Doxsey, S. (2001). "Re-evaluating centrosome function." Nat Rev Mol Cell Biol **2**(9): 688-698.

Erkelenz, S., W. F. Mueller, et al. (2013). "Position-dependent splicing activation and repression by SR and hnRNP proteins rely on common mechanisms." RNA **19**(1): 96-102.

Evans, R. J., N. Schwarz, et al. (2010). "The retinitis pigmentosa protein RP2 links pericentriolar vesicle transport between the Golgi and the primary cilium." Hum Mol Genet **19**(7): 1358-1367.

Fanarraga, M. L., J. Bellido, et al. (2010). "TBCD links centriologenesi, spindle microtubule dynamics, and midbody abscission in human cells." PLoS One **5**(1): e8846.

Faustino, N. A. and T. A. Cooper (2003). "Pre-mRNA splicing and human disease." Genes Dev **17**(4): 419-437.

Feldman, J. L. and W. F. Marshall (2009). "ASQ2 encodes a TBCC-like protein required for mother-daughter centriole linkage and mitotic spindle orientation." Curr Biol **19**(14): 1238-1243.

Fletcher, D. A. and R. D. Mullins (2010). "Cell mechanics and the cytoskeleton." Nature **463**(7280): 485-492.

Fuller, S. D., B. E. Gowen, et al. (1995). "The core of the mammalian centriole contains gamma-tubulin." Curr Biol **5**(12): 1384-1393.

Gonçalves, J. (2010). "Caracterização da proteína centrossomal TBCCD1, em linhas celulares humanas e durante o desenvolvimento de Danio rerio." Tese de Doutorado em Genética Molecular, FC, UL.

Gonçalves, J., S. Nolasco, et al. (2010a). "TBCCD1, a new centrosomal protein, is required for centrosome and Golgi apparatus positioning." EMBO Rep **11**(3): 194-200.

Gonçalves, J., A. Tavares, et al. (2010b). "Revisiting the tubulin folding pathway: new roles in centrosomes and cilia." BioMolecular Concepts **1**: 423-434.

Goren, A., O. Ram, et al. (2006). "Comparative analysis identifies exonic splicing regulatory sequences--The complex definition of enhancers and silencers." Mol Cell **22**(6): 769-781.

Graveley, B. R. (2000). "Sorting out the complexity of SR protein functions." RNA **6**(9): 1197-1211.

Grayson, C., F. Bartolini, et al. (2002). "Localization in the human retina of the X-linked retinitis pigmentosa protein RP2, its homologue cofactor C and the RP2 interacting protein Arl3." Hum Mol Genet **11**(24): 3065-3074.

Guillet, V., M. Knibiehler, et al. (2011). "Crystal structure of gamma-tubulin complex protein GCP4 provides insight into microtubule nucleation." Nat Struct Mol Biol **18**(8): 915-919.

Hage-Sleiman, R., S. Herveau, et al. (2010). "Tubulin binding cofactor C (TBCC) suppresses tumor growth and enhances chemosensitivity in human breast cancer cells." BMC Cancer **10**: 135.

Haren, L., T. Stearns, et al. (2009). "Plk1-dependent recruitment of gamma-tubulin complexes to mitotic centrosomes involves multiple PCM components." PLoS One **4**(6): e5976.

Hastings, M. L. and A. R. Krainer (2001). "Pre-mRNA splicing in the new millennium." Curr Opin Cell Biol **13**(3): 302-309.

Heald, R. and E. Nogales (2002). "Microtubule dynamics." J Cell Sci **115**(Pt 1): 3-4.

Hiller, M., Z. Zhang, et al. (2007). "Pre-mRNA secondary structures influence exon recognition." PLoS Genet **3**(11): e204.

Hubberstey, A. V. and E. P. Mottillo (2002). "Cyclase-associated proteins: CAPacity for linking signal transduction and actin polymerization." FASEB J **16**(6): 487-499.

Iacono, M., F. Mignone, et al. (2005). "uAUG and uORFs in human and rodent 5'untranslated mRNAs." Gene **349**: 97-105.

Irimia, M., D. Penny, et al. (2007a). "Coevolution of genomic intron number and splice sites." Trends Genet **23**(7): 321-325.

Irimia, M., J. L. Rukov, et al. (2007b). "Functional and evolutionary analysis of alternatively spliced genes is consistent with an early eukaryotic origin of alternative splicing." BMC Evol Biol **7**: 188.

Jin, Y., Y. Yang, et al. (2011). "New insights into RNA secondary structure in the alternative splicing of pre-mRNAs." RNA Biol **8**(3): 450-457.

Jurica, M. S. and M. J. Moore (2003). "Pre-mRNA splicing: awash in a sea of proteins." Mol Cell **12**(1): 5-14.

Ke, S. and L. A. Chasin (2011). "Context-dependent splicing regulation: exon definition, co-occurring motif pairs and tissue specificity." RNA Biol **8**(3): 384-388.

Kelemen, O., P. Convertini, et al. (2013). "Function of alternative splicing." Gene **514**(1): 1-30.

Keren, H., G. Lev-Maor, et al. (2010). "Alternative splicing and evolution: diversification, exon definition and function." Nat Rev Genet **11**(5): 345-355.

Kim, E., A. Goren, et al. (2008). "Alternative splicing: current perspectives." Bioessays **30**(1): 38-47.

Kobayashi, T. and B. D. Dynlacht (2011). "Regulating the transition from centriole to basal body." J Cell Biol **193**(3): 435-444.

Kochetov, A. V. (2008). "Alternative translation start sites and hidden coding potential of eukaryotic mRNAs." Bioessays **30**(7): 683-691.

Kollman, J. M., A. Merdes, et al. (2011). "Microtubule nucleation by gamma-tubulin complexes." Nat Rev Mol Cell Biol **12**(11): 709-721.

Kollman, J. M., J. K. Polka, et al. (2010). "Microtubule nucleating gamma-TuSC assembles structures with 13-fold microtubule-like symmetry." Nature **466**(7308): 879-882.

Kortazar, D., M. L. Fanarraga, et al. (2007). "Role of cofactors B (TBCB) and E (TBCE) in tubulin heterodimer dissociation." Exp Cell Res **313**(3): 425-436.

Kozak, M. (1987). "An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs." Nucleic Acids Res **15**(20): 8125-8148.

Ladd, A. N. and T. A. Cooper (2002). "Finding signals that regulate alternative splicing in the post-genomic era." Genome Biol **3**(11): reviews0008.

