UNICAMP

UNIVERSIDADE ESTADUAL DE CAMPINAS

Instituto de Biologia

PEDRO FARIA ZENI

ESTUDO DA ATIVAÇÃO DO LINFÓCITO: PAPEL BIOLÓGICO DA RETENÇÃO INTRÔNICA COMO SPLICING ALTERNATIVO

Campinas

2018

PEDRO FARIA ZENI

ESTUDO DA ATIVAÇÃO DO LINFÓCITO: PAPEL BIOLÓGICO DA RETENÇÃO INTRÔNICA COMO SPLICING ALTERNATIVO

Dissertação apresentada ao Instituto de Biologia da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Mestre em Biologia Funcional e Molecular, na Área de Fisiologia.

Supervisor/Orientador: PEDRO OTAVIO DE CAMPOS LIMA

ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL DA DISSERTAÇÃO DEFENDIDA PELO ALUNO PEDRO FARIA ZENI, E ORIENTADO PELO PROF. DR. PEDRO OTAVIO DE CAMPOS LIMA

Campinas

2018

Agência(s) de fomento e nº(s) de processo(s): FAPESP, 2014/04412-4

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Biologia
Mara Janaina de Oliveira - CRB 8/6972

Zeni, Pedro Faria, 1993-
Z43e Estudo da ativação do linfócito : papel biológico da retenção intrônica como *splicing* alternativo / Pedro Faria Zeni. – Campinas, SP : [s.n.], 2018.

Orientador: Pedro Otavio de Campos Lima.
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Biologia.

1. Íntrons. 2. Processamento alternativo. 3. Transcrição genética. I. de Campos-Lima, Pedro Otavio. II. Universidade Estadual de Campinas. Instituto de Biologia. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Lymphocyte activation : biological role of intron retention as alternative splicing
Palavras-chave em inglês:
Introns
Alternative splicing
Transcription, Genetic
Área de concentração: Fisiologia
Titulação: Mestre em Biologia Funcional e Molecular
Banca examinadora:
Pedro Otavio de Campos Lima [Orientador]
José Andrés Yunes
Leonardo dos Reis Silveira
Data de defesa: 30-05-2018
Programa de Pós-Graduação: Biologia Funcional e Molecular

Campinas, 30 de maio de 2018

COMISSÃO EXAMINADORA

Prof. Dr. Pedro Otavio de Campos Lima (Orientador)

Prof. Dr. José Andres Yunes

Prof. Dr. Leonardo dos Reis Silveira

Os membros da Comissão Examinadora acima assinaram a Ata de Defesa, que se encontra no processo de vida acadêmica do aluno.

DEDICATÓRIA

Dedico essa dissertação à minha mãe que sempre esteve ao meu lado me apoiando em todas as minhas escolhas e trajetórias.

AGRADECIMENTOS

Gostaria de agradecer à minha mãe, **Rosana Mara de Freitas Faria**, por todo esse tempo de parceria. Me incentivando a continuar minha carreira acadêmica e me dando todo suporte que eu precisava.

Ao meu avô, **Antônio Faria**, que não se encontra mais entre nós, mas certamente estaria muito orgulhoso de seu neto estar terminando seu mestrado. Um homem de respeito, que me ensinou a importância da família e como levar a vida de maneira alegre. À minha avó, **Edna Faria**, que sempre me incentivou a correr atrás dos estudos e sempre esteve torcendo e orando por mim.

À minha família, aos meus tios, primos e parentes. Em especial ao meu tio **Alberto Riecken** que sempre acreditou no meu potencial.

Aos meus professores, principalmente ao MSc **Décio Fragata da Silva** que me apoiou na minha vinda à UNICAMP. À minha ex-orientadora Prof. Drª **Maria Lucila Hernández Macedo,** que me deu a oportunidade de conhecer o Centro Infantil Boldrini e me deu o *background* necessário para começar minha carreira em biologia molecular. Agradeço também aos professores da pós-graduação da Biologia Funcional e Molecular que me ensinaram bastante.

Agradeço ao meu orientador, Dr. **Pedro Otavio de Campos Lima**, que diante de toda adversidade lutou para que tivéssemos uma boa educação durante a pós-graduação. Foi um orientador capaz de mudar minha perspectiva e visão de mundo na ciência, me fazendo crescer bastante em pouco tempo.

Aos meus amigos e colegas de laboratório: **Leonardo Pissinato**, **Guilherme Navarro**, MSc. **Gabriel Centoducatte**, MSc. **Carolina Dias**, **Caroline Lopes**, Drª. **Mônica Ganazza**, **Natália Paiva**, **Juliana Ronchi**, **Leonardo Ártico**, **Patrícia Jotta**, **Amanda Biscaro**, Drª. **Ana Seidinger**, MSc. **Livia Campos**, MSc. **Natacha Migita**, MSc. **Nathália Cury**, **Sabrina Morelli**, **Victor Sande**, Drª **Bruna Lima**, MSc. **Diego Maldonado**, **Mayara Euzébio**, **Thais Guimarães**.

Ao Dr. **Rafael Canevarolo**, que foi um exemplo de pessoa, caráter e pesquisador. Um amigo que me ajudou e orientou em diversas ocasiões no trabalho e na vida. À Drª.**Priscilla Zenatti** que foi bastante receptiva e também é um exemplo de pesquisadora.

Ao pessoal da casa jabuticabeira e agregados, que tornou minha experiência em campinas muito divertida. Em particular, ao **João Piñheiro** que virou um grande amigo.

À república morango do nordeste: à **Gisele Silvestre** que me acolheu, ao **Ridley Gadelha** que foi muito amigo e me mostrou muitas coisas, ao **Gilson Boticeli** e ao **Geovani Rocha**. Tenho uma imensa gratidão ao **Rômulo Vieira,** que tanto me ajudou e ensinou nessa jornada.

Um agradecimento especial aos meus amigos do coração: **Raphael Ferreira**, **Diogo Nishikawa**, **João Nunes**, **Mariene Amorim**, **Katharina Oliveira**. Espero levar para vida toda esses amigos, que estiveram dividindo tristezas e felicidades ao meu lado − vibrando com cada passo e conquista que eu dava e torcendo pelo meu sucesso e felicidade. Sem vocês, certamente a vida seria muito mais difícil.

À **Jullyana Matos** foi uma pessoa crucial para eu chegar aonde cheguei e me tornar o que me tornei, embora inúmeros obstáculos da vida tenham se oposto a nós. Ela sempre torceu pela minha felicidade, e acima de tudo, pelo meu crescimento profissional. Também me deu apoio emocional e calma para lidar com todos os problemas da vida.

Agradeço ao Centro Infantil Boldrini por proporcionar um ambiente de aprendizado e de trabalho.

E por fim agradeço à CAPES e a Universidade Estadual de Campinas por terem me proporcionado um auxílio financeiro e uma educação de qualidade.

Obrigado a todos.

RESUMO

A transcrição é a etapa necessária para a expressão gênica que envolve a produção de uma fita de RNA usando o DNA como molde. Estima-se que 95% dos genes humanos sejam afetados por um processo de *splicing* alternativo. Este processo é uma das ferramentas de que a célula dispõe para a produção de diferentes tipos de transcritos derivados de um mesmo gene. O linfócito T é um modelo bastante útil para o estudo do processamento transcricional, de que nos servimos para analisar um tipo emergente de *splicing* alternativo ainda pouco caracterizado: a retenção intrônica. Criamos um programa algorítimo para identificar os transcritos com retenção intrônica, usando dados de sequenciamento de RNA, obtidos de linfócitos T CD4+ em repouso e ativados. A retenção intrônica mostrou-se ser o tipo de *splicing* alternativo mais frequente dos eventos transcricionais, com pouca divergência em termos de classificação ontológica das isoformas diferencialmente expressas nos fenótipos de ativação celular estudados. Entretanto, provou-se que esse tipo de *splicing* não é um 'ruído' transcricional generalizado. Ademais, constatou-se a existência de diversidade desses transcritos, com oscilação quanto à carga intrônica retida durante a ativação do linfócito. Identificamos, todavia, isoformas com retenção intrônica diretamente ligadas ao fenótipo de repouso e de ativação celular, nos quais a célula parece regular seletivamente os transcritos, de forma similar, mesmo em diferentes indivíduos.

ABSTRACT

Transcription is a necessary step for gene expression involving the production of RNA molecules from DNA templates. It is estimated that 95% of the human genes are affected by some sort of alternative splicing. Alternative splicing is a tool that the cell uses to produce different types of transcripts from the same gene. T lymphocytes are particularly suitable for studies on transcriptional processing, and we have used this model to analyze a new emergent type of alternative splicing: intron retention. We have created a program to identify intron retention transcripts using data from RNA sequencing obtained from resting and activated CD4+ cells. Intron retention was shown to be the most common type of alternative splicing, with small differences in ontological classification of the transcripts differentially expressed in the two activation phenotypes. We have shown that this kind of alternative splicing is not transcriptional 'noise' in its entirety. In addition, there is transcriptional diversity as regards the level of intron retention in the two states of activation of the lymphocyte. However, we have identified specific isoforms with intron retention directly connected to the resting or activated cell phenotypes, suggesting the existence of selective regulation of these transcripts, whose pattern is conserved even in different individuals.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

5’CAP – guanosida metilada na extremidade 5’

3’ss – sítio splice 3’

5’ss – sítio splice 5’

A3 – sítio de splice alternativo 5’

A5 – sítio de splice alternativo 3’

AP-1 – proteína ativadora 1

ATSS – sítio alternativo de iniciação transcricional

ATTS – sítio alternativo de terminação transcricional

CA – Célula ativada

cDNA – DNA complementar

CR – Célula em repouso

CTB – domínio carboxi-terminal

CTP – código de terminação prematuro

DNA – ácido desoxirribonucleico

ES – éxon *skipping*

ESE – *enhancers* de *splicing* exônico

GRE – Elemento responsivo a glicocorticoide

H3K4m3 – trimetilação na H3K4

IFN-γ – interferon-γ

IL-2 – interleucina 2

MES – múltiplos éxon skipping

MHC – complexo de histocompatibilidade

miRNA – microRNA

MPT – modificações pós-traducionais

mRNA – RNA mensageiro

NCBI – Centro Nacional de Informação Biotecnológica dos Estados Unidos

NFAT – fator nuclear ativador de células T

NMD – *nonsense mediated decay*

ORF – *open reading frame*

pb – pares de base

PBMC – células mononucleares do sangue periférico

PCR – reação em cadeia da polimerase

PIC – pré-iniciação

polyA – calda de adenina na extremidade 3'

RI – retenção intrônica

RISC – complexo de indução de silenciamento de RNA

RPKM – *reads per kilobase million*

RNA – ácido ribonucleico

RNAi – RNA interferante

rRNA – RNA ribossomal

RNApol – RNA polimerase

RNA-seq – sequenciamento de RNA

SF1 – fator de *splicing* 1

siRNA – pequeno RNA interferente

snRNP – ribonucleoproteínas

SA – *splicing* alternativo

SR – proteína serina-arginina

SRA – arquivos de sequências de reads

T_{CM} – célula T de memória central

TCR – receptor de células T

TEM – célula T de memória efetora

T_{SCM} – célula T tronco de memória central

U2AF – fator auxiliar da U2

UTR – região não traduzida

SUMÁRIO

1 – INTRODUÇÃO

As células são unidades funcionais de um organismo que precisam ser sensitivas e estarem prontas para responder aos estímulos do ambiente, tais como temperatura e disponibilidade de nutrientes, modelando sua capacidade de adaptação e sobrevivência (SAUNDERS e colab., 2006). Tais estímulos são transmitidos por moléculas sinalizadoras que determinam quando a célula precisa se dividir, migrar ou até mesmo morrer. Essa adaptação requer a produção de proteínas, o que frequentemente depende da capacidade do estímulo induzir a transcrição gênica (WEAKE e WORKMAN, 2010).

Uma vez que os genes são ativados, inicia-se o fluxo de transmissão da informação genética, no qual o DNA (ácido desoxirribonucleico) funciona como um molde para a produção de RNA (ácido ribonucleico) e este último constitui o diagrama de construção das proteínas durante o processo de tradução. Este fluxo de informação se constitui no princípio do dogma central da biologia molecular (SAUNDERS e colab., 2006) (Figura 01), sendo a etapa de transcrição no qual é gerada a molécula de RNA a partir da fita molde de DNA uma fase importante para resposta celular.

Figura 01. Dogma central da biologia molecular. A partir de uma fita molde de DNA, a informação genética é transcrevida em uma molécula de RNA no núcleo celular. RNAs codificantes deverão deixar o núcleo para ir ao citoplasma aonde levarão informações necessárias para sua tradução em proteínas.

A transcrição é feita pelas RNA polimerases (RNApol) I, II e III, que produzem RNAs ribossomais (rRNA), transcritos codificadores de proteínas e pequenos RNAs não codificadores, respectivamente (SAINSBURY e colab., 2015). O processo se desenvolve em três fases: iniciação, elongação e terminação, tendo como produto o pré-mRNA, o qual deve ainda ser processado em um RNA maduro ou RNA mensageiro (mRNA), para funcionar como "receita genética" na tradução de proteínas. Inúmeros fatores influenciam a habilidade da RNA polimerase de ter acesso e se ligar aos genes, bem como de transcrevê-los em resposta a algum sinal (SAUNDERS e colab., 2006). Estas enzimas são formadas por múltiplas subunidades, sendo Rpb1 a maior delas e possuidora de sequências repetitivas (YSPTSPS) (HSIN e MANLEY, 2012) no domínio carboxi-terminal (CTB). Modificações pós-traducionais deste

heptapeptídeo, tais como fosforilação, modulam a capacidade de interação com outras proteínas. Embora CTB não tenha função catalítica, funciona como plataforma de acoplamento (*docking*) para diferentes tipos de quinases (Ser2, CDK7, TFIIH, CDK8) e fosfatases, que influenciam diretamente o papel da RNApol II (ROEDER, 2005; SAINSBURY e colab., 2015; WEAKE e WORKMAN, 2010).

Durante a fase inicial da indução gênica, a RNApol II liga-se a sequências específicas do DNA, interagindo com fatores gerais de transcrição (TFIIB, TFIID, TFIIE, TFIIF, TFIIH) e componentes da maquinaria transcricional. Este processo resulta na formação do complexo de pré-iniciação (PIC) no promotor do gene ativado (PTASHNE e GANN, 1997; ROEDER, 2005). A subunidade DNA-helicase da TFIIH hidrolisa ATP para remodelar o PIC. No começo do processo de remodelamento, aproximadamente 11 a 15 pares de bases (pb) do DNA situados próximo ao sítio de iniciação são desenrolados, permitindo a ligação da RNApol II ao molde de DNA (CRAMER, 2004; WANG, W e colab., 1992) e o rápido início da síntese de RNA.

A RNApol II promove a elongação completa do transcrito com alta possessividade, sem dissociação do molde de DNA ou liberação do RNA nascente. Entretanto, para que isso ocorra, é necessário que a síntese de aproximadamente 15 nucleotídeos já tenham ocorrido. Até então, a síntese é altamente instável e dependente de fatores de transcrição auxiliares, com alta propensão de parada proximal (*promoter-proximal arrest*) ou aborto da transcrição. Esta fase de instabilidade funcional da RNApol II ou *promoter escape* deve ser superada para que seja formado o complexo de elongação transcricional (CET) (DVIR, 2002).

Assim, a molécula nascente de RNA recebe a adição de uma guanosida metilada, na extremidade 5’ (5’CAP) e, quando completada a transcrição, ganha uma cauda de adenina na extremidade 3’ (calda polyA) (ZHANG, Zhiqiang e colab., 2004). Essas modificações são importantes para a etapa de terminação, pois contribuem para o processo de dissociação do complexo transcricional da RNApol II (LUO e BENTLEY, 2004; ZHANG, Zhiqiang e colab., 2004).

Em eucariotos, a RNApol II tem que lidar com a transcrição do molde de DNA no contexto de uma série de complexos proteicos a ele ligados, que constituem a cromatina e regulam diretamente a transcrição. A cromatina tem como principal componente o nucleossoma, que é constituído por 146 pb de DNA, enrolados em um octâmero de histonas. O octâmero é formado por duas cópias de histonas canônicas, conhecidas como H3, H4, H2A e

H2B (KORNBERG, 1974). É conhecido que regiões descondensadas da cromatina têm genes mais transcritos, com menos nucleossomas nas regiões intergênicas. Os promotores de genes ativamente transcritos também tendem a apresentar densidades mais baixas de histonas e, consequentemente, terem diminuição de nucleossomas quando comparados às regiões a jusante (POKHOLOK e colab., 2005).

Diversos fatores catalíticos regulam a cromatina através de alterações químicas das histonas. Essas modificações pós-traducionais (MPT) afetam a estabilidade do nucleossoma, alterando as interações físico-químicas do mesmo ou dos vizinhos e resultando em uma cromatina mais aberta ou fechada. De maneira geral, os nucleossomas representam uma barreira transcricional, que deve ser movida ou desfeita, para que a polimerase possa percorrer o DNA (CLOSE e colab., 2006; KRISTJUHAN e colab., 2002). A acetilação das histonas H3 e H4 é tipicamente associada com genes ativos (CHIH e colab., 2005; POKHOLOK e colab., 2005), sendo importante para a transcrição na etapa de elongação (CLOSE e colab., 2006; KRISTJUHAN e colab., 2002). A trimetilação na H3K4 (H3K4me3) é normalmente concentrada na extremidade 5’ do gene e tem forte correlação com ativação transcricional. Igualmente, a metilação de H3K36 também pode ser evidenciada em vários genes codificadores (KROGAN e colab., 2003; POKHOLOK e colab., 2005; RAO e colab., 2005; XIAO e colab., 2003).

Essas modificações de histonas citadas são apenas alguns exemplos de estratégia de que a célula dispõe, para regular a transcrição e expressão dos genes antes de processar o RNA em sua maturação. Outros mecanismos regulatórios pós-transcricionais incluem os pequenos RNAs interferentes (siRNA) e os microRNAs (microRNA), que regulam a expressão gênica após a transcrição, via complexo de indução de silenciamento de RNA (RISC) (BARTEL, 2004; FILIPOWICZ, 2005). Contudo, a regulação pode ocorrer através do processamento por *splicing* do precursor do RNA mensageiro.