Lange, B. M. (2002). "Integration of the centrosome in cell cycle control, stress response and signal transduction pathways." Curr Opin Cell Biol **14**(1): 35-43.

Lawo, S., M. Hasegan, et al. (2012). "Subdiffraction imaging of centrosomes reveals higher-order organizational features of pericentriolar material." Nat Cell Biol **14**(11): 1148-1158.

Letunic, I., R. R. Copley, et al. (2002). "Common exon duplication in animals and its role in alternative splicing." Hum Mol Genet **11**(13): 1561-1567.

Li, Q., J. A. Lee, et al. (2007). "Neuronal regulation of alternative pre-mRNA splicing." Nat Rev Neurosci **8**(11): 819-831.

Li, R. and G. G. Gundersen (2008). "Beyond polymer polarity: how the cytoskeleton builds a polarized cell." Nat Rev Mol Cell Biol **9**(11): 860-873.

Lim, L. P. and C. B. Burge (2001). "A computational analysis of sequence features involved in recognition of short introns." Proc Natl Acad Sci U S A **98**(20): 11193-11198.

Llorca, O., J. Martin-Benito, et al. (2000). "Eukaryotic chaperonin CCT stabilizes actin and tubulin folding intermediates in open quasi-native conformations." EMBO J **19**(22): 5971-5979.

Lopez-Fanarraga, M., J. Avila, et al. (2001). "Review: postchaperonin tubulin folding cofactors and their role in microtubule dynamics." J Struct Biol **135**(2): 219-229.

Luco, R. F., M. Allo, et al. (2011). "Epigenetics in alternative pre-mRNA splicing." Cell **144**(1): 16-26.

Luders, J. and T. Stearns (2007). "Microtubule-organizing centres: a re-evaluation." Nat Rev Mol Cell Biol **8**(2): 161-167.

Maris, C., C. Dominguez, et al. (2005). "The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression." FEBS J **272**(9): 2118-2131.

Maston, G. A., S. K. Evans, et al. (2006). "Transcriptional regulatory elements in the human genome." Annu Rev Genomics Hum Genet **7**: 29-59.

Matlin, A. J., F. Clark, et al. (2005). "Understanding alternative splicing: towards a cellular code." Nat Rev Mol Cell Biol **6**(5): 386-398.

Meijer, H. A. and A. A. Thomas (2002). "Control of eukaryotic protein synthesis by upstream open reading frames in the 5'-untranslated region of an mRNA." Biochem J **367**(Pt 1): 1-11.

Misteli, T. and D. L. Spector (1998). "The cellular organization of gene expression." Curr Opin Cell Biol **10**(3): 323-331.

Morris, D. R. and A. P. Geballe (2000). "Upstream open reading frames as regulators of mRNA translation." Mol Cell Biol **20**(23): 8635-8642.

Nagao, I. and J. Obokata (2006). "In vitro selection of translational regulatory elements." Anal Biochem **354**(1): 1-7.

Neto Ade, M., D. A. Alvarenga, et al. (2012). "Improving N-terminal protein annotation of Plasmodium species based on signal peptide prediction of orthologous proteins." Malar J **11**: 375.

Nilsen, T. W. and B. R. Graveley (2010). "Expansion of the eukaryotic proteome by alternative splicing." Nature **463**(7280): 457-463.

Nogales, E. (2001). "Structural insight into microtubule function." Annu Rev Biophys Biomol Struct **30**: 397-420.

Nolasco, S., J. Bellido, et al. (2005). "Tubulin cofactor A gene silencing in mammalian cells induces changes in microtubule cytoskeleton, cell cycle arrest and cell death." FEBS Lett **579**(17): 3515-3524.

Ota, T., Y. Suzuki, et al. (2004). "Complete sequencing and characterization of 21,243 full-length human cDNAs." Nat Genet **36**(1): 40-45.

Pagani, F. and F. E. Baralle (2004). "Genomic variants in exons and introns: identifying the splicing spoilers." Nat Rev Genet **5**(5): 389-396.

Pelletier, L. and Y. M. Yamashita (2012). "Centrosome asymmetry and inheritance during animal development." Curr Opin Cell Biol **24**(4): 541-546.

Pereira, G. and E. Schiebel (1997). "Centrosome-microtubule nucleation." J Cell Sci **110** ( Pt 3): 295-300.

Pereira, I. (2011). "Estudo da proteína humana centrossomal TBCCD1: Determinação de domínios funcionais." Tese de Mestrado em Bioquímica, FC,UL.

Pleiss, J. A., G. B. Whitworth, et al. (2007). "Transcript specificity in yeast pre-mRNA splicing revealed by mutations in core spliceosomal components." PLoS Biol **5**(4): e90.

Raynaud-Messina, B. and A. Merdes (2007). "Gamma-tubulin complexes and microtubule organization." Curr Opin Cell Biol **19**(1): 24-30.

Remy, M. H., A. Merdes, et al. (2013). "Assembly of gamma-tubulin ring complexes: implications for cell biology and disease." Prog Mol Biol Transl Sci **117**: 511-530.

Sachs, M. S. and A. P. Geballe (2006). "Downstream control of upstream open reading frames." Genes Dev **20**(8): 915-921.

Satir, P. and S. T. Christensen (2007). "Overview of structure and function of mammalian cilia." Annu Rev Physiol **69**: 377-400.

Schmucker, D., J. C. Clemens, et al. (2000). "Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity." Cell **101**(6): 671-684.

Schwahn, U., S. Lenzner, et al. (1998). "Positional cloning of the gene for X-linked retinitis pigmentosa 2." Nat Genet **19**(4): 327-332.

Shepard, P. J. and K. J. Hertel (2008). "Conserved RNA secondary structures promote alternative splicing." RNA **14**(8): 1463-1469.

Singer, R. H. and M. R. Green (1997). "Compartmentalization of eukaryotic gene expression: causes and effects." Cell **91**(3): 291-294.

Singh, R. and J. Valcarcel (2005). "Building specificity with nonspecific RNA-binding proteins." Nat Struct Mol Biol **12**(8): 645-653.

Sluder, G. and A. Khodjakov (2010). "Centriole duplication: analogue control in a digital age." Cell Biol Int **34**(12): 1239-1245.

Smith, C. W. and J. Valcarcel (2000). "Alternative pre-mRNA splicing: the logic of combinatorial control." Trends Biochem Sci **25**(8): 381-388.

Stephan, A., S. Vaughan, et al. (2007). "An essential quality control mechanism at the eukaryotic basal body prior to intraflagellar transport." Traffic **8**(10): 1323-1330.