Splicing é um processo que remove os íntrons do pré-mRNA e junta os éxons para produção de mRNA maduros (CHOW e colab., 1977; STERNER e colab., 1996). Tradicionalmente, os éxons são definidos como as sequências codificadoras dos genes que são representadas ininterruptamente em um mRNA e os íntrons são sequências que intervém entre os éxons no pré-mRNA e no molde de DNA. Cada íntron tem pelo menos três elementos cis: um sítio de *splice* 5’ (5’ss), um sítio de ramificação e um sítio de *splice* 3’ (3’ss). Este último é chamado de trato de polipirimidina e contém 15-40 pirimidinas (usualmente uracilas) (TARN e STEITZ, 1996; TURUNEN e colab., 2013). Uma vez que o ser humano tem 20.000-25.000

genes codificadores de proteínas e apenas 700 que não contém íntron (BUSCH e HERTEL, 2013; LANDER e colab., 2001; LOUHICHI e colab., 2011), o *splicing* representa uma etapa fundamental para o bom funcionamento celular.

O processo de *splicing* envolve diferentes etapas de alto consumo de ATP (PAN e colab., 2008), que são catalisadas pelo spliceossomo – uma maquinaria de pequenas partículas de ribonucleoproteínas (snRNP), conhecidas como U1, U2, U3, U4, U5 e U6 (OHSHIMA e colab., 1981). Primeiramente, U1 snRNP reconhece 5'ss via pareamento de bases com o pré-mRNA, formando o complexo inicial (complexo E). O fator de *splicing* 1 (SF1) liga-se ao sítio de ramificação; o fator auxiliar da U2 (U2AF) liga-se ao trato de polipirimidina e ao 3'ss; as proteínas serina-arginina (SR) ligam-se a *enhancers* de *splicing* exônico (ESE) (SHEN e colab., 2004). Depois dessas interações, o complexo E está destinado à via de *splicing*, com subsequente ação catalisada pelas helicases Pp5 e Sub2. O complexo E é convertido em complexo A pela substituição do SF1 pela U2 no sítio de ramificação (CHEN e MANLEY, 2009). Após a montagem do complexo A, U5 e U6 snRNPs são recrutadas para formação do complexo B, onde também há reconhecimento das regiões intrônicas (RAGHUNATHAN e GUTHRIE, 1998). Várias helicases de RNA participam do complexo B, resultando em um rearranjo das formas U2-U6 snRNPs, liberando U4 e U1 e iniciando vários processos catalíticos que amadurecem o mRNA no complexo C (ILAGAN e colab., 2013; SCHWER e GROSS, 1998). A formação dos complexos E, A e B são possíveis pontos de regulação do *splicing* alternativo. As reações do processo de *splicing* estão representadas de maneira simplificada na figura 02.

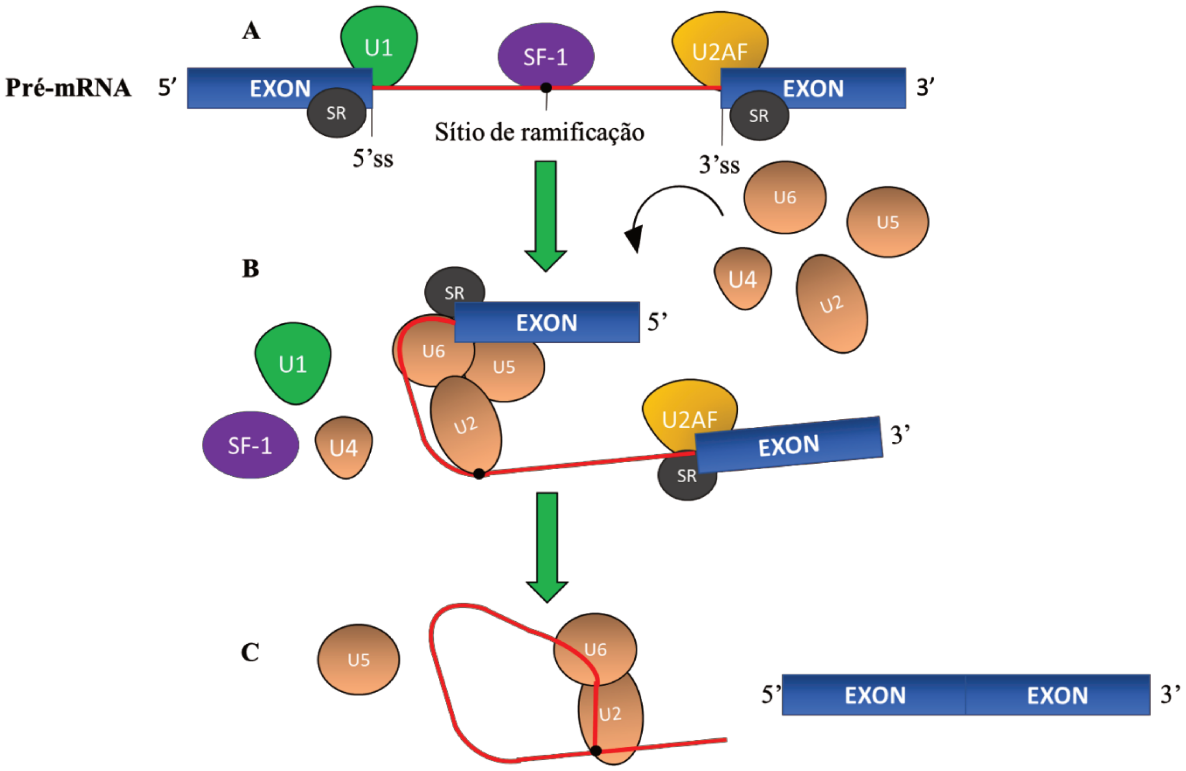


Figura 02. Resumo das reações realizadas pelas proteínas da maquinaria de *splicing*. **A**: A formação do complexo E inicia com a U1 ligando ao 5'ss, SF-1 ao sítio de ramificação e U2AF ao trato de polipirimidina. As proteínas SR reconhecem os *enhancers* de *splicing* exônico. **B**: A formação do complexo A começa com U1 e U2. U4-U6 são recrutadas para formação do complexo B. A interação entre as snRNPs liberam U1, SF-1 e U4, seguido de mudanças conformacionais que culminam na formação do complexo C. **C**: U5 é liberada e logo após os cassetes de éxons também o são.

O splicing alternativo (SA) permite à célula regular a expressão gênica com a produção de mais de uma espécie de mRNA a partir de um único gene. Esses mRNAs podem diferir em sua região não traduzida (UTRs) e nas sequências codificadoras em consequência de mecanismos de éxon *skipping* (ES), múltiplos éxon *skipping* (MES), utilização de sítio doador (A5) ou aceptor alternativo (A3) e retenção intrônica (RI) (Figura 03) (BLACK, 2003; MATLIN e colab., 2005; PAN e colab., 2008; SAMMETH e colab., 2008). Essa diferença no processamento entre as espécies do mesmo pré-mRNA têm impacto direto em sua estabilidade, localização ou tradução. O *splicing* alternativo dos pré-mRNA cria diferentes janelas de leitura e/ou combinações de éxons, que são traduzidas em diferentes variantes proteicas potencialmente possuidoras de diversa funcionalidade e localização. Cerca de 95% dos genes humanos sofrem *splicing* alternativo e aproximadamente 37% dos mesmos possuem isoformas proteicas correspondentes, contribuindo para a complexidade do proteoma no contexto de

diversos processos biológicos (BLACK, 2003; PAN e colab., 2008; WAHL e colab., 2009; WANG, Eric T e colab., 2008).

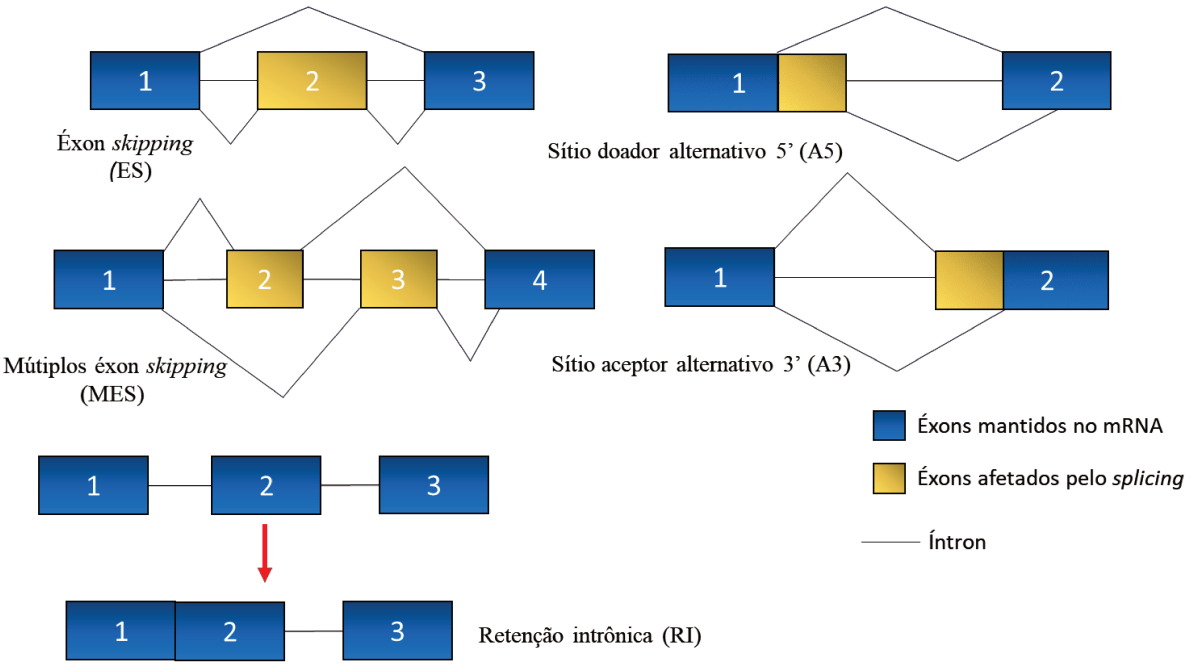


Figura 03. Classificação tradicional dos tipos básicos de eventos de *splicing*. Éxons representados em azul são mantidos até o final do processamento do mRNA. **A**: Éxon *skipping* mantém os cassetes de éxons representados em azul, pulando um éxon. **B**: Múltiplos éxons *skipping* mantém cassetes de éxons, pulando mais de um éxon. **C**: Retenção intrônica preserva um ou mais cassetes intrônicos. **D** e **E:** Sítio de *splice* alternativo que permite encurtamento ou alongamento de um éxon.

Até recentemente, os estudos de *splicing* eram limitados a genes individuais, mas foi possível iniciar uma nova era de melhor caracterização do transcriptoma, com o advento do sequenciamento de nova geração associado à análise computacional (PAN e colab., 2008; WANG, Eric T e colab., 2008). Embora já houvesse vários estudos reportando éxon *skipping* único, múltiplos éxons *skipping* e *splicing* alternativo dos sítios 5' e 3', relatos de retenção intrônica eram raros e negligenciados, sendo frequentemente interpretados como contaminação de moléculas de pré-mRNA (NILSEN e GRAVELEY, 2010). Agora, com uma visão mais ampliada, oferecida pelas novas tecnologias de sequenciamento de RNA (RNA-seq), a retenção intrônica se tornou um tipo de *splicing* alternativo emergente, que parece contribuir bastante para a diversidade do transcriptoma, em função do tipo de tecido, fase do desenvolvimento e padrões de estresse, sugerindo uma regulação fina pela maquinaria de *splicing* (BOUTZ e colab., 2015; CHO e colab., 2014; SHALGI e colab., 2013). A prevalência da RI é maior em

plantas e menor em humanos (MARQUEZ e colab., 2012; REDDY e colab., 2012), nos fazendo pensar na existência de uma diferença entre as espécies no reconhecimento de éxons e íntrons durante a implementação do *splicing* alternativo (REDDY e colab., 2012).

A retenção intrônica poderia ser um mecanismo pós-transcricional de regulação negativa da expressão gênica via indução da maquinaria de vigilância de degradação (YAP e colab., 2012) ou por *Nonsense Mediated Decay* (NMD), através da introdução de um codón de terminação prematuro (CTP) (WONG e colab., 2013). Alternativamente, a retenção intrônica poderia fazer parte de um mecanismo de regulação de expressão gênica, no qual o bloco intrônico seria liberado em resposta à determinada sinalização (NINOMIYA e colab., 2011), ou estímulo de desenvolvimento (BOOTHBY e colab., 2013).

Muitos processos biológicos que determinam funções celulares essenciais são atribuídos a compartimentos genéricos da célula, partindo da premissa de que as organelas e subcompartimentos celulares sejam estáticos. Esta perspectiva não poderia estar mais distante da realidade, uma vez que uma dinâmica e complexa relação com o citoesqueleto define o posicionamento dessas organelas e microestruturas celulares, que servem de base estrutural dos subcompartimentos onde as reações e processos biológicos se desenvolvem (VAN BERGEIJK e colab., 2016). Macromoléculas, tais como RNA e proteínas, devem navegar a microanatomia intracelular para atingir seus sítios de ação, sendo fundamental que compreendamos melhor as regras que governam este trânsito. Alguns estudos reportaram que a retenção de íntrons pode funcionar como sinal de localização subcelular dos transcritos que os contém. A RI foi essencial para o direcionamento de 33 mRNAs neurais para os dendritos, ao ponto de afetar a localização onde aconteceria sua tradução (BUCKLEY e colab., 2011). Além disso, foi observado que um transcrito com retenção intrônica que codifica uma variante da chaperona GRP78 (GRP78va) atua de forma difusa no citoplasma, enquanto o produto de sua isoforma canônica fica retido no lúmen do retículo endoplasmático, exercendo ali sua função (NI, Min e colab., 2009). A complexidade desta área de investigação é ilustrada pela recente identificação de sequências intrônicas codificadoras de proteínas, as quais receberam uma nomenclatura própria, sendo chamadas de exitrons (STAIGER e SIMPSON, 2015). Sendo assim, alguns dogmas começam a ser quebrados e haverá muito debate na literatura nos próximos anos sobre as hipóteses de funcionamento dos transcritos com retenção intrônica (Figura 04). Esses transcritos em geral costumam ter pequenos íntrons com alto teor de GC (AMIT e colab., 2012; KEREN e colab., 2010; SAKABE e DE SOUZA, 2007; WONG e colab., 2013), os quais frequentemente resultam de sítios de *splicing* fracos que escapam do spliceossomo (SAKABE e DE SOUZA,

2007). A cromatina também pode influenciar este proce sso, uma vez que regiões com alta densidade de nucleossomas favorecem a retenção intrônica, assim como o fazem distintas modificações de histonas nos éxons (ANDERSSON e colab., 2009; SCHWARTZ e colab., 2009; SPIES e colab., 2009) e marcas epigenéticas específicas (LUCO e colab., 2010).

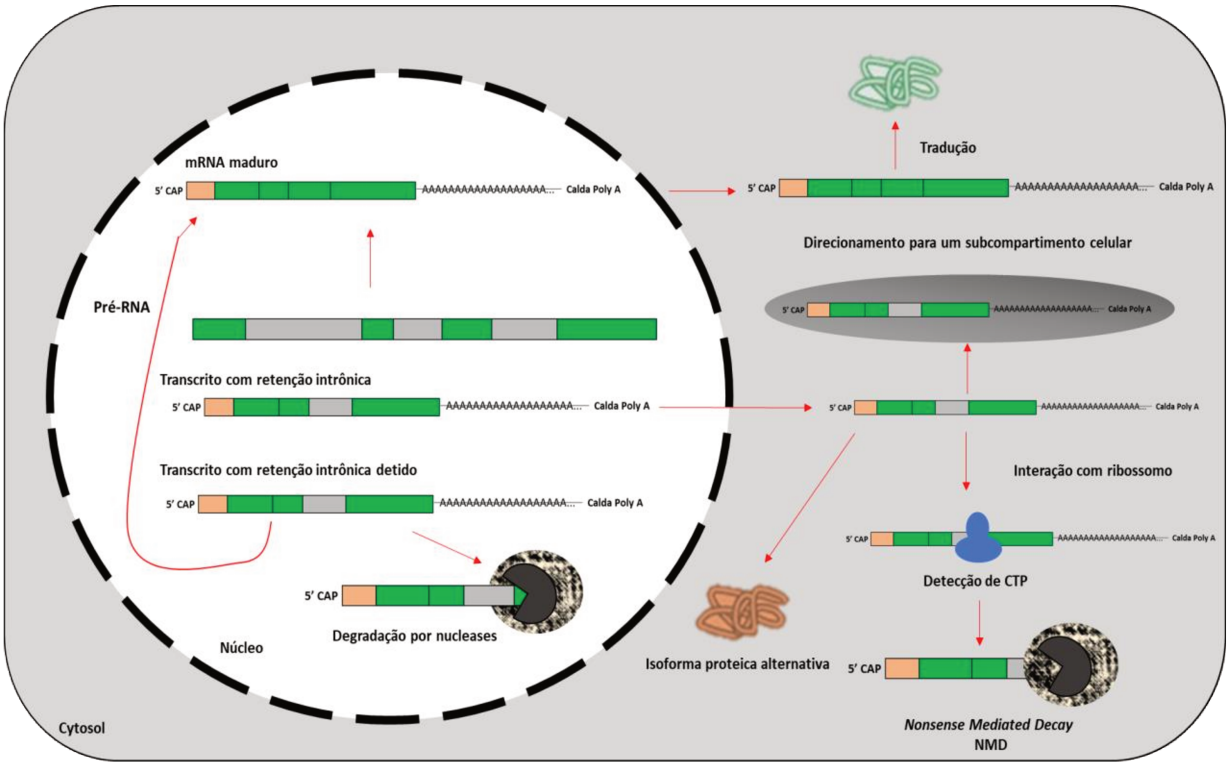


Figure 04. Hipóteses sobre os destinos dos transcritos com retenção intrônica. O transcrito sem este evento de *splicing* alternativo de retenção intrônica em geral deixa o núcleo para ser traduzido em proteína no citoplasma. Um transcrito com um ou mais íntrons retidos pode haver diferentes destinos. Um transcrito com retenção que contem um código prematuro pode ser reconhecido pela maquinaria da NMD e então ser degradado. Outro possível destino do transcrito é codificar para uma isoforma proteica alternativa. Há relatos que transcritos com retenção que podem ser direcionados a um subcompartimento celular aonde deverão ser traduzidos. Alguns destes transcritos costumam ficar detidos no núcleo para posteriormente serem degradados por nucleases. Contudo, há também um mecanismo descrito que permite a excisão do íntron do transcrito detido no núcleo para exporta-lo ao citoplasma e traduzi-lo em proteína. Figura baseada em Vanichkina e colaboradores (VANICHKINA e colab., 2017).

O linfócito T é um modelo clássico de estudo de *splicing* alternativo, em que se descobriu que o processamento do pré-mRNA do gene *Ptprc* produz diferentes isoformas proteicas de uma tirosina-fosfatase de membrana (CD45) (HERMISTON e colab., 2003;

MARTINEZ e LYNCH, 2013) durante as fases de diferenciação celular. Para que ocorra diferenciação, é necessária a ativação do linfócito T pelo sinal um, que é desencadeado pelo reconhecimento por parte do receptor de células T (TCR) de pequenos fragmentos antigênicos apresentados pelo complexo de histocompatibilidade (MHC) expresso nas células apresentadoras de antígeno, bem como por um segundo sinal produzido pelo engajamento de um receptor coestimulador como, por exemplo, o CD28. O disparo simultâneo dos sinais um e dois desencadeia o processo de maturação do linfócito, que parte do fenótipo da célula *naive* em repouso e atravessa várias etapas progressivas de diferenciação, para atingir o fenótipo do linfócito armado efetor. Uma vez que as células atingem a diferenciação terminal, há três rotas possíveis: o exercício da função efetora, a morte por apoptose e o recolhimento ao compartimento de memória, nos formatos de células T tronco de memória central (T_{SCM}) (ZHANG, Yi e colab., 2005), células T de memória central (T_{CM}) e células T de memória efetora (T_{EM}) (GATTINONI, L e colab., 2011; GATTINONI, Luca e colab., 2006; ZHANG, Yi e colab., 2005). A ativação do linfócito T está associada à produção de granzimas, perforina, citocinas efetoras, espécies reativas de oxigênio, bem como à proliferação celular (SMITH-GARVIN e colab., 2009).

A ativação do linfócito T é um dos modelos fenotípicos mais estudados, no que diz respeito à cromatina, com vários trabalhos de pesquisa em epigenômica na literatura (BARSKI e colab., 2007; POKHOLOK e colab., 2005; WANG, Zhibin e colab., 2008). Esses estudos possibilitaram alcançar um melhor entendimento sobre a assinatura epigenética dos linfócitos, com a identificação de anotações funcionais envolvidas na regulação e execução transcricional (ERNST e KELLIS, 2012; WANG, Zhibin e colab., 2008). Estudos genômicos da ativação do linfócito T revelaram que muitos dos genes silentes nas células T em repouso já carregam marcas epigenéticas na cromatina, com altos níveis de RNApol II no promotor, que são indicadoras de previa ativação. Curiosamente, após ativação do linfócito T, observa-se pouca ou quase nenhuma modificação no estado da cromatina, o que parece indicar que certos genes já poderiam estar preparados para transcrição rápida em resposta ao estímulo ativador (BARSKI e colab., 2009; CUDDAPAH e colab., 2010).

Foi demostrado que a retenção intrônica é prevalente nas células T CD4+ em repouso, mas os transcritos produzidos por esse tipo de *splicing* alternativo desaparecem, quando os linfócitos são ativados através do TCR e CD28. Esse estudo revelou também que os transcritos com retenção intrônica têm maior *turnover* nas células em repouso e sugere que o

processamento adicional dos mRNAs durante a ativação do linfócito CD4+ poderia estabilizá-los, afetando o estado das isoformas proteicas codificadas (NI, Ting e colab., 2016).

Neste trabalho, abordamos indiretamente uma questão fundamental sobre a relevância biológica da retenção intrônica, que surgiu depois da descoberta de que esta forma de *splicing* alternativo se estende além da biologia das plantas ao transcriptoma de mamíferos. Procuramos responder se a retenção intrônica representa alguma forma de ruído *(background)* transcricional e o quanto ela é seletiva para fenótipos definidos da célula, assim como se existe variação interindividual.

2. HIPÓTESE

Lançamos a hipótese de que o *splicing* alternativo por retenção intrônica não seja um mero ruído transcricional que acompanha passivamente a flutuação da expressão gênica global e que este tipo de *splicing* alternativo possa ocorrer de maneira seletiva durante os estados de repouso e ativação do linfócito T.

3. OBJETIVOS

- Construir uma ferramenta algorítmica para identificação e análise de retenção intrônica;

- Validar a ferramenta;

- Processar e construir o transcriptoma do linfócito T CD4+ exposto à ativação do TCR e CD28;

- Determinar o padrão de transcritos com retenção intrônica nos genes analisados por RNA-seq

- Testar e validar os transcritos contendo retenção intrônica por RT-PCR;

- Estudar a ocorrência da retenção intrônica em diferentes estados de ativação dos linfócitos T CD4+.

4. MATERIAIS E MÉTODOS

4.1 Construção de uma ferramenta algorítmica para análise de retenção intrônica

Geramos um programa de identificação de transcritos que determina as formas de *splicing* alternativo e, dentre elas, aquelas que exibem retenção intrônica. Empregamos dois pacotes essenciais nesta tarefa. Usamos o cummeRbund para criar os bancos de dados que descrevem as relações entre os genes, transcritos e regiões dos transcritos (COSSIO e colab., 2012). Para identificação dos diferentes tipos de *splicing* alternativo, usamos spliceR (VITTING-SEERUP e colab., 2014). Construímos um código capaz de assimilar os arquivos gerados pelo processamento dos dados brutos de RNA-seq e gerar uma lista de isoformas contendo retenção intrônica, conforme nossas perguntas biológicas, assim como, tabelas essenciais e figuras necessárias para a realização deste trabalho.

4.2 Validação e padronização do processamento dos dados

Utilizamos a ferramenta criada para processar os dados de RNA-seq com alta cobertura (~150 milhões de *reads*), adquiridos no Arquivo de Sequências de Reads (SRA) do Centro Nacional de Informação Biotecnológica dos Estados Unidos (NCBI). A indexação do genoma humano UCSC hg19 foi realizada com o bowtie2. Logo então, o mapeamento dos reads foi conduzido com o Tophat2 e a montagem dos transcritos realizada pelo cufflinks. Com os arquivos gerados, foi possível juntar as amostras com o cuffmerge e processar as tabelas dos genes diferencialmente expressos com o cuffdiff (COSSIO e colab., 2012). Fizemos mudanças nos parâmetros desta *pipeline*, de acordo com Dvinge & Bradley (DVINGE e BRADLEY, 2015): –min-anchor-length 6, --splice-mismatches 0, --min-intron-length 10, --max-intron—length 1,000,000, --min-isoform-fraction 0.0.

4.3 Isolamento de células T CD4+

Células mononucleares do sangue periférico (PBMC) foram isoladas através de centrifugação em gradiente de Ficoll, de acordo com o protocolo do fabricante (GE Healthcare). Foi feita a depleção de monócitos por aderência diferencial ao plástico durante 1h a 37°C, em uma atmosfera de 5% CO_2. As células mononucleares resultantes foram isoladas, usando Dynabeads Untouched Human CD4+ T cells kit (ThermoFisher Cat. N. 1135D), sem serem ativadas.

4.4 Ativação dos linfócitos T CD4+

As células T CD4+ foram cultivadas em meio RPMI com 10% de soro fetal bovino em placas de 24 poços. A ativação foi feita após o isolamento das células CD4+, usando Gibco Dynabeads Human T-Activator CD3/CD28 (ThermoFisher Cat. N. 111.31D), para efetuar o disparo dos sinais um e dois e conduzir à completa ativação do linfócito T.

4.5 Verificação da ativação e isolamento das células T CD4+

Após 18h de incubação das células T com os Dynabeads ativadores, parte das amostras foi coletada para verificação da qualidade do processo de isolamento celular, assim como do nível de ativação celular. Para isso, as células foram marcadas com um anticorpo anti-CD4 para verificar a qualidade do isolamento, tendo as células com o fenótipo de interesse que contemplar 90% da população depois da separação. Já para a verificação do estado de ativação, foi utilizado um anticorpo anti-CD25. As células T em repouso costumam ter expressões baixas deste marcador de superfície, enquanto células T ativadas costumam mostrar um aumento dramático da expressão do mesmo, pouco tempo após a ativação.

4.6 Extração de RNA e síntese de cDNA

A extração de RNA foi realizada após 18h de ativação das células T, utilizando o reagente Trizol. A qualidade do RNA foi testada em gel de eletroforese e espectrofotometria (Quibit da Thermo Fisher). As amostras utilizadas não exibiram sinais de degradação ou contaminação. Para síntese de cDNA, usou-se o kit Improm II Reverse Transcription System (Promega Cat. N. A3800). Empregamos oligo(dT), ao invés de *random primers,* para assegurar maior confiabilidade dos resultados, uma vez que são excluídos transcritos sem cauda polyA.

4.7 Validação dos genes com retenção intrônica em seus transcritos

A validação das isoformas intrônicas foi realizada através do método de reação em cadeia da polimerase (PCR), para amplificar os transcritos escolhidos. Nesta estratégia, foi usado *nested*-PCR, uma vez que o isolamento dos transcritos relevantes é complicado pela expressão diluída dos mesmos no conjunto de transcritos de cada gene. Desta forma, foi feita uma amplificação inicial das amostras, utilizando *primers* externos desenhados nas regiões *upstream* do éxon e *downstream* do íntron. Para a amplificação da segunda reação do *nested*-PCR, utilizou-se 0,5 µl da primeira reação, um *primer* que reconhece uma sequência dentro do íntron e outro que reconhece o éxon adjacente. Garantimos, assim, uma pré-amplificação dos templatos, para realizar uma amplificação específica do transcrito com retenção intrônica numa segunda etapa. O número de ciclos foi adaptado para cada gene, sendo TIMMDC1, GAS5, LY9, PIK3R1, CCNL1, FYN, RHOA, HLA-B amplificados com 30 ciclos na primeira reação e ANKZF1, UBXN11, SH3BRGL3 com 35 ciclos. Já na segunda reação, usamos 30 ciclos para TIMMDC1, GAS5, LY9, PIK3R1, ANKZF1, UBXN11 e 35 ciclos para CCNL1, FYN, RHOA, HLA-B, SH3BRGL3. A lista de *primers* encontra-se nas tabelas abaixo:

Tabela 1 – Lista de *Primers* para a primeira reação do *nested*-PCR

	Primer Forward	***Primer Reverse***
ANKZF1	CAAGAGCTACAGCGTGTGCTCC	GAGAGGTAGGGGCTCCCAACTG
GAS5	ACCACTGGGAGGCTGAGGAT	GGTTTTGGTGGCATATACACCTT
LY9	CCATGCCTACGTGTGCTCAG	AGATCATTCTGTTGTGGTGGTGG
PIK3R1	GCCAAGGAAACTCTTGCACA	GCCCAAAACCCTAATTTCAGC
TIMMDC1	GGTCAAATGCACGGATTCTCAC	CACTCTTCCAGTTTTAGCTCATGG
UBXN11	GCTGCCTGCTGCTCTGTTTC	CGAGGAAGTTGTGGGACCTG
SH3BGL3	GCGTCTACAGCACGTCGGT	TGGCTGATGGAACCAACAGA
HLA-B	CCTTCCCGTTCTCCAGGTATC	AGTCTCCTCAGACGCCGAGAT
FYN	TTCGGATTTGGGGACAAGTG	CCCTGTACGGGAGGTTCACA
CCNL1	GCCAGTTTTCCAGCCCAAG	CGTAATCAAACCCTGGTTCAAA
RHOA	TGGTGTGTCAGGTGGGAGTG	ATCGAGGTGGATGGAAAGCA

Tabela 2 – Lista de *Primers* para segunda reação do *nested*-PCR

	Primer Forward	***Primer Reverse***
ANKZF1	CCTTGGTCTGGTTGGGTGAT	GTGATGCCTTCCCACTCTGC
GAS5	GCACTCTAGCTTGGGTGAGG	TCATCTTCAAACAGGCAGACC
LY9	ATCACCCCAACCTCACATGC	GCCTGTCCTCATCCTCCTCA
PIK3R1	CACTGCCTCCTAAACCACCAA	GCCCAAAACCCTAATTTCAGC
TIMMDC1	GAGTCTTTGCTGCCGAAGCT	CTTTGCCAAACAGCTCCCG
UBXN11	CCCTTCAAAATCCCCAAAGG	CAGGTGAAGGCCCAGACTGA
SH3BGL3	GGACATCTCCCAGGACAACG	GTTCCACAGCCTCCACGAAG
HLA-B	GGATGGGGAGTCGTGACCT	GAACCGTCCTCCTGCTGCT
FYN	TTGGTACTGGGGCTCTGTCG	GCATGAACAACCGGGAGGT
CCNL1	CCCATTGTAAGGGCTTTTGGA	GCATGTGCTTGCATCTACCTTG
RHOA	CAGACGGGTTGGACATCGTT	TTCGGAATGATGAGCACACAA

4.8 Análise estatística

Em diferentes seguimentos do trabalho de dissertação aqui apresentado, o teste *t* de Student foi empregado para comparar dois grupos experimentais. As diferenças foram consideradas significativas quando $p < 0,05$. Para avalição estatística descritiva em ambiente de programação R, aplicamos o coeficiente de correlação de Spearman, onde o $r = 1$ significa uma correlação positiva perfeita, $r = -1$ indica uma correlação negativa perfeita e $r = 0$ significa que as duas variáveis não dependem linearmente uma da outra. A ferramenta Cuffdiff foi usada para calcular a expressão diferencial dos transcritos. Para isso, o Cuffdiff realiza um test *t* unilateral baseado na métrica de Jensen-Shannon.(COSSIO e colab., 2012).

5. RESULTADOS

5.1 Processamento dos dados e programação da ferramenta de análise de *splicing*

O processamento foi efetuado em *unix commands,* utilizando dados de RNA-seq, baixados em FastQ do banco de arquivos SRA-NCBI. Os dados brutos referentes às amostras de linfócitos T CD4+ em repouso (SRR2033046) e ativados (SRR2033047) foram gerados através de um protocolo de sequenciamento de RNA por ligação direcional (Deli-seq) *single end* com purificação dupla de polyA. Os arquivos foram baixados através do SRA toolkit (versão 2.8.2-1), com código fastq-dump – 'código SRR da amostra'. Esses arquivos foram podados pela ferramenta Trimmomatic (versão 0.36), para remoção de adaptadores, bases de baixa qualidade de leitura ou bases N, bem como *reads* abaixo de 36 pares de base, usando o código descrito abaixo:

```
$ java -jar trimmomatic-0.36.jar SE -phred33 SRR2033046.fastq SRR2033046_trimmed.fastq ILLUMINACLIP:TruSeq3-SE:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36

$ java -jar trimmomatic-0.36.jar SE -phred33 SRR2033047.fastq SRR2033047_trimmed.fastq ILLUMINACLIP:TruSeq3-SE:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36
```

Após o emprego deste filtro, foi verificada a qualidade das amostras usando o FastQC (versão 0.11.5). O FastQC gera um *boxplot* para cada amostra que identifica o intervalo interquartil de 25% a 75% em amarelo; em vermelho é indicado a mediana; em preto estão os *whiskers* que vão de 10% a 90%; e em azul está a linha que representa a média da qualidade dada pelo programa. Uma vez que os *scores* obtidos por este software eram > 20 (Figura 05 e Figura 06), o que é de boa qualidade segundo os criadores do programa (Babraham Bioinfomatics), prosseguimos com a execução do protocolo *Tuxedo*. Este protocolo requer que geremos arquivos compatíveis com o alinhamento genômico, o que foi realizado através do comando bowtie2-build do programa Bowtie2 (versão 2.3.2).