Szolajska, E. and J. Chroboczek (2011). "Faithful chaperones." Cell Mol Life Sci **68**(20): 3307-3322.

Tian, G., A. Bhamidipati, et al. (1999). "Tubulin folding cofactors as GTPase-activating proteins. GTP hydrolysis and the assembly of the alpha/beta-tubulin heterodimer." J Biol Chem **274**(34): 24054-24058.

Tian, G., Y. Huang, et al. (1996). "Pathway leading to correctly folded beta-tubulin." Cell **86**(2): 287-296.

Tian, G., S. A. Lewis, et al. (1997). "Tubulin subunits exist in an activated conformational state generated and maintained by protein cofactors." J Cell Biol **138**(4): 821-832.

Verollet, C., N. Colombie, et al. (2006). "Drosophila melanogaster gamma-TuRC is dispensable for targeting gamma-tubulin to the centrosome and microtubule nucleation." J Cell Biol **172**(4): 517-528.

Wade, R. H. (2009). "On and around microtubules: an overview." Mol Biotechnol **43**(2): 177-191.

Wang, E., W. F. Mueller, et al. (2011). "G Run-mediated recognition of proteolipid protein and DM20 5' splice sites by U1 small nuclear RNA is regulated by context and proximity to the splice site." J Biol Chem **286**(6): 4059-4071.

Wang, E. T., R. Sandberg, et al. (2008). "Alternative isoform regulation in human tissue transcriptomes." Nature **456**(7221): 470-476.

Wang, G. S. and T. A. Cooper (2007). "Splicing in disease: disruption of the splicing code and the decoding machinery." Nat Rev Genet **8**(10): 749-761.

Wang, Z. and C. B. Burge (2008). "Splicing regulation: from a parts list of regulatory elements to an integrated splicing code." RNA **14**(5): 802-813.

Wiese, C. and Y. Zheng (2006). "Microtubule nucleation: gamma-tubulin and beyond." J Cell Sci **119**(Pt 20): 4143-4153.

Will, C. L. and R. Luhrmann (2011). "Spliceosome structure and function." Cold Spring Harb Perspect Biol **3**(7).

Xu, Q., B. Modrek, et al. (2002). "Genome-wide detection of tissue-specific alternative splicing in the human transcriptome." Nucleic Acids Res **30**(17): 3754-3766.

Yilmaz, A. and E. Grotewold (2010). "Components and mechanisms of regulation of gene expression." Methods Mol Biol **674**: 23-32.

Yoon, J. H., J. Qiu, et al. (2006). "The retinitis pigmentosa-mutated RP2 protein exhibits exonuclease activity and translocates to the nucleus in response to DNA damage." Exp Cell Res **312**(8): 1323-1334.

Zahler, A. M. (2005). "Alternative splicing in *C. elegans*." WormBook: 1-13.

Zheng, C. L., X. D. Fu, et al. (2005). "Characteristics and regulatory elements defining constitutive splicing and different modes of alternative splicing in human and mouse." RNA **11**(12): 1777-1787.

# Anexos

## 1. Sequência nucleotídica da região do gene *tbbcd1* onde ocorre o *splicing* alternativo dos transcritos descritos ao longo do presente trabalho