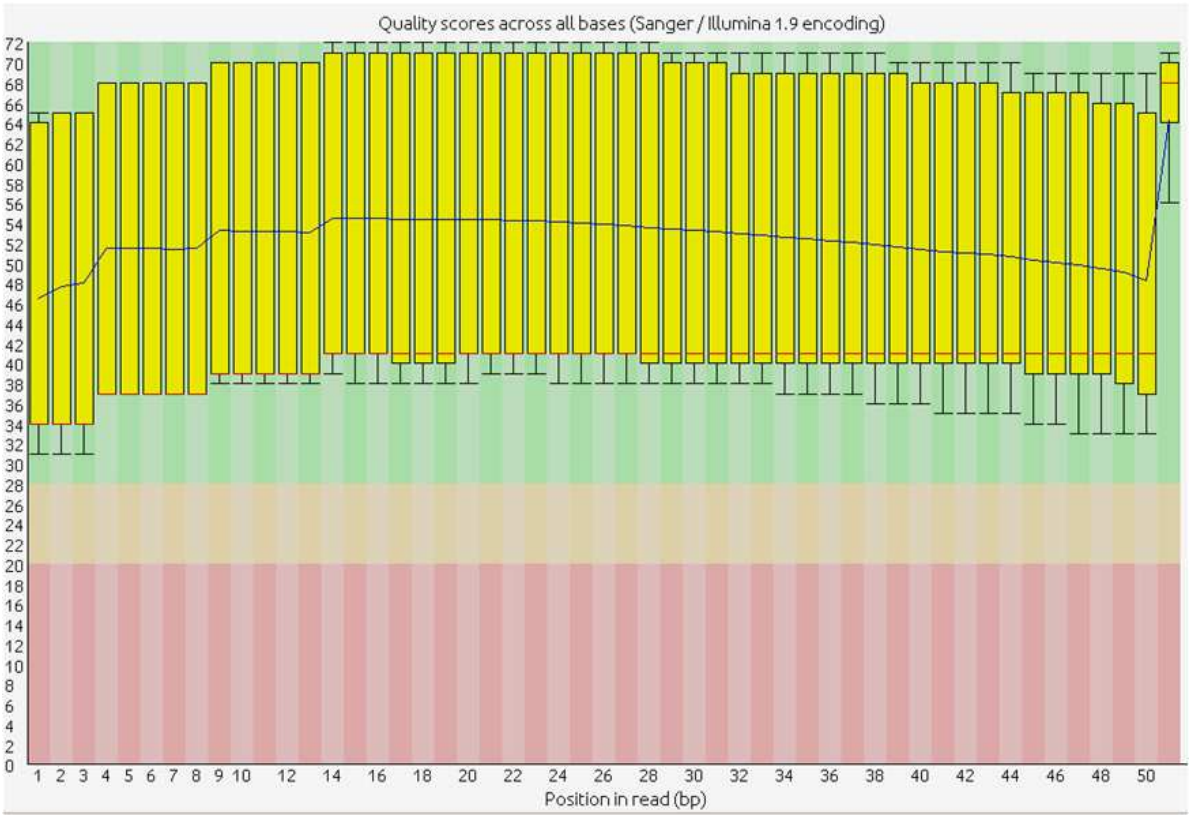


Figura 05. Resultado da análise de qualidade das bases identificadas no sequenciamento da amostra SRR2033046.fastq (T CD4+ em repouso) pelo software FastQC. Pode-se observar uma pontuação de qualidade (*Quality score*) acima de 30, o que representa uma boa média para análises posteriores.

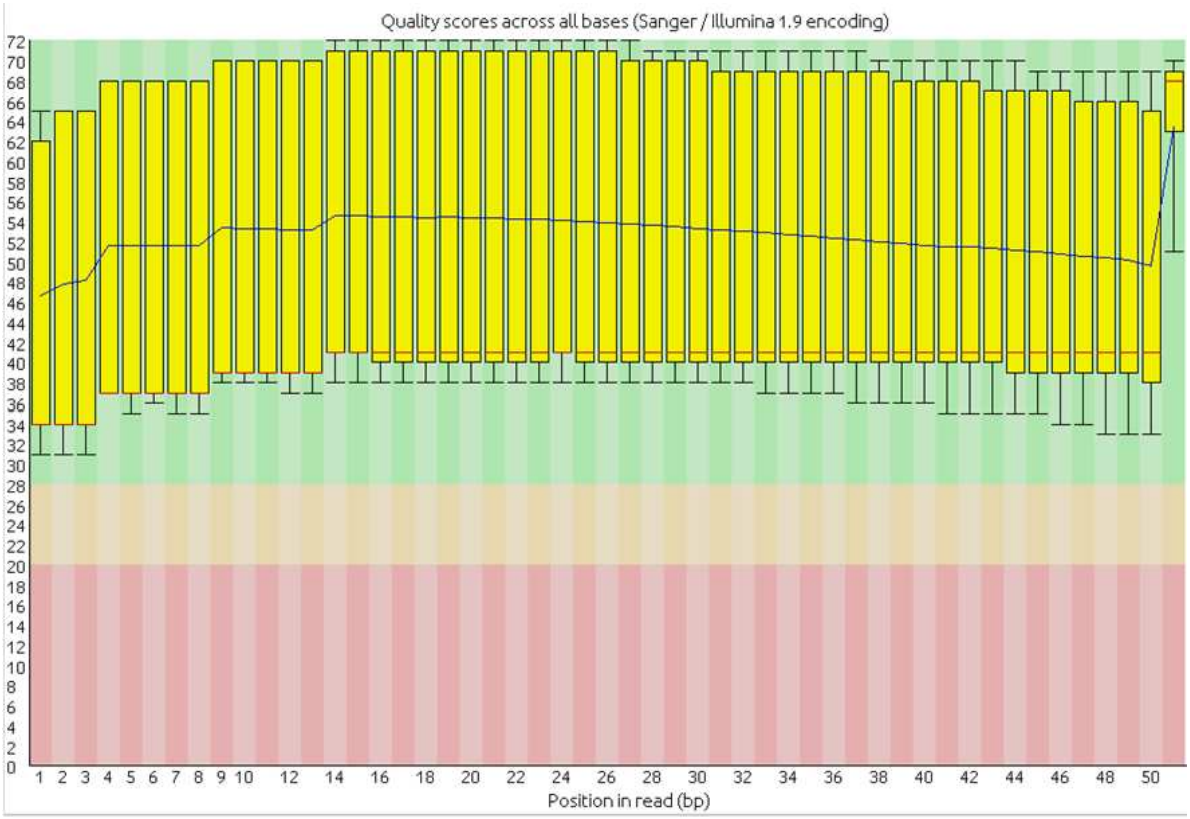


Figure 06. Resultado da análise de qualidade das bases identificadas no sequenciamento da SRR2033047.fastq (T CD4+ ativado) pelo software FastQC. Pode-se observar uma pontuação de qualidade acima de 30, o que também indica uma boa qualidade para análises posteriores.

O alinhamento das amostras foi conduzido pelo Tophat2 (versão 2.1.1), utilizando parâmetros um pouco diferentes do padrão, para conseguir detectar transcritos contendo retenção intrônica. Utilizamos um mínimo de âncora de 6, um erro de pareamento de *splicing* de 0, mínimo de tamanho de retenção intrônica de 10 e máximo de 1.000.000. A indexação genômica foi feita com o genoma de referência UCSC hg19. Tophat2 gera arquivos comprimidos (.bam) no formato de alinhamento de sequenciamento (.sam). Os dados obtidos pelo alinhamento realizado pelo programa Tophat2 se revelaram de boa qualidade (> 97% de mapeamento) (tabela 03), permitindo prosseguir com o protocolo. O código utilizado é descrito abaixo:

```
$  ./tophat2 -o  ~Bioinfo/Tophat2-Bowtie2/Results/Resting_Cell -a 6 -m 0 -i 10 -I 1000000
-p 12 -hg19 SRR2033046_trimmed.fastq

$ ./tophat2 -o  ~Bioinfo/Tophat2-Bowtie2/Results/Activated_Cell -a 6 -m 0 -i 10 -I
1000000 -p 12 hg19 SRR2033047_trimmed.fastq
```

Tabela 03 – Resultados obtidos a partir do alinhamento com o programa Tophat2

	Célula T CD4+ em Repouso	**Célula T CD4+ Ativada**
Entrada:	146155086 *reads*	179467597 *reads*
% de *reads* mapeados:	141936045 (97,1%)	174589910 (97,3%)
% de *reads* com múltiplos alinhamentos	16364593 (11,5%) tem múltiplos alinhamentos (546131 tem > 2).	17991064 (10,3%) tem múltiplos alinhamentos (310815 tem > 2)

Utilizamos o cufflinks (versão 2.2.1) para montagem dos transcritos e estimar as isoformas transcricionais. O arquivo de saída em .gtf é unido através do cuffmerge ao mesmo tempo em que os nomes dos genes para cada *locus* são anotados. Por último, lançamos mão do cuffdiff para calcular a expressão diferencial entre as amostras. Os códigos estão descritos abaixo:

Cufflinks:

$ ./cufflinks -o ~/Bioinfo/cufflinks-2.2.1/Resting_Cell -p 12 -F 0. ~/Bioinfo/Tophat2-Bowtie2/Results/Resting_Cell/accepted_hits.bam

$ ./cufflinks -o ~/Bioinfo/cufflinks-2.2.1/Activated_Cell -p 12 -F 0.0 ~/Bioinfo/Tophat2-Bowtie2/Results/Activated_Cell/accepted_hits.bam

Cuffmerge:

$ ./cuffmerge -o ~/Bioinfo/cufflinks-2.2.1/Cuffmerge --min-isoform-fraction 0.0 -g genes.gtf -p 12 assembly_list.txt

Cuffdiff:

$ ./cuffdiff -o ./ -L Resting_CD4+_T_cell,Activated_CD4+_T_cell -p 24 ~/Bioinfo/cufflinks-2.2.1/CuffmergeNV/merged.gtf ~/Bioinfo/Tophat2 Bowtie2/Results/ Resting_Cell/accepted_hits.bam ~/Bioinfo/Tophat2-Bowtie2/Results/Activated_Cell/ accepted_hits.bam

O segundo passo foi a construção de uma ferramenta algorítmica para identificação dos transcritos com retenção intrônica. Por esta razão, criamos uma *pipeline* que, a partir dos dados de RNA-seq processados, nos permitiu gerar tabelas com o nível de expressão dos genes

e separar aqueles com diferentes tipos de *splicing*. Então, escrevemos um programa em linguagem R com a finalidade de nos servirmos dos pacotes cummeRbund e spliceR. O cummeRbund é um pacote que permite o armazenamento, acesso e exploração dos dados de sequenciamento derivados do cufflinks, com algumas funções de visualização comumente usadas na área. Já o spliceR é outro pacote para classificação de *splicing* alternativo para dados derivados de RNA-seq.

A *pipeline* foi criada começando pela leitura dos dados derivados do cufflinks e do cuffdiff. Então, foram produzidos *objetivos de GRanges* (intervalos genômicos) para compor uma lista, em que foram anotados todos os objetos e isoformas transcricionais contendo as classes de *splicing* alternativo, assim como outros metadados. Uma vez compilada a lista de todas as isoformas transcricionais e suas anotações, o código permitiu gerar sublistas, tais como aquela de transcritos contendo retenção intrônica ou a sublista apenas de genes. Outras duas importantes finalidades desta ferramenta foram criar a possibilidade de identificar isoformas transcricionais que aparecem somente em um dos grupos analisados, bem como fazer a visualização dos dados. Segue o código baixo:

```
#================IDENTIFICAÇÃO DE RETENÇÃO INTRÔNICA================#


##Instalar e carregar os pacotes necessários:
source("https://bioconductor.org/biocLite.R")
biocLite(c("spliceR", "cummeRbund"))
n


library(spliceR, cummeRbund)

##Criar um banco de dados cuff (cuffDB): gtf merged file + arquivos cds/gen
e/isoforms/TSS/promoters/splicing do cuffdiff.


#Usar genoma hg19 como referência e reconstruir. Ex:
cuffDB <- readCufflinks(dir= "C:/Users/Pedro/Documents/Cuffdiff", gtfFile =
"C:/Users/Pedro/Documents/Cuffdiff/merged.gtf", genome = 'hg19', rebuild =
TRUE)


#Preparar cuffDB para o spliceR:
cuffDB_spliceR <- prepareCuff(cuffDB)


#Correr o spliceR para criação dos Granges e listas.
```

```
mySpliceRList <- spliceR(cuffDB_spliceR, compareTo='preTranscript', filters
=c('expressedGenes','geneOK', 'isoOK', 'expressedIso', 'isoClass'))


#Visualizar um diagrama de Venn com isoformas globais com retenção intrônic
a

mySpliceRList <- spliceRPlot(mySpliceRList, evaluate="nr_AS", asType="ISI")


#Visualizar números de transcritos por gene em diagrama de venn

#Transcript_gene <- spliceRPlot(mySpliceRList, evaluate="nr_transcript_pr_g
ene")


#Você pode mudar a cor se preferir:

#mySpliceRList <- spliceRPlot(mySpliceRList, evaluate="nr_AS", asType="ISI"
, colors = c("green", "blue", "orange"), alpha = 0.5)


#Identificar as isoformas e lista isoformas com retenção intrônica em conju
nto de dados

#Criar uma tabela com as isoformas:

isoforms_list = as.data.frame(mySpliceRList[[6]]$isoforms)



#Identificar genes em um conjuto de dados:

genes_list = as.data.frame(mySpliceRList[[6]]$genes)


#Filtrar tabelas de isoformas e genes com retenção intrônica. Criar lista d
e isoformas com retenção intrônica:

#Não precisamos de ISI (retenção intrôniaca) = 0


intron_retention_list = subset(isoforms_list, isoforms_list$ISI > 0)


#Recuperar IDs dos genes (XLOC)

get_gene_id <- getSig(cuffDB,level="genes")


#Usar para criar o retorno de idenficadores do objeto CuffgeneSet com todas
informações relevantes do gene:

#gene_short_name values (and corresponding XLOC_* values) can be retrieved
from the CuffGeneSet by using:

#nomes do gene_short_name (correspondendes a XLOC) podem ser recuparados pe
lo CuffGeneSet usando:

names <- as.data.frame(featureNames(get_genes))

#Mudar o nome do tracking_id para gene_id

names(names)[1] <- paste ("gene_id")
```

```
#Unir XLOC com nome do gene

genes_list <- merge.data.frame(names, genes_list, by = "gene_id", all = TRU
E)


#Unir XLOC com nome dos genes nas listas

intron_retention_list <- merge.data.frame(names, intron_retention_list, by
= "gene_id", all = TRUE)


isoform_WO_intron_retention <- merge.data.frame(names, isoform_WO_intron_re
tention, by = "gene_id", all = TRUE)


isoforms_list <- merge.data.frame(names, isoforms_list, by = "gene_id", all
= TRUE)


#Filtrar ISI:

intron_retention_list = subset(intron_retention_list, intron_retention_list
$ISI > 0)


#Troca o 0 para NA na tabela, caso seja melhor:

#intron_retention_list[intron_retention_list == 0] <- NA



#Ajustar as tabelas e gerar documentos finais:

diff_expression <- read.table("C:/Users/Pedro/Documents/Cuffdiff/isoform_ex
p.diff")


names(intron_retention_list)[3] <- paste("isoform_id")

intron_retention_list <- merge.data.frame(diff_expression, intron_retention
_list, by = "isoform_id")

intron_retention_list$`Sample 1` <- NULL

intron_retention_list$`Sample 2` <- NULL

intron_retention_list$Status <- NULL

intron_retention_list$`value 1` <- NULL

intron_retention_list$`value 2` <- NULL

intron_retention_list$test_stat <- NULL

intron_retention_list$p_value <- NULL

intron_retention_list$q_value <- NULL

intron_retention_list$significant <- NULL

intron_retention_list$gene_id.y <- NULL

intron_retention_list$gene_short_name <- NULL

intron_retention_list$`log2(fold_change)` <- NULL
```

```
names(intron_retention_list)[2] <- paste("gene_id")
write.csv2(intron_retention_list, file = "intron_retention_list")


names(gene_list)[3] <- paste("isoform_id")
genes_list <- merge.data.frame(diff_expression, genes_list, by = "isoform_i
d")
genes_list$`Sample 1` <- NULL
genes_list$`Sample 2` <- NULL
genes_list$Status <- NULL
genes_list$`value 1` <- NULL
genes_list$`value 2` <- NULL
genes_list$`log2(fold_change)` <- NULL
genes_list$test_stat <- NULL
genes_list$p_value <- NULL
genes_list$q_value <- NULL
genes$significant <- NULL
genes_list$gene_id.y <- NULL
names(genes_list)[2] <- paste("gene_id")
write.csv2(genes_list, file = "genes_list")



names(isoforms_list)[3] <- paste("isoform_id")
isoforms_list <- merge.data.frame(diff_expression, isoforms_list, by = "iso
form_id")
isoforms_list$`Sample 1` <- NULL
isoforms_list$`Sample 2` <- NULL
isoforms_list$Status <- NULL
isoforms_list$`value 1` <- NULL
isoforms_list$`value 2` <- NULL
isoforms_list$`log2(fold_change)` <- NULL
isoforms_list$test_stat <- NULL
isoforms_list$p_value <- NULL
isoforms_list$q_value <- NULL
isoforms_list$significant <- NULL
isoforms_list$gene_id.y <- NULL
isoforms_list$gene_id.y <- NULL
names(isoforms_list)[2] <- paste("gene_id")
write.csv2(isoforms_list, file = "isoforms_list")
```

```
#========ACHAR TRANSCRITOS DIFERENTES COM RETENÇÃO INTRÔNICA==========#

#Victor Sande me ajudou com este código:

#i = interator
#k = interator
k = 1
i = 1
diff_transcripts_w_intron_retention = "x"

intron_retention_list <- as.matrix(intron_retention_list)
for (i in 1:dim(intron_retention_list)[1]){

if (as.numeric(intron_retention_list[i,"Resting_CD4__T_cell"]) != 0 && as.n
umeric(intron_retention_list[i,"Activated_CD4__T_cell "]) == 0){
diff_transcripts_w_intron_retention[k]=intron_retention_list[i,3]
k = k + 1
}
}
intron_retention_list <- as.data.frame(intro_retention_list)
diff_transcripts_w_intron_retention



#==================Visualização dos Transcritos=======================#

#Visualizar região do gene:
#Primeiro, escolha o gene de interesse:
mygeneid <- "GENENAME"
mygene <- getGene(cuffDB, mygeneid)

#Checar anotação:
mygene

#Plottrack

tracklist <- list()
```

```
mystart <- min(features(mygene)$start)
myend <- max(features(mygene)$end)
mychr <- unique(features(mygene)$seqnames)
genome <- 'hg19'
ideotrack <- IdeogramTrack(genome = genome, chromosome = mychr)
tracklist <- c(tracklist, ideotrack)
axtrack <- GenomeAxisTrack()
tracklist <- c(tracklist, axtrack)
genetrack <- makeGeneRegionTrack(mygene)

tracklist <- c(tracklist, genetrack)

#=========================================================================#

#Adicionar transcritos do ENSEMBL
biomtrack <- BiomartGeneRegionTrack(genome = genome, chromosome = as.charac
ter(mychr),
               start= mystart, end = myend, name = "ENSEMBL", showId = T)
tracklist <- c(tracklist, biomtrack)

#=========================================================================#

#Adiciocionar picos de conservação
conservation <- UcscTrack(genome = genome, chromosome = mychr,
                 track = "Conservation", table = "phyloP100wayAll",
                 from = mystart-2000, to = myend+2000, trackType = "DataTra
ck",
                 start = "start", end = "end", data = "score",
                 type = "hist", window = "auto", col.histogram = "darkblue"
,
                 fill.histogram = "darkblue", ylim = c(-3.7, 4),
                 name = "Conservation")
tracklist <- c(tracklist, conservation)

#=========================================================================#

#Visualizar transcritos
plotTracks(tracklist, from=mystart-2000, to = myend+2000, main="GENE transc
ripts",col.main="Black", title.width = 0.2, fontcolor="black", fill = "blac
k", col.line = "darkgray", col.title = "transparent", background.title = "t
ransparent", col.axis = "darkgray", fontsize = 14)
```

```
#Checar id para ver sequência no ensemble e NCBI
features(mygene)
```

```
#========Criando tabelas e Analisando expressão diferencial===========#

library(spliceR, cummeRbund)

cuffDB <- readCufflinks(dir= "C:/Users/Pedro/Documents/Cufflinks_CD4_cells"
, gtfFile = "C:/Users/Pedro/Documents/Cufflinks_CD4_cells/merged.gtf", geno
me = 'hg19', rebuild = TRUE)

diff_expression <- read.table("C:/Users/Pedro/Documents/Cuffdiff/isoform_ex
p.diff")

#====================Expressão diferencial ============================#

#Ajustar nomes das colunas:
names(diff_expression)[1] <- paste("isoform_id")
names(diff_expression)[2] <- paste("gene_id")
names(diff_expression)[3] <- paste("Gene")
names(diff_expression)[4] <- paste("Locus")
names(diff_expression)[5] <- paste("Sample 1")
names(diff_expression)[6] <- paste("Sample 2")
names(diff_expression)[7] <- paste("Status")
names(diff_expression)[8] <- paste("value 1")
names(diff_expression)[9] <- paste("value 2")
names(diff_expression)[10] <- paste("log2(fold_change)")
names(diff_expression)[11] <- paste("test_stat")
names(diff_expression)[12] <- paste("p_value")
names(diff_expression)[13] <- paste("q_value")
names(diff_expression)[14] <- paste("significant")

#Transfromas p valor em número
diff_expression$p_value <- as.numeric(as.character(diff_expression$p_value)
)

#Filtrar a tabela com base na significância da expressão diferencial
```

```
diff_expression_significant <- subset(diff_expression, diff_expression$sign
ificant == "yes")

#Abrir intron_retenion_list

intron_retention_list <- read.csv2("C:/Users/Pedro/Documents/intron_retenti
on_list")


#Unir as tabelas:

diff_expression_IR_significant <- merge.data.frame(diff_expression_signific
ant, intron_retention_list, by = "isoform_id")


diff_expression_IR_significant$gene_id.y <- NULL

diff_expression_IR_significant$Locus.y <- NULL

diff_expression_IR_significant$Resting_CD4__T_cell <- NULL

diff_expression_IR_significant$Activated_CD4__T_cell <- NULL

diff_expression_IR_significant$Gene.y <- NULL


#=========================Gene Ontoogy================================#

#Fazer Gene Ontology Analysis no David Website e baixar os arquivos. Abrir
e salvar no excel como .CSV e abrir no R


GO_gene_exp <- read.csv2("C:/Users/Pedro/Documents/GO_gene_exp.csv")

GO_gene_AS  <- read.csv2("C:/Users/Pedro/Documents/GO_gene_AS.csv")

GO_gene_IR  <- read.csv2("C:/Users/Pedro/Documents/GO_gene_IR.csv")

#Transformar o pvalues e Benjamini values em números


GO_gene_exp$PValue <- as.numeric(as.character(GO_gene_exp$PValue))

GO_gene_exp$Benjamini <- as.numeric(as.character(GO_gene_exp$Benjamini))


GO_AS$PValue <- as.numeric(as.character(GO_AS$PValue))

GO_AS$Benjamini <- as.numeric(as.character(GO_AS$Benjamini))


GO_IR$PValue <- as.numeric(as.character(GO_IR$PValue))

GO_IR$Benjamini <- as.numeric(as.character(GO_IR$Benjamini))

#Criar um datafram com valores acima pvalor 0.01

GO_gene_exp <- subset(subset(GO_gene_exp, GO_gene_exp$PValue < 0.01))

GO_AS <- subset(subset(GO_AS, GO_AS$PValue < 0.01))

GO_IR <- subset(subset(GO_IR, GO_IR$PValue < 0.01))
```

```
#Criar dataframe com valores acima 0.05 para benjamini

GO_gene_exp <- subset(subset(GO_gene_exp, GO_gene_exp$Benjamini < 0.05))

GO_AS <- subset(subset(GO_AS, GO_AS$Benjamini < 0.05))

GO_IR <- subset(subset(GO_IR, GO_IR$Benjamini < 0.05))

#Visualizar barplot com contagem de genes com análise de Gene Ontology pelo
pvalor



barplot(order(go_list$Count, decreasing = TRUE), main="Gene Ontology", hori
z=TRUE, names.arg=(go_list$Term), cex.names=0.8)


barplot(go_list$PValue, main="Gene Ontology", horiz=TRUE, names.arg=(go_lis
t$Term), cex.names=0.8)



#==========Achar Premature Termination Codon(PTC)====================#

#Abrir o BSgenome do H.sapiens

library(BSgenome.Hsapiens.UCSC.hg19)


require("BSgenome.Hsapiens.UCSC.hg19",character.only = TRUE)

ucscCDS <- getCDS(selectedGenome="hg19", repoName="UCSC")


#Adquirir PTC

cuffDB_spliceR_PTC <- annotatePTC(cuffDB_spliceR, cds=ucscCDS, Hsapiens, PT
CDistance=50)

cuffDB_spliceR_PTC_grange <- as.data.frame(cuffDB_spliceR_PTC[[1]])

cuffDB_spliceR_PTC_intermed <- data.frame(cuffDB_spliceR_PTC_grange$isoform
_id, cuffDB_spliceR_PTC_grange$spliceR.PTC)

names(cuffDB_spliceR_PTC_intermed)[1] <- paste("isoform_id")

names(cuffDB_spliceR_PTC_intermed)[2] <- paste("PTC - NMD")



#Achar os genes diferencialmente expressos contendo rentenção intrônica com
PTC

diff_expression_IR_significant <- merge.data.frame(diff_expression_IRp0.05,
cuffDB_spliceR_PTC_intermed, by = "isoform_id")
```

5.2 Análise dos genes e eventos com retenção intrônica

O processamento dos dados de expressão das amostras analisadas com a ferramenta algorítmica aqui desenvolvida permitiu que se perguntasse primeiramente se existia alguma diferença global entre os tipos de *splicing* alternativo nos estados de repouso e ativação dos linfócitos T CD4+, contando cada evento transcricional (isoforma) de *splicing* alternativo, desconsiderando a expressão dos transcritos. Para nossa surpresa, não detectamos nenhuma diferença nos dois estados fenotípicos no que diz respeito à frequência global de eventos dos diferentes tipos de *splicing* alternativo (Figura 07, *t* test: $p > 0,05$), uma vez que a maior parte dos mesmos são compartilhados entre os dois fenótipos da célula. Contudo, pode-se afirmar que o tipo de evento de *splicing* alternativo mais comum seja o de retenção intrônica.

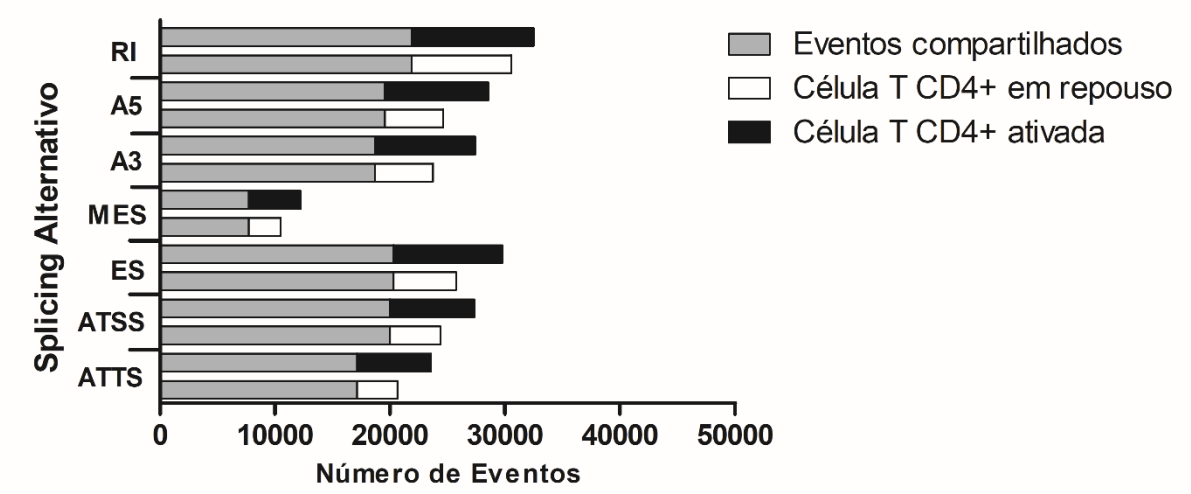


Figura 07. Número de eventos transcricionais (isoformas) globais em células T CD4+. RI: Retenção intrônica; A5: Sítio de *splice* alternativo 5'; A3: Sítio de *splice* alternativo 3'; MES: Múltiplos éxons *skipping*; ES: Éxon *skipping;* ATTS: Sítio alternativo de terminação transcricional; ATSS: Sítio alternativo de iniciação transcricional.

A análise dos dados obtidos revela que os padrões de expressão gênica das células T CD4+ em repouso e ativadas são relativamente parecidos, observando-se expressão diferencial ($p < 0,05$) de apenas 784 dos cerca de 327.000 transcritos analisados. Esta "pequena" diferença é, contudo, capaz de causar uma mudança fenotípica importante nos linfócitos (Figura 08).

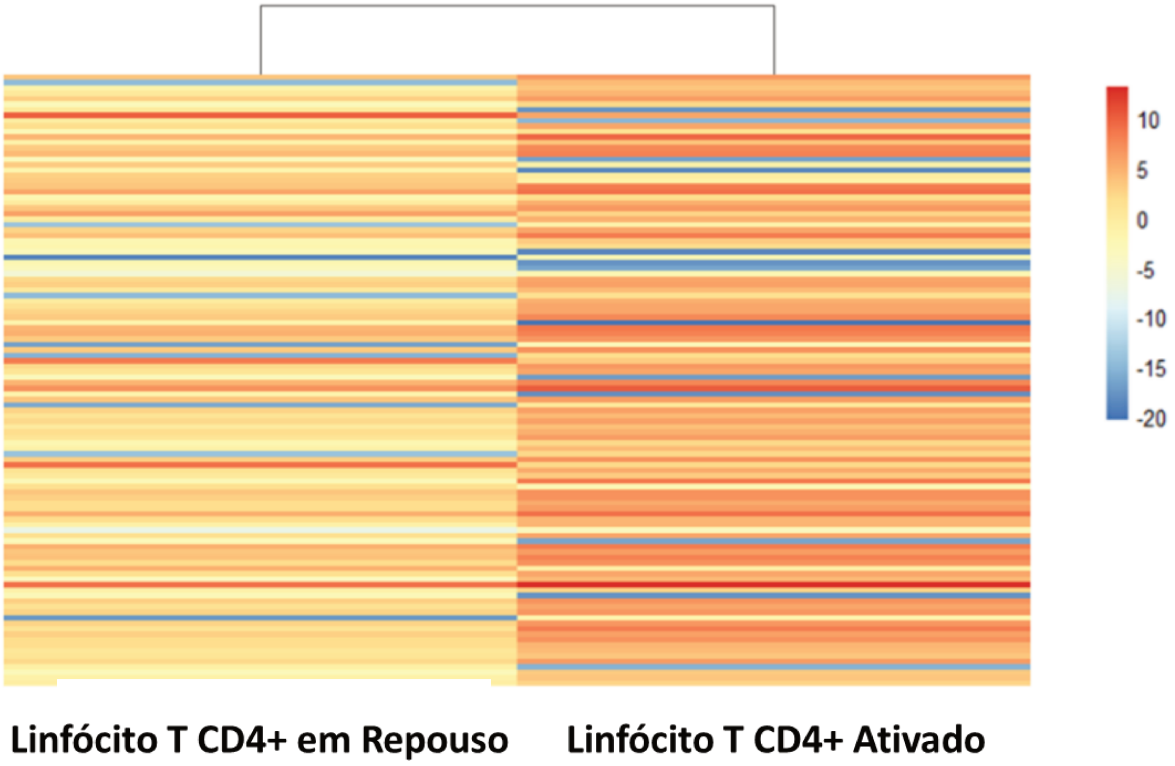


Figura 08. Heatmap dos transcritos diferencialmente expressos nas células T CD4+ em repouso e após ativação. Embora haja inúmeros transcritos expressos nos dois estados de ativação do linfócito, é plausível que as diferenças fenotípicas sejam devidas a expressão diferencial destes transcritos.

Com o intuito de identificar os transcritos de importância biológica num contexto funcional, as isoformas que possuíam expressão diferencial nos estados de repouso e ativação dos linfócitos T CD4+ apresentadas na figura anterior foram separadas ontologicamente (Figura 08). Isolamos também os transcritos que continham retenção intrônica de forma independente do estado de ativação celular em outra lista, para estratificar a ocorrência deste tipo de *splicing* alternativo de acordo com os processos biológicos essenciais à economia da célula (Figura 09). Este exercício de classificação ontológica permitiu identificar um viés de participação em processos metabólicos, tanto dos transcritos diferencialmente expressos de maneira dependente do estado de ativação (independentemente de ter ou não RI), quanto daqueles que apresentam retenção intrônica (independente do estado de ativação do linfócito). Embora trate-se de uma análise qualitativa bem genérica, a mesma sugere a possibilidade de que um determinado gene possa se expressar através de diferentes transcritos para o desempenho de uma função definida, sendo alguns desses transcritos completamente processados e ativos, enquanto aqueles com RI poderiam simplesmente representar um ruído *(background)* transcricional. Todavia, contra esta

interpretação, foi registrada uma maior prevalência de transcritos com retenção intrônica no subgrupo ontológico associado ao processamento de RNA.

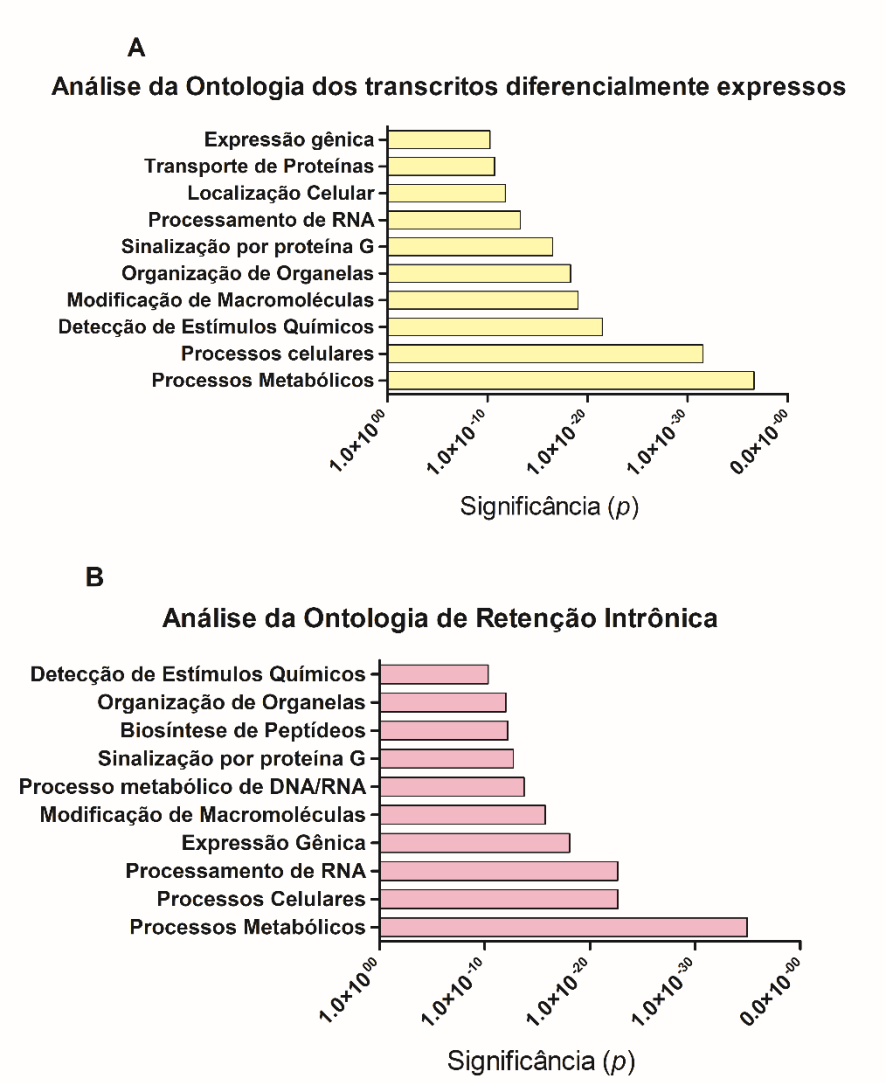


Figura 09. Análise ontológica de genes em processos biológicos. **A**: Genes diferencialmente expressos em linfócitos ativados. **B**: Genes contendo retenção intrônica. A maioria dos genes estão envolvidos em processos metabólicos e celulares, contemplando uma sobreposição da análise ontológica dos genes em **A** e **B** da figura. Nota-se uma maior representatividade de genes possuidores de retenção intrônica relacionados com o metabolismo, bem como aqueles relacionados com o processamento de RNA, quando comparado com a mesma análise feita com os transcritos diferencialmente expressos.