>chromosome:GRCh37:3:186263262:186288932:-1

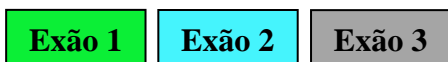
```
1 GTTCTCCTCTCACCCCACTAGCCTGGCCATACTACCCCTTCTCCGCCACCCCGCCAAC 60
61 CCACAGCGATGCCCCCTCCACTCCGTCAGTCTGGCAATGCCACTTCCCCACCCCTCC 120
121 CACCCCCAGCAAAGGCTGACCCACAGCCCCGTTGTCTGCGAGTGTCTGTTCCCTACCC 180
181 GGCAATCACCGCCCCGATGAGGTATAGCAGCAACAGGCAAAAGGTGCTCAGGTTCTGCGG 240
241 AGCCATGGTTCTCTGTCCCGGTCTGTCCACACACTCCTCACAGCCTCCTCCGCCGC 300
301 CGCGTCGGCTCGCCAGCCACCCGGCCCTGTGAGGCCGCTCACACCGGGGGGGCGCGG 360
361 GGTCTCCTTGGTCCGCAGAGACAGCTAGCTAGCCCCCTTCTGCCGGTCCACTTCCCGG 420
421 GAGTCCCGAGTCCCGGTGAGAGAAGCCCCGGTAGGCCGAGAGCCAATCGTTGCCCGGGAC 480
481 ATCACACGCGCCGGCGCTGGAAGCCAATCAGCGCATACACTTACCTCCGAGGCGGC 540
541 ACGGGCAACCAGTTAGGCTTCAAGCTGCAAGTGCAGCTCTTGACCAATAGAAAGCGGAG 600
601 ACGTCAGAGGGTCGCGCCGCTGCAGGAACCCGCTCTGCCTGAGCTACCGCGAGTCGTAGA 660
661 AAATTACGTAATGATAGGGGGCCGGTCAGAGTGAAGGCACCTGAGGTACACGAAATCCCT 720
721 CCGGCTTGCCCACTTCTCCGACTTTCCACCAATCAGGAGCTTCTCCCCAGCACCAACC 780
781 AATTGTGGGAGAGAAGTTTTCTGTGTCTACAAGCCCGGAATTTCAATCATACCCAGCTA 840
841 CGCTCCGAATGAGGGAGCAGAGGTACAGCGGCCGAATGAGCGGGAACAGTTATTCCACGA 900
901 CAGGTGTACATCAGACGCATTGCTTAGCCCTCTTATTTTACGTTCTCCAAGCAACTCCAT 960
961 AAGACAACGTCTCGGAAAGGCTACAGTGATGACAGACCGACTTAAGTTCGGATTCCGGCT 1020
1021 TCCACCGCTTCTTAGTGCTGAGACCCCCACCCCTCCCTCCTCCCAAGTCTGAGCCTAA 1080
1081 GTTTATCAACAAAACCTGAGGCAAAGCACTTATTGGCAGTCTGTGTAATTTAACAGGGAA 1140
1141 AAAAATCGGTAATCATCGGGCGTATAGTAAGCAGGCAACGAGTAAAAACCATATTACG 1200
1201 GTTCAAGCAGAACTCTTTGATTCCCCATCTTCCCCCGATTTTGCTCATCGCAGTTACAC 1260
1261 AACACCTAGTTGATCAGGACAAAGCTTAGGTGTTTTTATTCTTTTATGCATAAGCTAT 1320
1321 CCAACTCTTTACCTCTGAACATACTTCAATCTGTTCACTACATCTCCATTACCAGGATT 1380
1381 CTGGGTACAGTGTATCTCTCACCTGGCCAGGGCCTCCTAACTGGCCTCCTTGCCCCAC 1440
1441 TCTTAATATCTTCTTAATATCTTAATATATGTTGCTCATCCCTGTCCCTCACTCCATGA 1500
1501 TTTAAAAAAGCCGCGGGCGCGGTGGCTCACCCACCCAAAGTGAATCAGC 1560
1561 ACTTTGCGAGGCCGAGACGGGCGGAGCACCTGAGGTGAGGAGCTTGAACAGGCTGGCC 1620
1621 AACATGATGAAACCCGTCTCTACTAAAACTACAAAAATTAGCCAGGTGTGGTGGCGAGC 1680
1681 GCCGTAAACCCAGCTGCTCGGGAGCCTGAGGCAGGAGAATTGCTTGAACCTGGGAGGTG 1740
1741 GAGGTTGCAGTGAGCCGAGTTCCTGCCACTGCACTCCAGCCTGGGCGACAGAGTGACACA 1800
1801 CCGTTTCAAATAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1860
1861 CCCTGTTTAAAAGCCTGGAATGGCTCTTCAATTTGCACCTAGAATAAAATCCCAACTCCAT 1920
1921 GCTAGTTTACAAAGCCGTTTATGCTTTGTAACCTATATATACTCCTAACCATGCTCCAGC 1980
1981 AAAACTTGCTTCTTTCTTTTTTTTTGTATACACCGAGCTCATTCCTGCCGAGGGCCTTTG 2040
2041 TGCTAGCTATTGCCTTTGCCTGAAACACTGTACTCCTCATGTGCTAGCATGGGTGAT 2100
2101 TCCTTCTTTAACTCATATATCAGCTTAAAGGTCGCCACCTCAGGGAGGACTTTCCCTGAC 2160
2161 CAATTAATCAAATTTAGCCCTACTGAGTTCCTTACTCTTACACCACCTGTTTTCTTTT 2220
2221 CTTTATTTCTTTTTTTTTTCTGAGACAGAGTCTCGCTCTGTGCGCCAGGCTGGAGAGCAG 2280
2281 TAGCGTATCTTGGCTCACTGCAACCTCCGCTCCTGGGTTCAAGCGATTCTCCTGCCTC 2340
2341 AGCCTCCCGAGTAGCTGGGACTACAGGCGCACACCACCATGTCCGTCTAATTTTTTTGTA 2400
2401 TTTTGTAGTAGAGACTGGGTTAACCATGTTGGCCAGGCTGGTCTCGAATCCTGACCTCA 2460
2461 AGTGATCCACCCGCTCAGCCTCCAACTGCTGGGATTATAGGCATGAGCCACCCAGCC 2520
2521 CGGCCTCCTATTTTTTTTTCATCCCTGGTCCCTAGACTTACACACTGAAATGCCAGTTTCAT 2580
2581 GAGAGCATGAATCTATATGTTCTGTTCACAGCTGTATAAAAGAGCACCTTAGAACAGTG 2640
2641 CCTGGCATGGAGTAGGTGCTGAAGTACTTTTGAGTCAAAGTTACTGGGATGGTGATCATT 2700
2701 GTATGCTGGCTTCTTACTCTTCAAGGCTCAGCTCAAATTTGGCACCACAAAGTGGCCTT 2760
2761 CCCTGACCCCCGCCCCAATTTTCATAGCATTTTTCACTATTTGAAATTATCTTATTTTA 2820
2821 TTCTTAATATGTATGTCTATTTCCCTGTTATACACTCCAAGAAAGCAAGGACCTTGT 2880
2881 CTGTTTGTTCACCTTCTGTCACTTCTAGTGGCTACCACTGTGCCAGACACACAGTGGTACT 2940
2941 CAACAAGTAACAGCAGGGTCACTGAGCTAATAAGTGGTGCAGCCTGGATCCTAACCCAGG 3000
```

3001 TATGTGTAATTTGAAGCCCCTGCTTTCAGGAAGATACCCAGTAGGGACAATGGCCACAC 3060  
3061 AAAAGAGGAACTAAATTTGCATTTTACGGTCTAATTTACCTTTGTATCTATCACACCTAGA 3120  
3121 GAGGCCATAAAGGAACTAAAAAAGCAGGAAATGTCCAGAAATTTCTGAAATGGAAGTTTA 3180  
3181 AAGGAAGGAGTGATGGCTTCTGCCCGTAATCTGAAAAAGCTTCATGAAGGTGGAACAC 3240  
3241 GGCAGGAAAGCTAACAGAGTTAAGGAGCTGACACCGAGGGGCGGCAGGGAGTGCCAGTCT 3300  
3301 TGCTTCAGACGCCCTGGGCTTGAATCTTGGCTCCATCTCTTGTGTGACCTTAGGGTTGTG 3360  
3361 TTATTTTACCTTTCTAGGCCCTAGTTTCTTTCATCAGTAAAACGAATGTACTTAAAGGGC 3420  
3421 TCTTGACAAGACTTAGGAAATACACATAAAGTTGCTTTACACAGTGCTGAGACCATAGTA 3480  
3481 AGTGCTCTATAAATGCAAACTGCCATTAGTATTAGTAGCATTCTGATGAAAGCGTGCTG 3540  
3541 ATCTTACCAGCAGCTATTAGCTGACCACAGTTATGCGGAGCAACAGAAAGCTTCTACCC 3600  
3601 CCCCAGGTTGAAAGGCGGACACCAGAAAGCGGTGGGTCTGGCTCCAGGTGCCACTGTTG 3660  
3661 CGGACTCTCGACCCAGCCTAGATTTTCTCAGACGCTTGAGCCGCAATCTGCCGCTCTAG 3720  
3721 CAACGGGGTCCGCCAGGTCTCGGTGGTGGGGGCTCGCGCTTGCGCAGTGGGCACGGCA 3780  
3781 GCGGCGCGCG**AGGAGGGT****GAGGGGCGGGCCAGCGAGCGGACGCCGGGCGGGCGGGC** 3840  
3841 **GCGGAGAA****GTGCGGCGGAGCGGGCCTGCATTAGCAG**GTAAACCGCGCACTGGGTGCGCCG 3900  
3901 GGGAGGGCGGTGTGGGTTAGTCAGGCTTCAGGCTTGGGTGCGAGGAGGCTGCCTTCGCGCT 3960  
3961 TACCGGCTCTCGGGTCAGCTACCACCCGTGTCAGCGTGAGCGTCTGTGTCTCCTGAATCC 4020  
4021 GGTGCGGACGCTGGGGCGGCCACGCAAGTCCCACCCGGATCGCGTCCAAGGTGGCGG 4080  
4081 CGAAGCCCAGGACAGCCGCTTGGACCCAGCCGATCTCCCGACCTGAAAAACCTTAAC 4140  
4141 CTCAGGGCACCGCCAGGGTCAGAGCCCCAGTTTATGTTTGTTCAGGATTGAGCCCTCTC 4200  
4201 GATTCTCGGAGTTAGATGGAGGGCTCATCTCTGGGATCTGGCTGTTGACAGGGGCTCAG 4260  
4261 TACCTCACAGAGCAGCCTCTTTACATTTTGGAGAGAACCAGGCTTCTCCTCGCCTAACTT 4320  
4321 TTATCGTCTGGTTTTACTCAAACGACTCCAGTCTCTCCACATTTGGAAGACAGCTGCCCT 4380  
4381 TTGCCCTTCCACGCCCTTCTTTTGGAGTTTACTCTCCCTGGTCTTTTCACTGCTTCTCTG 4440  
4441 TCTAAGACTCTTTATTTCTGAAATTTCTTCTGGATCCACTCCCGTTTTTAAGAGATTCT 4500  
4501 CCAAAACTGGATTCTTAGAGTACATTTATAGTAATGCAGTTCTCAAGTCAGACCTGAAT 4560  
4561 CCTAATCTAGAGCTGTGCCAGCTAAGTAGGTGTCTTAGGGAAATCCCTTAACTTCGCTA 4620  
4621 AGCCTCATTTTTTTTTTTCATCGAAAATGTGTACAGTCTTAAAGTACTGTATCTCCAGGTGGT 4680  
4681 TGTGAGGATGAAAGGAGGTGAAGTATGCAAAATCATGAGCATGATACCTGGTACATAGTA 4740  
4741 AGCATCGATTGATGTAGTATTTTTTCCGAACAGATGATTCTCGAGAGAGGCACAGTTG 4800  
4801 GCAGGCGGGTGGGTGGGTGCGGCTGTAATCCAGCACTTTGGAAGGCCAAGGACAGGC 4860  
4861 GGATCACCTGAGGTGAGGGAGTTCAAGACCAGCCTGGCCAACATGGTGAACCCCTGTCTG 4920  
4921 TACTAAAAAATACAAAAATTAGCCTGGCGTGGTGGCGGGAGCCTGTAATCCAGCTACT 4980  
4981 TGAGAGGCTGAGGCAGGAGAATCGTTTGAACCTGGGAGGTGGAGGTTGCAATGAGCCGAG 5040  
5041 ATCGAGCTATTGCACTCCAGCCTGGGCAACAGAGCGGAGACTCCGTCTCAAAAAAAAAAAAA 5100  
5101 AAGAGAGAAGCACAGTTTTCTTAGAATCAAATGCAAAATATAATAATAGCAATAAAATTT 5160  
5161 TCATCAGTCATATAAATGATTTTTTAAATAGTTTTTATATTTTTTAAAAAATGATGTTG 5220  
5221 CTAATGTTGAGGAGTGACACAGTGAAGTAGGTACCAGAGTCAAAGGATGGTACAATTTGAT 5280  
5281 AAAATCATTTTGAAGCAACAAGGATATATTTTACAAGCTTTAGAAATGTCAATTACTTAA 5340  
5341 TTTACAAAAGTACATCTGAAAATTTTACACAGAAATAATTCGTAATGCAGAAAATATG 5400  
5401 AAGAAAGATGGTAGTGATAGCATTATTTTACACTTGTGACAACATTTGAACTTTTGTAA 5460  
5461 AAATTTTTTAAACATACAAAAATAGAATCAGTGAGCCCCCACGTACCCATTACCCAC 5520  
5521 TTCAATAATGATCAGTATTTTGCCTATGTTGTTTTTGTGAAAAATCTGAAACCTAGTAAG 5580  
5581 TTGCTTAAAAATGGGAGAAGTTGAAATAAAATGGGATACATTTATTTAGAATACTATAA 5640  
5641 AGCTTTTGAACAGTACAGAATTGTGTATGTAACATCACAACCTGTGAAAAATCAAAAACT 5700  
5701 AACATCTAATTTATGTTGTAAGAAAAAGACAAAGAGGCCAGGCATAGTGGCTCAAGCCTGT 5760  
5761 AATCCTAGCACTTTGGGAAGCTGAGGCGAGGGATCACTTGGAGTCAAGGATTTGAGATCA 5820  
5821 GCCTGGCCAACATAGTTAAACCCATCTCTACTAAAAATACAAAAATTAGCCAGGTGTGG 5880  
5881 TGGTGGACGCCTGTAGTCCCAGCTACTGGGGAGGCTGAGGCATGAGAAATCGCTTGAACCT 5940  
5941 GGGAGGTGGAGGTTGCAGTGAGCTGAGATCATGCCACTGCACCTAGCCTGGGCAACAGA 6000  
6001 GGGAACTCCATCTCAAAAAAAAAAAAAAAAAAAGAGAAGAAGGAAATACAATAAA 6060  
6061 ATATTAATTTGTTGCCCTTTACTTGTGGGATTGTGGATGGTTCCCTGTTTTTGTATTTTA 6120  
6121 TAAATATTCATAAAAAATCATGTATCATTTGTTACAACAAAAAATCTAAAAAATCACA 6180  
6181 CCTGAACTGAGCAGGATCAAATATGCTAGAGTGGGTTCACTGTTATAGAGAATAAAGATG 6240  
6241 CTGTAACCTTCTCCTGATCTATTCTTGCAGTCTACATTTACATTAGCATTCTTGCCTGCTG 6300  
6301 CCTCACATTAGCACTTCATATTGAATTTATATGCTCAGCTAACATGTCCAGATCTTTAAC 6360  
6361 CAACTGTCTTCTGTGTAAGTGATTTAACTCTGTGAAATGTATTTCCCTTATTTTGAACAG 6420  
6421 TGTTCGCCCTCAGAGTAGAAAAAGGTTTCTGTGATAGGATGACTTTTTTCCCTCATTC 6480  
6481 ATTTCTTCTCCTGGGATCTCACTCAGACTTCATGCATCTACACACTTCTTAAACAAGG 6540  
6541 ATTTCTAACAAATTCAGTAAGATAATTTGTTTTATCACAATTTTTTCTCAGATTATAA 6600  
6601 TGAGATCTGATTAAAAAAATGTTCTGCTCTTGGTGGTGAACCTAGAGCTGCCTTTTAA 6660  
6661 ATGATGATGTAGTACATATGACACAGTTGAGAGAGAGGGGTACCTTCCAAGAAAAAATG 6720  
6721 CATAAATGAGACTTGTTCAAATCAGTTTATGAAGAACTTTGTTTCCTAG**GTATGCAA** 6780  
6781 **GAAGCCTTTTACCCCTGATGTCTTTAGAGATAATATGGATCAGTCCAGAGTTCTCCTCTG** 6840  
6841 **GGTGAAAGCAGAACCCTTTATAGTGGGTGCCTTGAGGTCCCCCTCCATCCAAGTTTAG** 6900