5.3 Estudo do ruído (*background*) transcricional

A fim de estudar o ruído transcricional relacionado com a expressão da retenção intrônica, somamos todas as isoformas transcricionais com retenção intrônica de cada gene em

um único valor, que denominamos de *expressão intrônica*. Ruído transcricional nesta dissertação, refere-se ao produto de transcrição sem nenhuma consequência biológica produzido por consequência natural da expressão gênica (o que, por sua vez, representa o valor de expressão global de todos os transcritos de um gene). Um ruído transcricional seria evidente, se quando a expressão gênica estivesse regulada positivamente ou negativamente durante os estados fenotípicos da ativação do linfócito, a expressão intrônica a acompanhasse. Pensando nisto, utilizamos o *fold change* da expressão gênica e da expressão intrônica para verificar a existência de uma correlação. Desta forma, então, foi possível indagar se o aumento ou diminuição da expressão das isoformas transcricionais do gene (expressão gênica) também estaria diretamente relacionado com as isoformas transcricionais intrônicas (expressão intrônica). A partir dos dados de RNA-seq, fizemos um teste de correlação de Spearman – onde 1 e -1 seriam os extremos positivo e negativo e zero o mínimo de correlação –, excluindo expressões de 0 RPKM *(reads per kilobase million)*.

A princípio, esta análise revelou não haver uma correlação forte entre expressão gênica e expressão intrônica (r = 0.6) (Figura 10). Essa descoberta nos permite afirmar a existência de um desacoplamento de parte dos transcritos com RI da expressão gênica, o que derruba a suposição de que a retenção intrônica seja completamente ligada a uma ineficiência de maquinaria de *splicing*.

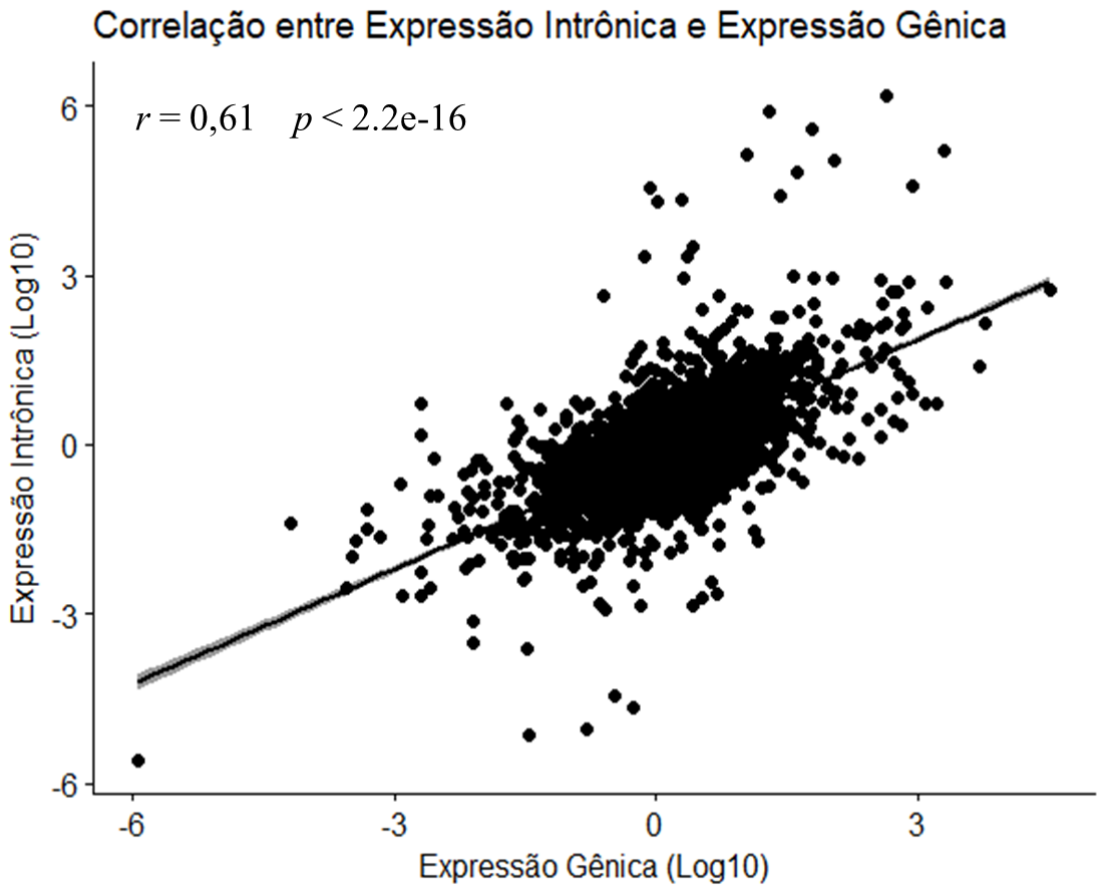


Figura 10. Expressão gênica *vs* expressão intrônica. O teste de Spearman foi utilizado para correlacionar a expressão gênica com a regulação da expressão intrônica no linfócito T CD4+. Não foi observada uma correlação forte entre as duas expressões.

Embora o resultado anterior mostre uma correlação moderada, em que o nível de retenção intrônica sugira que uma fração de produtos transcricionais não tenha uma relevância biológica maior, o fato de não evidenciarmos um paralelismo forte entre a retenção intrônica e o volume global de transcrição não exclui que, para certos genes, a retenção intrônica seja de fato fortemente acoplada a um ruído que sobe e desce com a transcrição. Tal raciocínio não seria surpreendente se considerarmos a diversidade do transcriptoma humano e a quantidade de genes expressos sob diferentes tipos de regulação. Para evidenciarmos estes genes, calculamos a razão da expressão intrônica entre as células em repouso e ativadas (expressão intrônica da célula em repouso ÷ expressão intrônica da célula ativada). Este exercício permitiu identificar os 10 genes mais regulados positivamente e os 10 genes mais regulados negativamente quanto à sua expressão intrônica. Uma vez identificados os extremos da curva de distribuição no que diz respeito à retenção intrônica, refizemos o teste de Spearman, o qual nos permitiu visualizar uma forte correlação entre as variáveis comparadas (Figura 11). Este resultado nos revela que

a transcrição de alguns grupos de genes é acompanhada de um ruído de expressão intrônica com uma forte correlação ($r = 0.9$).

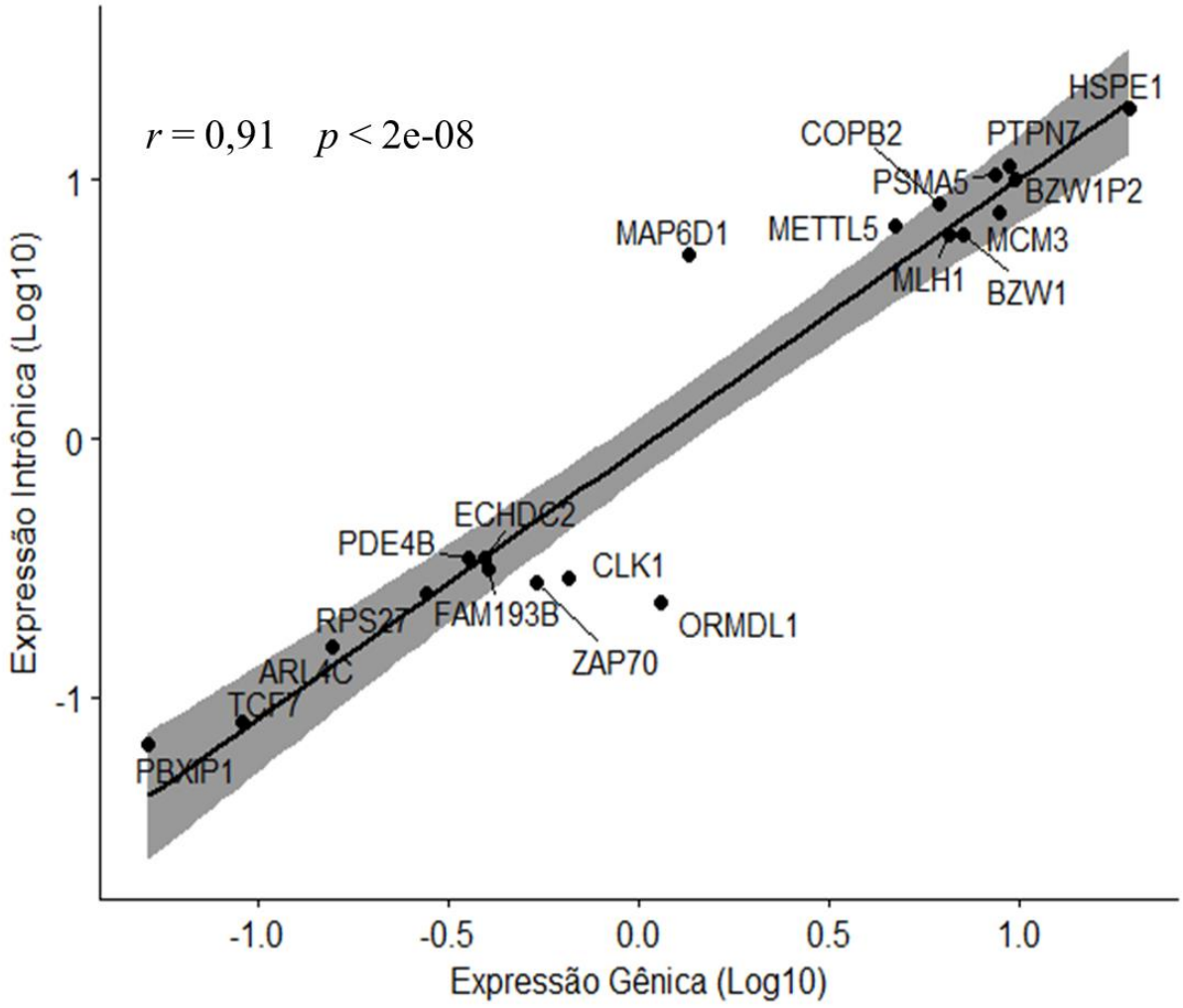


Figura 11. Correlação entre expressão gênica e intrônica dos genes mais afetados pela mudança de expressão intrônica. Quando analisados os genes que apresentam uma mudança extrema de expressão gênica entre os fenótipos de ativação do linfócito, observa-se uma alta correlação positiva com a expressão intrônica.

Observamos que os genes que sofrem uma drástica mudança de expressão gênica são naturalmente acompanhados por uma alteração de expressão de transcritos com RI, com alta correlação positiva. Entretanto, quando analisamos o contexto global, percebemos a ocorrência de uma correlação moderada entre a expressão gênica e a expressão intrônica, o que nos permite especular sobre a existência de uma força moduladora sobre determinados grupos de genes. Ademais, resta a ser estabelecido se a modulação da retenção intrônica é determinante ou consequência do fenótipo celular.

5.4 Regulação fina da retenção intrônica

Considerando que parte dos transcritos com retenção intrônica não estão diretamente ligados ao aumento ou diminuição da expressão gênica global, como demonstrado na figura 10, especulamos que os mesmo devam sofrer algum tipo de regulação fina do processo de *splicing,* para seleção das variantes com RI. Na tentativa de identificar os genes mais "desacoplados" e que estivessem excluídos do ruído transcricional, elaboramos uma sublista de genes da qual foram excluídas as isoformas transcricionais com retenção intrônica de RPKM < 10.

Entendemos que a totalidade dos transcritos com RI de cada gene tem o potencial de causar impacto na funcionalidade da célula. Então, somamos os valores de RPKM de todos os transcritos com RI de um mesmo gene para gerar novamente o que chamamos de expressão intrônica. Filtramos, em seguida, os genes de acordo com sua presença no estado de repouso ou ativação do linfócito. É importante considerarmos aqui a expressão intrônica em contexto com a expressão gênica, pois assim encontraremos quais genes podem estar sendo modulados durante a transcrição. Para isso, calculamos o percentual de RI relativo à sua expressão gênica e filtramos os genes em que a RI correspondesse a > 20% da transcrição. A escolha do *cut-off* de 20% se justifica, porque há precedente na literatura que demonstra que este nível de RI pode gerar uma degradação por NMD comparável àquela obtida de um silenciamento funcional por RNAi (TRAVELLA, 2006). De posse destas sublistas, calculamos o que definimos como *"Índice de Desacoplamento Transcricional"* ou IDT, que reflete a incidência relativa de RI associada ao estado de ativação do linfócito. O IDT foi determinado para os genes de cada sublista em se dividindo o valor percentual de RI por aquele correspondente do estado de ativação alternativo (%RI-repouso ÷ %RI-ativação ou %RI-ativação ÷ %RI-repouso).

As tabelas 4 e 5 nos permitem identificar genes claramente desacoplados, alguns deles de importância já conhecida na fisiologia do linfócito e para a célula em geral. Notam-se quatro padrões de desacoplamento: (i) expressão gênica similar nos dois estados de ativação e queda do % de RI durante a ativação (HMGN3 e ORMDL1); (ii) expressão gênica similar nos dois estados de ativação e aumento do % de RI durante a ativação (GAS5); (iii) expressão gênica aumentada na ativação com queda do % de RI na ativação (POLR2B, SFPQ, PSMB4, RER1, SP140 e MRPS25) e (iv) expressão gênica aumentada na ativação com aumento desproporcional do % de RI na ativação (IQCG, ADAM15, FASTKD3, EXO1, UBE2F, NIF3L1 e MRPL22).

Identificamos casos extremos de desacoplamento, o que sugere fortemente que os transcritos com retenção intrônica possam ser regulados de forma independente da transcrição global do gene pela maquinaria de *splicing*. Assim, o gene HMGN3 apresenta expressão global pouco afetada pela ativação, mas ganha um IDT de 16, ou seja, sua expressão intrônica cai drasticamente (cerca de 16 vezes) na ativação. Já o gene POLR2B aumenta cerca 3,5 vezes sua expressão gênica e inversamente reduz cerca de 8 vezes o % de RI na ativação. Finalmente, o gene EXO1 aumenta 20 vezes com a ativação linfocitária, mas o incremento na % de RI é desproporcionalmente modesto, com um IDT de 3,78.

Tabela 4 – Genes com alto IDT na célula em repouso (RPKM)

Nome do gene	Expressão intrônica – Célula em repouso	Expressão intrônica – Célula ativada	Expressão gênica – Célula em repouso	Expressão gênica – Célula ativada	% da expressão intrônica do gene – Célula em repouso	% da expressão intrônica do gene – Célula ativada	Razão IDT
HMGN3	33,89	2,69	37,16	47,76	91,21%	5,64%	16,16
POLR2B	17,68	7,81	73,44	258,80	24,07%	3,021%	7,96
SFPQ	90,14	46,49	304,04	1079,84	29,64%	4,30%	6,88
PSMB4	65,48	47,79	115,02	428,81	56,93%	11,14%	5,10
LDLRAP1	17,93	0,84	55,02	13,11	32,59%	6,43%	5,06
ORMDL1	56,33	13,12	110,13	125,26	51,14%	10,47%	4,88
RER1	12,11	11,57	45,34	207,55	26,71%	5,57%	4,78
SP140	18,13	9,66	27,96	69,39	64,84%	13,93%	4,65
MRPS25	13,18	12,61	18,06	79,62	72,96%	15,83%	4,60
SNRPE	10,19	15,36	20,52	138,61	49,64%	11,08%	4,47

Tabela 5 – Genes com alto IDT na célula ativada (RPKM)

Nome do gene	Expressão intrônica – Célula em repouso	Expressão intrônica – Célula ativada	Expressão gênica – Célula em repouso	Expressão gênica – Célula ativada	% da expressão intrônica no gene – Célula em repouso	% da expressão intrônica no gene – Célula ativada	Razão IDT
IQCG	0,79	14,50	6,15	19,23	12,84%	75,42%	5,87
ADAM15	3,14	28,06	22,77	47,43	13,80%	59,16%	4,28
GAS5	62,78	351,49	447,66	593,47	14,02%	59,22%	4,22
FASTKD3	0,25	11,13	2,00	22,31	12,73%	49,87%	3,91
EXO1	0,36	28,85	1,47	30,50	25,00%	94,58%	3,78
RUNX3	5,89	43,73	66,99	149,04	8,80%	29,34%	3,33
UBE2F	1,58	25,44	16,08	78,93	9,88%	32,23%	3,26
WDR41	1,21	10,85	8,57	23,70	14,20%	45,78%	3,22
NIF3L1	1,13	15,92	6,45	33,05	17,53%	48,18%	2,74
MRPL22	3,42	56,60	19,46	118,50	17,59%	47,76%	2,71

Os resultados apresentados até o presente momento se referem aos transcritos mais frequentes que são compartilhados, embora não necessariamente de forma igualitária, pelos dois estados de ativação do linfócito. Há, todavia, eventos transcricionais de cada tipo de *splicing* alternativo que são exclusivos de cada fenótipo. No intuito de melhor dissecar a regulação fina da RI, usamos nossa ferramenta algorítmica para selecionar transcritos com retenção intrônica, cujas isoformas fossem exclusivas de um dos fenótipos estudados. Descobrimos uma janela de variação em RPKM na transcrição das isoformas exclusivas identificadas que varia de 6-208 RPKM (mediana de 14,16 e média de 28,45). Para nossa surpresa, encontramos alguns transcritos com retenção intrônica de alta representatividade em relação a expressão gênica, ultrapassando 20% em vários casos (Figura 12). Vale ressaltar que também foi possível identificar a introdução de um código prematuro de terminação (CPT) em múltiplas ocasiões, o que nos permite dizer que o transcrito tem grandes chances de ter um destino para degradação via NMD.