6901	<b>TCTTCACTATCTCAGGAAGATATCCACCTATGTGCAAATCCGGGCCACAGAAGGAGCTTA</b>	6960
6961	<b>CCCGCGCTCTACTGGTCTACATGGAGGCACATCGCTTGTGGGAAGCTGCAGTTGGCCAA</b>	7020
7021	<b>GGACCTGGCGTGGCTTACTTCGAAATATTTGATAGTCTTTCAATGAAGACACCTGAGGA</b>	7080
7081	<b>GCGCTGGAAATGGTCTGAGGTTCTGTCCAATGCATGTCTGAGGAGGAAGTTGAAAAGCA</b>	7140
7141	<b>GAGAAATCAG</b> GTACGGATTGATATGTGTGTATACTTAATTACTTTTTGGAAAGAAACCATT	7200
7201	CTTTTTTAAATTAATAAAATGTTTTTGTCTTATTACAATTTTTAGATAGAAAACAGAGTTTCA	7260
7261	CTTTGTGGCCAGGCTGGTCTTGAACCTCCGGCTTCAAGCAGTCTTCCCTCGCTCGGCCCTC	7320
7321	CCAAAATGTTAGGATTACAGGCATGAGCCACCATGTCTGGCCCAATAAACCATTCTTAAT	7380
7381	TTTTGTATTAGCTGTAAAATTTTAGGTTTTAGGCTAATGGCCATTGTAATCTCAAGTTAT	7440
7441	CATTTGAGTTTTAGTAGTTCAGTCTTGGCAGTATTTTCATGTATTCAATGAATTAGTCCTT	7500
7501	ACTTACCAGACCTTACTCACCTACTCCCAAAGCAGTAATGTTACAGAGACCTGTTATTG	7560
7561	AAAAAGGGAAATATTTTACTGACTTTAAAAAGATTTTGGGGTGAATGAGTGTTATGAACA	7620
7621	ACTGATATGTTACAGTCAGAGCACCTGACTATTTTAAAAATGAACATCTGCTACATGATGG	7680
7681	GATTCAAAGATGATTAGGATAGCATCTCAACCCTCATGGAGCTTATGGACTAGTAGAGGA	7740
7741	AGCTGGTATTAATAAAAGATGCTATTTAATAGTGTGATATGTATAATAAAAGAATACACA	7800
7801	AAGTACAGAAATGAATAAAGAATGATTTATTTTCAGTGATAAGGAAAGACTTCACAGAGA	7860
7861	AAGTATATTAATCAAGTAGGCTGATATAGTAAATAGACAACCCAGTATTTTCAGTTGTA	7920
7921	GAACAGTAACAGTTTTATTTCTCATTACGGCAGCAATCTAACACAGTTGCTCCTGGTTGGA	7980
7981	CAAATCTCTGGGCTGCTGTCTATCCAAGCGATGACCAGGGATCCATGCTTCTATCACTTGG	8040
8041	AGTCACTGTTAATCAGCTTTGTTAAAAGCTGATTTTCATTCATTAAGAACAACAATGAA	8100
8101	ATTCAGTGGCTTACAGGGGAAAAGGTTATTTCTTGCCATTTTACATGTGGGCTTCAT	8160
8161	TTTTGCTATGGCATTGTCCAGGCTCCGTGTTTCCCATTTAGGGACACAGGCTAAAGAA	8220
8221	ATAGCTCTGTGTAGATCATTCCCATTCTTATAACAGAAGGGAAGATCAACAGAATGGAG	8280
8281	GAACCTCATGATGTCTCTTTTTAAATCCTTATTTTTATGTATACAGCTCAATCAGTTTTGG	8340
8341	GATAAGTATATGCTGTGAAACCATCATGACCGTCAAGGACATAGACATACCCATCACCCC	8400
8401	CCATAGGTTCCCCCTCCTCTTTTATGCTTGTATTTTTTGTGTGTATCTGTTAAGAACAC	8460
8461	TTAACATAAGATCTACTGTCTTAGAGAATTTTAAAGTATACAACATGTAATTGTTATCTCT	8520
8521	AGGCGTATGCTGTAGAGTAGATCTTCAAGACTTAATTTTCTTGCCATAACGGAAACTTTG	8580
8581	TAGCCTTTGACCATCAACCTCCCCATTTCCCCCTCCTCCTAGCCCTTGTCACTCACCCCT	8640
8641	CTGTCTGTTTTTATGAGTTTGACTATTTTAGATTCCACATATAAGTGAGATTACATGCT	8700
8701	TTGTCTTTCTGTGTCTGGCGTGTTTTTACTTAGCATAATGTCTCCAGGTCCATCCATGTT	8760
8761	GTTGAAAATGGCAGGATTTCTCTCTCTTTCTGTTTTTAAAGAGACAGGATCTCACTCTGTT	8820
8821	GCCCAAGCTGGAGTACAGGGATGTAATCATAGCTCACTATAGCCTGGAATTTCTGGGCAC	8880
8881	AAGCAATCCTTCCACCTCAGCCTCCTGAGTGGTTAGGACTACAGGTGTGTGCTATCATGC	8940
8941	CTGGCAAATTTTTGTATTTTTTTCATAGAGGTCAAGGATCTCACTGTGTGCCCAGGCTGG	9000
9001	TCTTGAACCTCTGGCCTCAAGTGATCCTCCCACCTCAGCCTCCCAAAGTGTGGGATTAC	9060
9061	AGCCATAACCACCTGTACCCAACATGATTTCTTTCTTTTTTAAAGCCAAATGGTACTCCA	9120
9121	TTATATATGCATGCCAAGTTTTCTTTATCTAGTCATCCATCTGTGGACATTTAGGCTGCT	9180
9181	TCCATATCTTGGCCATTATCAATAATGCTGCAATGAACATGGGAGTGCAGGATTTGTCAG	9240
9241	GATCCTGATTTCAATTCCTTTGGATAAATACCAGAAGTGGGATTGCTAGATCGTATGGT	9300
9301	AGTTCTCTTTTTTATTTTTTGGAGAAACCACTATACTATTTATTTATTTTTTTTTGAGAC	9360
9361	GGAGTCTCGCTCTGTGCGCCAGGCTGGAGCGCGGTGGCGGATCTCAGCTCATTGCAAGC	9420
9421	TCCACCTCCCAGGTTACACCATTCTCCTGCCTCAGCCTCCTGAGTAGCTGGGACTACAG	9480
9481	AGCCCAACCACCTGCCCCGCTAGTTTTGTGTATTTTTTAAATAGAGATGGGTTACCAT	9540
9541	GTTAGCCAGGATGGTCTCAATCTCCTGACCTCATGATCCGCTGCCTCGGCTCCCAAAG	9600
9601	TGCTGGGATTACAGGCGTGAGCCACCGCGCCCGGCTATTTTTTATAATGGCTGTACCA	9660
9661	GTTTACATTCTCACTGACAGTGTACCAGGTTCCCTTTCTTACATCCTTACTAATAACT	9720
9721	TGTTAATCTTTTGTAAACAAGTGTGAGGTGATATCTCATTTGTGGTTTTGCTTTATATTT	9780
9781	TCTAATGAGTAGTGTGTTGAGCACCTTTTCGTATACCTGTTGGCCATTTGTATGTCTTC	9840
9841	ATTGGAGAAATGTCATTCAGGTCCATAGTCCATTTTTTAAATTAGGTTATTTGGTTTTTC	9900
9901	TTGCTATTGAGCTGCAGGAGTTCATTTATATTTTTTCGATATTAACCCCTTACCAGATACA	9960
9961	TGGTTTACCTTTTCTCTGTTGATTGTTTCCTTAGTTGTGCAGAAGCTTTTTAGTTTGAT	10020
10021	ATAATCAACTTATTTTTGCATTGGTTGCTGTGCTTTTGGTGTATATCCAAACAATCA	10080
10081	TTACCAAGACCAATGTCAAGGAGCTTTTTCCTTATTTTTCTCTAGGAGTTTTACCATTT	10140
10141	ACCCCTTACATTTAAGTCTTTCTCCATTTTGAGTTGATTTGTGTATGGTATAAGCTATGG	10200
10201	TCCCAATTTCACTTTTGCATTTGTAGATCCAGTTTTCTCAGCATCATTTATTGAAGAG	10260
10261	ACTATTTCTGCGTTTTTGTATTCTTGACACCTTGTGGAAAATTAGTTGACCATATATGTGT	10320
10321	GAGTTTGTCTTCTGGACTCATGATGCCTCTTAAAACCTTTTGTCTCAGATTTGACATAAACTA	10380
10381	AGACTACTTTTATTCATTGGCTAAGGCATTTCCCATGAAGATGGGAAGGTAAACCAC	10440
10441	CTTTCAGGAGATAAGGAAAATCCCATGGCCATGGGTAGGGCTGTATAAGCCTCTTAGAGA	10500
10501	ATTTGAAAGTGAGCAATAGTAAATAATAATAACAATTTGCCACAGCCATGCCACTGGGGC	10560
10561	CTCAGAGTCCTTTTGCTTTTCATTTCTTTTGGAGAGTTTTATTTAGGCTTGAAAGTGATA	10620
10621	AAGATAAATTCATTCACATTCAGTTGCTGTAATTCATTTATGCCATATCTAATCCAAAG	10680
10681	GAAGATGGGAAATGCCATCTAATTACATACACAGGAAGAAAGAAAAAGTAGTAAGCATC	10740
10741	TAGCAGTCTCTGCCTTATAGGGTCATATCTAAACAACGTTTTGGAAATTGAAAGTTTTAT	10800

10801 CAAGCAATGAAGAGGGAAAGGGCAGTACATTTATTCATTAATCAAATATTTATTATCTAC 10860  
10861 TATGTAATCGGTATAGTAGTGGATGCTGGGCATATAGAGATGAATAAACCTTCAGGAAG 10920  
10921 TTCCAGATTTTTTGGAGTGCCTTTTATGTACCAGATGCTTTGCTAGTTACTAGGAGTATAA 10980  
10981 AAATGGAAAAGGCAAGGATCCTATAGTCTCCATGAAGAAATAATGATAATGTAACAAATA 11040  
11041 ATTGTTTAAAATGGAAGTATATATAAAAATCAGGGAATTACACTTAGTGAAAAAAGCCAAT 11100  
11101 CCCAAGAGGTTACATACTATATGATTCTATTTACCTAATGTTCTTGAAATAATAAATTA 11160  
11161 TAGAGATGGAGAACAAGTAAGTGGTTGCTAAGGGCCTAAGAAGGGGTGAGTGGGACAGAG 11220  
11221 AAGTGCATATGGCTGTAGAAGGGCAACATGAGAGATTGTTCTGTTGATGGAAATGTTCTC 11280  
11281 GTATCTTGACTGTATCGGTGTCAATATCCTGATTGTGATATTGTCCTATAGTTTGGCAAG 11340  
11341 GCGGGTAAAGGGTACATGGGATCTTTCTGTGTGAATTCTCACAACATATATCTCAATCTAC 11400  
11401 AGTTATCTCAGAAGTTCAATTAAGAAAAGTAAAAATTTATAAAAATATACAAATTCATGGGG 11460  
11461 CTTCACATTGTTAGTTGATTAAATAGGCTGGGTGATCCTGAGAATCTGTACTTTTAAGT 11520  
11521 TTCTCAGATGATAGTGTTTTTAGCCAGGTTGGGAATCAGGGTGAATATAGCTGTAGACCC 11580  
11581 AGATGTATTACTTTACTTTTTGCTCTCCTGTATGGGGGTGGGTAGATGAGGGTAATCA 11640  
11641 GGGGCTTCAAGTCGTGAGCTTCAACCACACCCTTCTTGAATAAGAAAATGACTATACATC 11700  
11701 TATCAGACCAACAGTGTTTTTGTCTTTAAGTGAACACCCTTTAAGAATTTGCTGAAAA 11760  
11761 CCTGAGCCTACTCCCTTGAAAAATGCACATATACCAAAAAATTACATACAATTTCCAGGT 11820  
11821 ATTGCACTGGTATCCTAAACTTTCTTGACAGGTTTGTGAGTCTTAAGTTAGACCTCTGTC 11880  
11881 TTAAGTGAATTTCTTCTTCTTACTTCCATCTACTTATTTATTCAAATTAAGGTTG 11940  
11941 GTTTTTTTTAAATTTAAGTTTAAACTTCTGAAATAATGTACTTAAAGTATGTACTACAT 12000  
12001 CCTTTCATCAGTTTTGAGAGATAATTTCTATAGCCATAGATGTTATACAGCTTTCTGTGA 12060  
12061 AGGTAAATTAATCCACTTTAATTAGTTTTAGGAACTCTTGGCCAGGCACAGGGGTTCCAG 12120  
12121 CCTGTAATCCCAGCACTTTGGGAAGCAAGCGGGCAGATCACCTGAGGTCAGGAGTTCA 12180  
12181 AGCCAGCCTGGCCAACATGGTGAAACCCTATCTCTACTAAAAATACAAAAAATTAGTT 12240  
12241 GGGCGTGGTGGCACATGCTTGTAATCCCAGCTACTCAGGAGGCAGAGGCAGGAGAATCTC 12300  
12301 GCTTGAACCAGGAAGCGGAGGTTGCACTGAGCTGAGATCGCATCATTGCACTCCAGCCT 12360  
12361 GGGCGACAAGAGTGAACCTCTGTCTCAAAAAAAGAAAAAGAACTTTACATCTGGG 12420  
12421 TACCAGAGACATACATATCAAAACTGGCAATGTTTCTAAGTTGGTTTTTTTAGGATCAGC 12480  
12481 CAACTTGTGAATTTACTATGTCTACTTAGTTGTTAGACTAGTGGTTGATGTTTTAAACG 12540  
12541 TCTATTCATCTTTTTATTTCTTCTTGGACAG**CTTTCAGTGGACACGCTACAGTTTCTGCT** 12600  
12601 **CTTCTTATACATTCAACAGTTGAACAAGGCTCTCCCTAAGGACATCTTTGATTGGCGAAGA** 12660  
12661 **GTGGCCAGTCCCAGAAACAATCTCAGTCTCCTGACCTGACTGAAAAATCTAATTGTCA** 12720  
12721 **TAATAAG**GTAATCCTTATATCTTGTATCTCATTCTCTAAAAGGGACAATTTGTGTGGGAGA 12780  
12781 TGGTTCACTTGCATTAATAAATATGATGAGATTTCTTACATAAGCCAGTGATTTGATGGC 12840