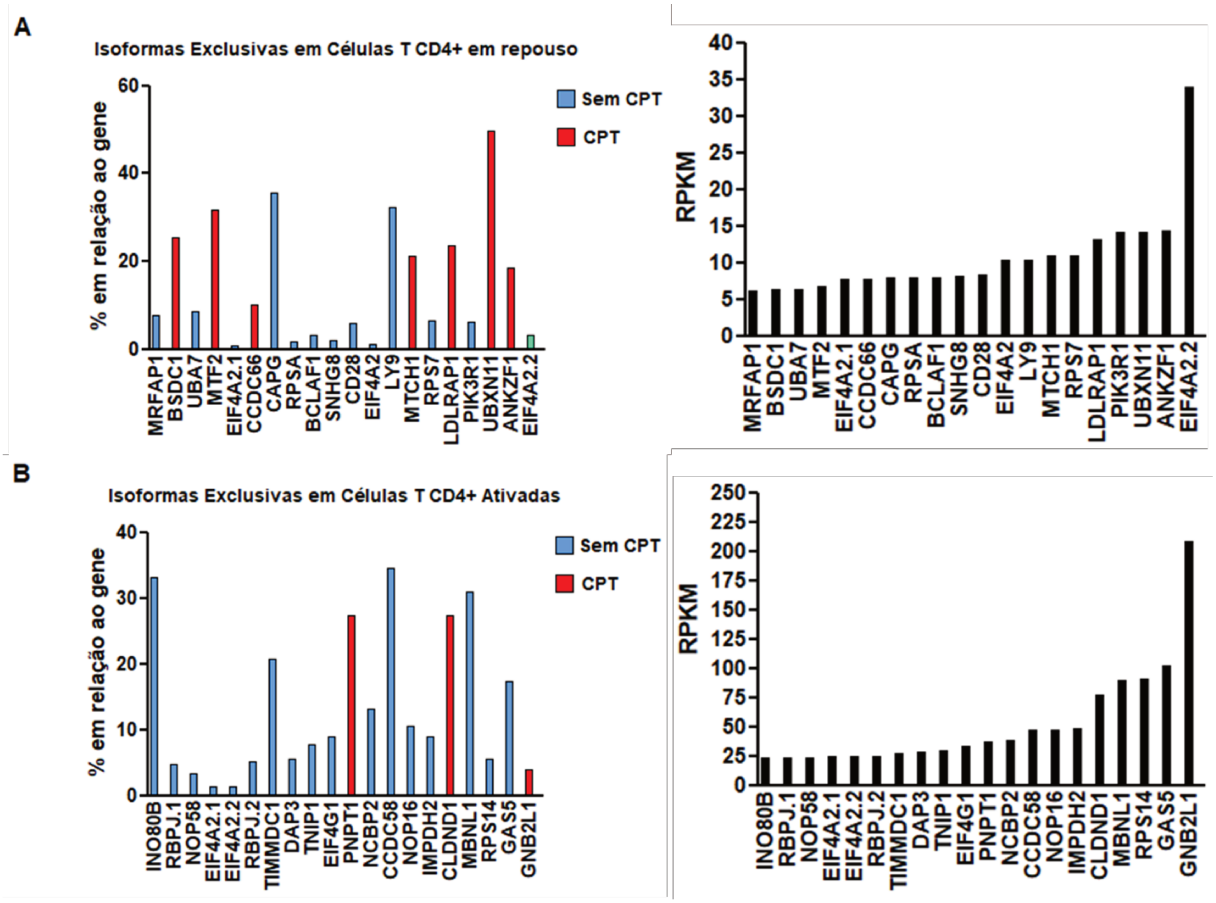


Figura 12. Isoformas transcricionais com retenção intrônica exclusivas. **A**: Isoformas exclusivas de células T CD4+ em repouso. **B**: Isoformas exclusivas de células T CD4+ ativadas. CPT: Código prematuro de terminação

Outra evidência que suporta a existência de uma regulação fina da retenção intrônica é a seletividade estrutural dos transcritos identificados. Em princípio, a célula teria a possibilidade de reter qualquer ou quaisquer íntron(s) durante o processamento do *splicing* alternativo. Todavia, observamos que frequentemente o(s) mesmo(s) íntron(s) são preservados nos transcritos. Há, portanto, predominância de certas isoformas, cuja expressão gênica não fica diluída na diversidade potencial oferecida pelas inúmeras combinações distintas de retenção intrônica de que as variantes podem usufruir.

5.5 Genes com alta expressão de retenção intrônica independentes do fenótipo

Nas seções anteriores relatamos os experimentos realizados para testar nossa hipótese de trabalho, os quais revelaram uma tendência moderada de associação entre retenção intrônica e volume global de transcrição. Todavia, esta tendência não é absoluta, sendo compatível com uma modulação ativa de *splicing* alternativo por RI de certos genes no trascriptoma dos linfócitos. Para complementar o cenário apenas descrito, buscamos encontrar quais transcritos com RI aparecem de maneira independente do fenótipo. Fizemos a primeira incursão neste sentido com o experimento relatado na figura 7, no qual procuramos identificar o número de eventos transcricionais (isoformas) de diferentes tipos de *splicing* alternativo segregados de acordo com o estado de ativação. Entretanto, aquela análise focou na ocorrência do evento e não no nível de expressão associado. Reabordamos a questão, desta vez de forma quantitativa, somando todos os transcritos com retenção intrônica de cada gene, para identificar quais os genes eram mais impactados com a retenção intrônica. A tabela 6 mostra os genes com os 10 maiores RPKMs de expressão intrônica. Com exceção da IL-2, que é expressa de maneira funcional apenas no linfócito ativado, todos os outros genes apresentaram alta expressão em RPKM de expressão intrônica em ambos estados de ativação (>50). Em determinados casos, alguns genes contêm uma expressão intrônica que representa a quase totalidade da expressão gênica (dados não mostrados) e, de maneira inesperada, encontramos genes fortemente associados à função do linfócito T, como se pode ver na tabela 6.

Tabela 6 – Genes com alta expressão intrônica independente do fenótipo de ativação

Nome do gene	Função	RPKM – Célula em repouso	RPKM – Célula ativada
(GNB2L1) RACK1	Metabolismo (PKC) e ciclo celular	935,47	1336,35
IL-2	Proliferação	2,18	1100,59
SH3BGRL3	Importante papel na via redox	339,60	447,86

TRAF3IP3	Media crescimento celular (Jun Kinase)	262,51	72,18
CSDE1	Repressão transcricional de c-myc	136,27	331,02
RHOA	Organização da actina	128,46	298,06
CD74	Envolvida no MHCII	204,59	218,15
FYN	Remodelação do citoesqueleto e Ativação do linfócito T	123,28	210,48
HLA-B	Envolvido no MHCI	162,86	150,26
CCNL1	Envolvido com CDKs e *splicing* alternativo	199,79	104,80

5.6 Validação experimental dos resultados

Após o exercício analítico do perfil de retenção intrônica no linfócito T CD4+ com o uso das ferramentas algorítmicas aqui desenvolvidas, avaliamos se os dados gerados na figura 12 e tabela 6 do RNA-seq tinham algum valor preditivo em diferentes doadores. Isolamos linfócitos T CD4+, que foram ativados (ou não) com anticorpos anti-CD3/anti-CD28 durante 18 horas. A confirmação da qualidade das amostras foi efetuada por citometria de fluxo, em se estimando a positividade do marcador CD4 e em se avaliando o estado de ativação linfocitária com o marcador CD25. Em seguida, extraímos o RNA para síntese de cDNA e subsequente amplificação por *nested*-PCR, com *primers* especificos para os genes selecionados.

Os testes feitos com o grupo de genes exclusivos da célula em repouso e da célula ativada em sua maioria identificaram as isoformas transcricionais previstas *in sílico,* seja pelo correto tamanho do *amplicon,* seja pelo direto sequenciamento das bandas amplificadas. GAS5 e LY9 mostraram concordância total, com exclusividade de expressão do transcrito com RI, condicionada pelo estado de ativação celular em 4 de 5 doadores para cada gene. O gene LY9 continha mais de uma isoforma exclusiva do fenótipo de repouso na análise. Esperava-se um *amplicon* na altura de 345 pb (TCONS_00013304) e possivelmente uma outra isoforma (TCONS_00013296) (ANEXO B – F), expressa em menor quantidade (0,3 RPKM) de 387 pb.

De acordo com a previsão *in sílico,* identificamos exclusividade de expressão do transcrito com RI em 2 de 7 doadores para o UBXN11 e em 3 de 7 doadores para o TIMMDC1. Entretanto, identificamos também alguma promiscuidade de expressão das isoformas com RI no que diz respeito ao estado de ativação do linfócito, que parece ocorrer de maneira dependente do doador em 4 de 7 indivíduos para o TIMMDC1 e em 5 de 7 indivíduos para o ANKFZ1 (Figura 13).

Não houve amplificação especifica para o gene PIK3R1, pois esperávamos amplificar uma isoforma pequena (TCONS_00244844) (ANEXO A e B - J), com apenas 148 pb (Figura 13C).

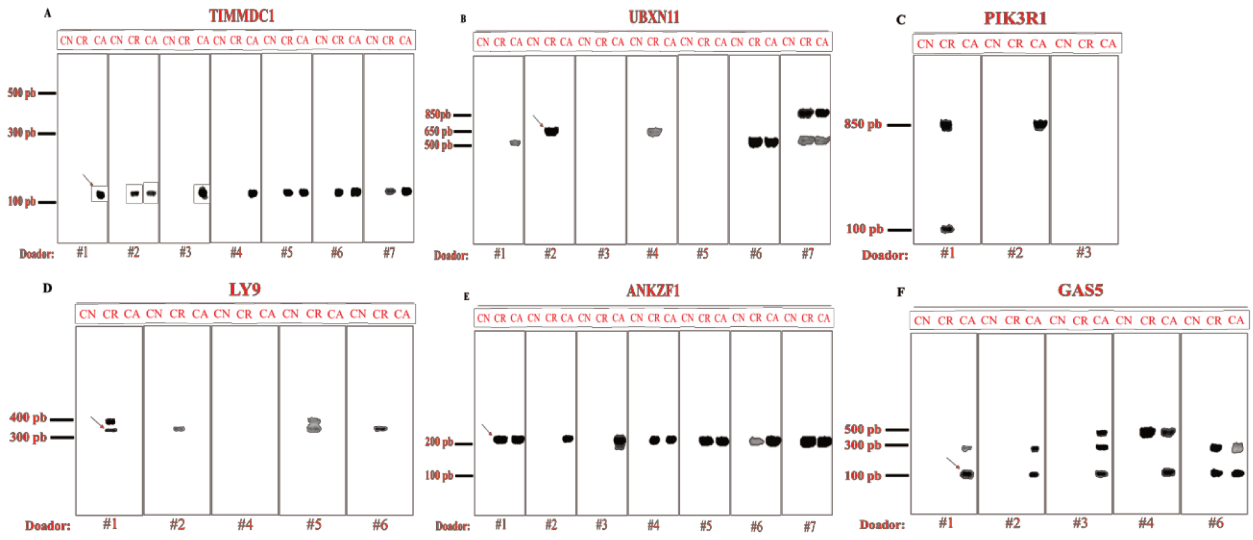


Figura 13. Eletroforese em gel de agarose 2% de diferentes genes do grupo de isoformas exclusivas. **A**: Eletroforese do gene TIMMDC1. **B**: Eletroforese do gene UBXN11. **C**: Eletroforese do gene PIK3R1. **D**: Eletroforese do gene LY9. **E**: Eletroforese do gene ANKFZ1. **F**: Eletroforese do gene GAS5. CN: controle negativo sem cDNA do teste *nested*-PCR. CR: célula em repouso. CA: célula ativada. Setas indicam o tamanho esperado do transcrito com retenção intrônica.

Da mesma forma, avaliou-se um grupo de cinco genes que, a princípio, deveriam ter alta expressão de transcritos com retenção intrônica tanto na célula ativada quando na célula em repouso (Figura 14). A transcrição de três deles seguiu o padrão previsto pela análise *in sílico,* com concordância total em 6/6 doadores para SH3BRGL3, em 3/4 doadores para FYN e em 3/6 doadores para HLA-B. No caso de CCNL1 e RHOA, não houve concordância, mas a sequência intrônica retida se localiza bem próxima do extremo 5’ dos transcritos, o que pode ter comprometido a capacidade de nosso algoritmo delimitar com precisão o início do íntron e, por consequência, o planejamento de nossa estratégia de amplificação.

Uma análise global dos resultados de validação experimental dos dados obtidos *in sílico* permite corroborar os seguintes pontos: (i) existência dos transcritos com retenção intrônica previstos *in sílico* através de identificação por tamanho e sequenciamento das variantes; (ii) exclusividade de expressão de isoformas com retenção intrônica de certos genes que segregam sua presença num estado de ativação celular específico; (iii) existência de genes

com alta expressão de variantes com retenção intrônica cuja presença independe do estado de ativação do linfócito e (iv) possibilidade de plasticidade interindividual dos mecanismos de *splicing* alternativo por retenção intrônica.

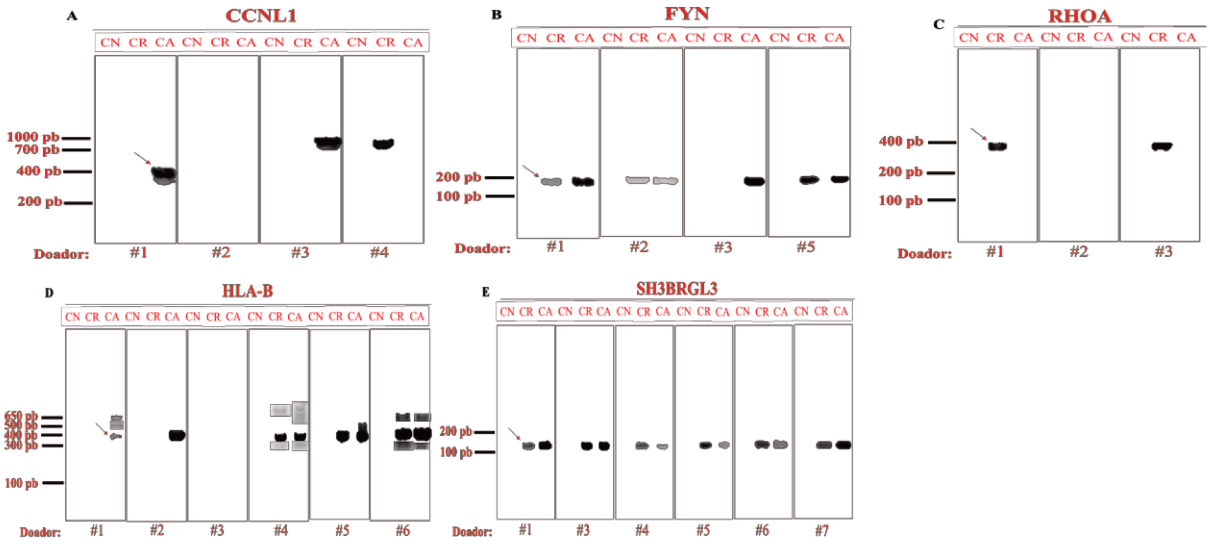


Figura 14. Eletroforese em gel de agarose 2% de diferentes genes do grupo de isoformas com alta expressão independentes do fenótipo. **A**: Eletroforese do gene CCNL1. **B**: Eletroforese do gene FYN. **C**: Eletroforese do gene RHOA. **D**: Eletroforese do gene HLA-B. **E**: Eletroforese do gene SH3BRGL3. CN: controle negativo sem cDNA do teste *nested*-PCR. CR: célula em repouso. CA: célula ativada. Setas indicam o tamanho esperado do transcrito com retenção intrônica.

6. DISCUSSÃO

A produção de novas proteínas em resposta a estímulos do microambiente é geralmente realizada através da transcrição, sendo este um processo usualmente rápido (SCHWARZL e colab., 2015). A transcrição gera um pré-mRNA que permite à célula ter um repertório de possíveis isoformas de mRNA com diferentes combinações de cassetes de éxons (e possivelmente também de íntrons) através do processo que chamamos de *splicing* alternativo. Para que este processo ocorra, há o envolvimento das várias proteínas que compõem o splisseossomo, o qual excisa os cassetes indesejados de maneira muito bem refinada.

A retenção intrônica é um tipo emergente de *splicing* alternativo que ainda não é muito bem compreendido quanto à sua funcionalidade, mas já é considerada a forma mais

comum de SA em plantas e em eucariotos (BRAUNSCHWEIG e colab., 2014). É possível ver o mesmo padrão de frequência nos resultados de *splicing* alternativo encontrados na célula T CD4+, sendo também o evento de *splicing* mais comum nessas células (Figura 07).

Os estudos mais detalhados sobre retenção intrônica existentes na literatura usaram a diferenciação celular como modelo. O primeiro trabalho que dissecou este fenômeno mostrou diferentes padrões de RI em cada uma das diferentes etapas da granulopoiese, sugerindo um papel importante na fisiologia deste processo (WONG e colab., 2013). Essa sugestão foi reforçada em estudos posteriores da eritropoiese (PIMENTEL e colab., 2016) e na espermatogênese (NARO e colab., 2017). Transcritos com RI poderiam estar reduzindo os níveis de expressão proteica menos importantes para a fisiologia da célula quando necessário (BRAUNSCHWEIG e colab., 2014). Estes estudos apontam para uma maior complexidade da dinâmica de transcrição e da regulação pós-transcricional do nosso transcriptoma do que pensado anteriormente.

A ativação do linfócito T é um processo muito importante para o sistema imune, em que as células em repouso se diferenciam e passam a proliferar e produzir moléculas efetoras, tais como citocinas (SMITH-GARVIN e colab., 2009). É sabido que as marcas epigenéticas entre estes dois estados do linfócito são semelhantes, sem exibir quase nenhuma diferença a nível de cromatina (BARSKI e colab., 2009; CUDDAPAH e colab., 2010). A mudança de expressão gênica do linfócito T (Figura 08) é consequência da sinalização mediada pela Proteína Ativadora-1 (AP-1), pelo Fator Nuclear Ativador das Células T (NFAT), pelo Fator Nuclear kappa B (NF-kB) e pela sinalização autócrina da interleucina 2 (IL-2) (MACIÁN e colab., 2001; NISHIMURA e colab., 2002).