Os códons de iniciação, presentes no exão 1 e 2, encontram-se sublinhados.



## 2. pIC111 (6xHis-PreScission-GFP) – vetor construído no vetor pcDNA3.1+ - vetor de expressão em mamífero

(Cheeseman and Desai 2005)

```

GAGCTCTCTGGCTACTAGAGAACCCACTGCTTACTGGCTTATCGAATTAATACGACTCACTATAGGGAGACCCAGCT
      ▲
      HindIII
NheI  PmeI  AfIII  KpnI  BamHI  SpeI  BstXI  EcoRI  PstI  EcoRV  BstXI
GGCTAGCGTTTAACTTAAGCTTGGTACCGAGCTCGGATCCACTAGTCCAGTGTGGTGGAAATCTGCAGATATCCAGCAC
  ▶ A F K L K L G T E L G S T S P V W W N S A D I Q H
      NotI  SacII
AGTGGCGGCGCGggcagcagccatcaatcaatcaatcaatcaacagcagcggcctggaagttctgttccagggggccctggaagt
▶ S G G R G S S S H H H H H S S G L E V L F Q G P L E V
      BspEI
ctgttccagggggccctCCGGAGGGATGGTGAGCAAGGGCGAGGAGCTGTTCCACGGGGTGGTGCCATCCTGGTCGAGC
▶ L F Q G P S G G M V S K G E E L F T G V V P I L V E
TGGACGGCGACGTAACCGGCCACAGTTCAGCGTGTCCGGGAGGGCGAGGGCGATGCCACCTACGGCAGCTGACCCCTG
▶ L D G D V N G H K F S V S G E G E G D A T Y G K L T L
AAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCTGGCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCCT
▶ K F I C T T G K L P V P W P T L V T T L T Y G V Q C F
CAGCGCTACCCGACCACATGAAGCAGCAGCTTCTTCAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCA
▶ S R Y P D H M K Q H D F F K S A M P E G Y V Q E R T
TCTTCTCAGGACGACGGCACTACAGACCCCGCCGAGGTGAGTTTCAGGGCGACACCCTGGTGAACCGCATCGAG
▶ I F F K D D G N Y K T R A E V K F E G D T L V N R I E
CTGACGGGACTTCAAGGAGGACGGCAGCATCCTGGGACACAGCTGGAGTACAACTACACAGCCACACGCTA
▶ L K G I D F K E D G N I L G H K L E Y N Y N S H N V Y
TATCATGGCCGACAGCAGAGAGACGGCATCAGGTGAACTTCAGATCCGCCACACATCGAGGACGGCAGCGTGCAGC
▶ I M A D K Q K N G I K V N F K I R H N I E D G S V Q
TCGCGACCACTACAGCAGAACCCCATCGGCGACGGCCCGTGTGCTGCTGCCCGACACCACTACCTGAGCACCCAG
▶ L A D H Y Q Q N T P I G D G P V L L P D N H Y L S T Q
TCCGCCCTGAGCAAGACCCCAACGAGAGCGCATCATGGTCTGCTGGAGTTCGTGACCGCCCGGGATCACTCT
▶ S A L S K D P N E K R D H M V L L E F V T A A G I T L
      XbaI  ApaI  PmeI
CGGCATGGACGAGCTGTACAGTAACTAGAGGGCCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCCCTTCTAGTTGCC
▶ G M D E L Y K ●
AGCCATCTGTTGTTTGGCCCTCCCGTGCCTTCCTTGACCCTGGAGGTGCCACTCCCCTGTCTTTCTAATAAAT
  
```

Laranja – 6xHis

Verde – PreScission “cleavage site” (2x)

Azul – eGFP

Seta – local de iniciação da transcrição