É fato que a distribuição ontológica dos transcritos com RI identificada nas células T CD4+ é semelhante àquela encontrada no estudo da eritropoiese (PIMENTEL e colab., 2016) e da espermatogênese (NARO e colab., 2017). Estes eventos envolvem grandes números de genes responsáveis pelo processamento de RNA, pelos processos celulares (ex: organização do citoesqueleto) e pela organização de organelas. Este padrão de participação da RI na transcrição gênica diferencia-se apenas no que diz respeito às sinalizações específicas do tecido. Até mesmo doenças como o câncer mantém o perfil de utilização da RI nos processos biológicos como, por exemplo, o de processamento de RNA (DVINGE e BRADLEY, 2015). Os resultados apresentados na figura 9 revelam que os processos biológicos associados com transcritos diferencialmente expressos no linfócito T CD4+, no que tange ao estado de ativação, são os mesmos processos biológicos mais relacionados com a RI. Embora esta tenha sido uma análise

qualitativa, ela era indicativa de que a RI fosse um ruído transcricional que acompanhava a oscilação da expressão gênica.

Com o intuito de abordar essa importante possibilidade, consideramos a expressão intrônica (soma de todos os transcritos com retenção intrônica do mesmo gene) e a comparamos com a expressão gênica (soma de todos transcritos do mesmo gene). Nós mostramos a existência de uma alta correlação entre RI e expressão gênica para os genes que aumentam ou diminuem nos extremos da curva gaussiana de expressão global (Figura 11). Por outro lado, quando fizemos a mesma análise com todos os genes, verificamos uma correlação moderada da expressão intrônica com a expressão gênica (Figura 10).

Portanto, nossos dados indicam que parte dos transcritos com RI podem estar 'desacoplados' da expressão gênica, podendo ser diferencialmente regulados. De forma interessante, um índice de RI calculado a partir das regiões intrônicas e exônicas compartilhadas por diferentes transcritos do mesmo gene mostrou-se reduzido no linfócito T CD4+ ativado, quando comparado à célula em repouso (NI, Ting e colab., 2016). Diante deste fato, um dado transcrito poderia estar sendo pré-selecionado para servir de sentinela capaz de sofrer rapidamente a excisão do bloco intrônico após uma estimulação, como já observado na atividade neuronal (MAUGER e colab., 2016).

Embora na análise global realizada por Ni e colaboradores (NI, Ting e colab., 2016) pareça que haja uma menor expressão dos transcritos com retenção intrônica na célula ativada, os nossos dados mostram que existem vários genes que aumentam a expressão intrônica na célula ativada quando comparado à célula em repouso, assim como o inverso (Tabela 4 e 5). Desta maneira, os genes identificados com o maior IDT (Índice de desacoplamento transcricional) nas tabelas 4 e 5 seriam passíveis de sofrer uma mudança da expressão intrônica de maneira dependente do fenótipo. Vale notar que em alguns genes não se observa alteração na expressão gênica global entre os fenótipos, mas há uma mudança drástica na expressão intrônica. Isso nos faz pensar que haja uma regulação seletiva de isoformas intrônicas.

Uma regulação seletiva explicaria os padrões de isoformas transcricionais com RI específicos, que já foram evidenciados em tecidos de mamífero (BRAUNSCHWEIG e colab., 2014) e em diferentes tipos de câncer (DVINGE e BRADLEY, 2015). Um estudo com células musculares diferenciadas sugeriu a existência de uma maquinaria de *splicing* adaptada ao tecido, capaz de aumentar os transcritos com RI sem que isto constituísse um defeito de

transcrição (LLORIAN e colab., 2016). Torna-se, assim, plausível a possibilidade do uso dessas isoformas como biomarcadores, para determinar fenótipos celulares.

Quando começamos nosso estudo, partimos da premissa de que era possível a existência de transcritos específicos de células T em repouso e de células T ativadas. Abordamos esse ponto, selecionando transcritos com RI exclusivos de cada estado celular com nosso código computacional. Os resultados de tal exercício sugerem que a célula tem um viés de escolha de qual íntron será retido para a transcrição de cada gene, de forma frequentemente dependente do estado de ativação celular (ANEXO A). Identificamos que alguns dos transcritos tinham boa representatividade em relação a expressão gênica (Figura 11) e apresentavam código prematuro de terminação, sugerindo uma regulação negativa através da degradação pela via NMD (BUCKLEY e colab., 2014).

Validamos os resultados gerados *in silico,* usando *nested*-PCR e visualizando os *amplicons* em eletroforese. Podemos dizer que estas isoformas transcricionais encontradas na análise do RNA-seq pela nossa ferramenta realmente existem (Figura 13 e Figura 14). Nas análises dos transcritos presentes com exclusividade nos fenótipos estudados, encontramos concordância em 4 de 5 doadores nos genes LY9 e GAS5 (Figura 13D e Figura 13F). LY9 é um gene que acredita-se regular negativamente a ativação do linfócito (MARTIN e colab., 2005; SINTES e colab., 2007) e inibir a secreção de interferon-γ (IFN- γ) (DE SALORT e colab., 2013). Já o GAS5 é um RNA não codificador, que se liga ao receptor de glicocorticoides, mimetizando a região do DNA que contém um elemento responsivo ao glucocorticoide (GRE). GAS5 inibe assim o GRE permitindo a sobrevivência celular durante a escassez de nutrientes (KINO e colab., 2010). A isoforma do gene LY9 é amplificada apenas na célula em repouso, enquanto a isoforma do gene GAS5 é amplificada no fenótipo ativado em 4 de 5 doadores. Ressalta-se que o GAS5 é um dos genes com alto IDT visto na tabela 5. Ele apresenta um nível similar de expressão gênica nos dois fenótipos celulares, mas há um aumento substancial da expressão intrônica na célula ativada. O perfil de expressão desses dois genes mostra a existência de isoformas específicas claramente ligadas ao fenótipo de ativação do linfócito.

Havíamos previsto transcritos com retenção intrônica compartilhados entre os dois fenótipos estudados do linfócito T CD4+ (Figura 07). Para enriquecer a informação sobre o cenário descrito até então, procuramos identificar os genes que exibiam a maior parte da expressão gênica feita por transcritos com RI de forma independente da ativação do linfócito. Análise experimental do gene SH3BRGL3 teve uma concordância total com os dados *in silico* (6/6 dos doadores) (Figura 14E). O gene FYN que é importante para ativação do linfócito T

(SMITH-GARVIN e colab., 2009) mostrou também boa concordância (3/4 dos doadores) (Figura 14B). HLA-B é um importante gene envolvido na apresentação de antígenos no sistema imune (EZQUERRA e colab., 1985); sua isoforma com retenção intrônica foi encontrada e está em parcial conformidade com a predição *in silico*, tendo sido amplificada no tamanho esperado de 367 pb (3/6 dos doadores) (TCONS_00274854, ANEXO B - H). No caso do HLA-B, houve também a amplificação de uma isoforma de 493 pb, identificada como AL671883.2 no banco de dados da ENSEMBL que julgamos ser um exitron, bem como de uma outra isoforma não identificada.

Com a exceção do gene da IL-2, aqueles listados na tabela 6 deveriam ser pouco susceptíveis a um suposto efeito sobre a RI desencadeado pelos sinais um e dois gerados pelo TCR e CD28, respectivamente. Este padrão de expressão intrônica do linfócito anteriormente descrito gera uma questão intrigante: por que a célula manteria uma alta expressão de certos transcritos "defeituosos" ou "imaturos", de maneira dissociada de seu fenótipo de ativação? Não faria muito sentido para a economia energética da célula fazê-lo, a menos que esses transcritos fossem traduzidos ou estivessem prontos para sofrer re-*splicing*. Embora esta última questão extrapole o propósito da tese, é plausível que os genes identificados representem ótimos alvos para estudos funcionais e elucidação da importância deste tipo de *splicing* alternativo na fisiologia do linfócito. FYN, por exemplo, tem uma enorme participação no processo da ativação dessa célula.

Alguns genes mostraram ter uma plasticidade interindividual no mecanismo de retenção intrônica dependente do estado de ativação do linfócito. A análise *in silico* prevê uma isoforma com RI do gene TIMMDC1, que codifica para uma chaperona envolvida na montagem mitocondrial do complexo I com grande importância na cadeia respiratória (ANDREWS e colab., 2013), cuja existência foi confirmada em 3 de 7 dos doadores na célula ativada, como esperado. Entretanto, 4 dos 7 indivíduos testados também apresentaram esta isoforma na célula em repouso (Figura 13A). Esse padrão repete-se no gene ANKFZ1, um gene de resposta ao peróxido de hidrogênio, importante para a manutenção da integridade da mitocôndria em condições de estresse celular (VAN HAAFTEN-VISSER e colab., 2017) (Figura 13E).

O processo de escolha dos transcritos pela célula é complicado, sendo estimado que cerca de 95% dos genes sofram *splicing* alternativo em seres humanos (PAN e colab., 2008; WANG, Eric T e colab., 2008), com a expressão variável conforme o tecido (YEO e colab., 2004). O projeto ENCODE avaliou 15 linhagens celulares com sequenciamento de alta profundidade, revelando que a expressão de um gene é feita em média por 12 isoformas por

gene, que não são equalizadas. Existe uma isoforma dominante que tipicamente representa 30% da expressão do gene, como vimos para alguns genes descritos neste trabalho (Figura 12) (DJEBALI e colab., 2012). Todavia, a complexidade por trás do *splicing* alternativo por retenção intrônica vai muito além. Em *S. cerevisiae* e outras leveduras, todos os íntrons catalisados pela U2 contêm sequências terminais canônicas GT-AG (BON e colab., 2003). Os sítios de *splice* desses íntrons são bastante conservados, com motivos de sequências específicos. Já em humanos, a maioria dos íntrons catalisados pela U2 são bastante degenerados, muitas vezes não tendo informação suficiente para sinalizar o 5'ss, o sítio de ramificação ou o 3'ss, sobretudo em longos íntrons que contêm vários elementos semelhantes (SHEPARD e HERTEL, 2009; VALCÁRCEL e GREEN, 1996). A título ilustrativo, se compararmos dois íntrons aleatórios de S. *cerevisiae*, haveria 58% de chance de que eles exibissem os mesmos 5'ss, enquanto que a probabilidade desceria para 5,5%, se fizéssemos a mesma comparação entre íntrons humanos. Já a probabilidade de íntrons humanos compartilharem o mesmo sítio de ramificação é <1%, enquanto que para *S. cerevisiae* a probabilidade é de 94% (GAO e colab., 2008; KOL e colab., 2005).

Dada sua complexidade, os íntrons retidos têm sido classificados em 3 tipos de classes, variando em sua porcentagem de retenção intrônica, conteúdo de GC, tamanho do íntron, histórico evolucionário e seu efeito no *open reading frame* (ORF). Os membros da classe A são os mais abundantes, tendo uma porcentagem baixa de retenção intrônica, conteúdo intermediário de GC, tamanho médio e origem de íntrons ancestrais. Aqueles da classe B têm alta porcentagem de retenção intrônica e de conteúdo GC, são curtos e parecem derivar de éxons ancestrais. Os íntrons da classe C são caracterizados por serem longos e adjacentes a cassetes de éxons de genes conhecidos, exibindo uma porcentagem intermediária de retenção intrônica, com baixo conteúdo GC (JACOB e SMITH, 2017). A heterogeneidade acima descrita foi vista num estudo que comparou pacientes com câncer de mama em indivíduos afro-americanos e caucasianos, constatando a diferença na ocorrência de retenção intrônica nas duas populações (KIM e colab., 2018), assim como nós constatamos em nossos doadores (Figura 12 e Figura 13). Tal estudo sugeriu o fato das marcas epigenéticas serem distintas entre indivíduos de diferentes etnias como possível motivo para a heterogeneidade na incidência de RI. Embora concordemos com esta interpretação, estima-se que menos de 20% dos transcritos com RI sejam explicáveis pelo efeito da dinâmica epigenética sobre a maquinaria de *splicing* (GUO e colab., 2014; WONG e colab., 2017), indicando que muito do mecanismo regulatório da RI ainda está por ser determinado. Pelo que nos consta, a heterogeneidade individual no uso da RI somente

foi abordada na literatura pelo estudo apenas descrito com pacientes portadores de câncer e pelo trabalho experimental desta dissertação nos estados fisiológicos de ativação do linfócito. Conjuntamente, os dois trabalhos indicam que a transcrição das isoformas com RI possa ter uma certa versatilidade interindividual.

Trabalhar com retenção intrônica é desafiador, com enumeras dificuldades técnicas. Precisa-se de precaução, uma vez que os produtos com RI podem ser idênticos ao DNA genômico ou ao pré-mRNA. O emprego de oligo dT durante a síntese de DNA complementar (cDNA) é uma garantia de que os RNAs estão poliadenilados; no entanto, existe uma possibilidade de perda de espécies nucleares nascentes de RNA. A obtenção de *reads* na junção éxon-íntron também é um problema. O íntron em geral é muito maior que o éxon ocasionando uma menor cobertura *(reads),* enquanto que a cobertura dos éxons é facilitada pelo tamanho médio de 150 nucleotídeos. Em outras palavras, os éxons facilmente serão cobertos num sequenciamento de alta cobertura, enquanto que os íntrons não, dificultando o mapeamento fidedigno de um transcrito com retenção intrônica. (JACOB e SMITH, 2017). Há, portanto, uma limitação importante da tecnologia de sequenciamento de RNA atual, uma vez que ela nos permite ler 75-150 pb em um experimento padrão de RNA-seq. Todavia, os experimentos da área estão usando alta cobertura para aumentar a detecção e quantificação dos tipos de *splicing*, sendo ideal o uso de aproximadamente 150 milhões de *reads* por amostra (BRAUNSCHWEIG e colab., 2014; GUEROUSSOV e colab., 2015). O fato do transcriptoma apresentar transcritos com sequências repetitivas e em tandem, muitas vezes com elementos interpassados (LINE e SINE) e transposons também constitui-se em um verdadeiro desafio para os algoritmos atuais de mapeamento como Tophat2 (DOBIN e GINGERAS, 2013), STAR (DOBIN e colab., 2013) e MapSplice (WANG, Kai e colab., 2010).

De fato, encontramos algumas dessas dificuldades durante as análises *in silico* neste trabalho. Um dos genes encontrado na célula ativada com um alto índice de expressão intrônica foi aquele que codifica para a IL-2. Nosso algoritmo previu, para a transcrição do gene desta citocina, a existência de uma isoforma intrônica (TCONS_00237719) (ANEXO B - Q), cuja expressão seria maior do que todas as outras isoformas transcricionais. Esta isoforma intrônica não foi amplificada, enquanto a segunda isoforma exônica mais expressa o foi (dado não mostrado). Este resultado leva a crer que o pequeno fragmento representado na figura do ANEXO B - Q foi confundido com um íntron, devido a um erro de processamento de *reads*. Podemos também afirmar que é difícil para o algoritmo identificar íntrons na extremidade 5' do primeiro éxon do gene, o que pode justificar a má amplificação/correlação encontrada para

os genes CCNL1 e RHOA (Figura 14A e 14C). Todavia, o programa tem excelente valor preditivo para genes com cassetes intrônicos mais bem definidos e distantes da região 5', como FYN e SH3BRGL3, com ótima correlação entre o esperado das análises computacionais e o identificado experimentalmente.

Além disso, a validação destes transcritos requer bastante trabalho e nem sempre é possível, considerando a complexidade e multiplicidade das isoformas transcricionais de certos genes (ANEXO B). Um gene pode se expressar em diferentes isoformas e quase nunca a isoforma intrônica é a mais frequente. Se o gene não tiver uma expressão comparável àquela de um gene *housekeeping,* suas diferentes isoformas transcricionais serão difíceis de serem amplificadas, o que somente nos foi possível fazer com o uso da estratégia do *nested*-PCR. O desenho de *primers* específicos em meio a tantas isoformas transcricionais possíveis é trabalho árduo. Num cenário ideal, deveríamos amplificar e sequenciar o transcrito inteiro ou, pelo menos, fazer leituras *(reads)* bem maiores no sequenciamento de RNA. O trabalho aqui apresentado é, até onde nós sabemos, um dos poucos que fizeram a validação experimental da retenção intrônica em diferentes doadores e nos indica a existência de transcritos com RI diretamente ligados ao fenótipo celular de ativação e outros que poderão variar entre indivíduos.

7. CONCLUSÃO

A retenção intrônica é o evento de *splicing* alternativo mais comum nos linfócitos T CD4+. O linfócito T regula, aumentando ou diminuindo a carga intrônica transcricional (RPKM) em determinados genes, de maneira dependente do seu estado de ativação. Comprovamos a existência de transcritos com retenção intrônica expressos de maneira exclusiva vinculada ao estado fenotípico de ativação do linfócito T CD4+. Nossos resultados indicam que a retenção intrônica pode representar um ruído transcricional. Entretanto, o linfócito também pode usar este tipo de *splicing* alternativo de maneira seletiva dependendo do estado de ativação

8. REFERÊNCIAS

AMIT, Maayan e colab. **Differential GC Content between Exons and Introns Establishes Distinct Strategies of Splice-Site Recognition**. Cell Reports, v. 1, n. 5, p. 543–556, 2012.

ANDERSSON, Robin e colab. **Nucleosomes are well positioned in exons and carry characteristic histone modifications**. Genome Research, v. 19, n. 10, p. 1732–1741, 2009.

ANDREWS, Byron e colab. **Assembly factors for the membrane arm of human complex I.** Proceedings of the National Academy of Sciences of the United States of America, v. 110, n. 47, p. 18934–9, 2013.

BARSKI, Artem e colab. **Chromatin poises miRNA-and protein-coding genes for expression**. Genome Research, v. 19, p. 1742–1751, 2009.

BARSKI, Artem e colab. **High-Resolution Profiling of Histone Methylations in the Human Genome**. Cell, v. 129, n. 4, p. 823–837, 2007.

BARTEL, David P. **MicroRNAs: Genomics, Biogenesis, Mechanism, and Function**. Cell, v. 116, n. 2, p. 281–297, 2004.

BLACK, Douglas L. **Mechanisms of Alternative Pre-Messenger RNA Splicing**. Annual Review of Biochemistry, v. 72, n. 1, p. 291–336, 2003.

BON, Elisabeth e colab. **Molecular evolution of eukaryotic genomes: Hemiascomycetous yeast spliceosomal introns**. Nucleic Acids Research, v. 31, n. 4, p. 1121–1135, 2003.

BOOTHBY, Thomas C. e colab. **Removal of Retained Introns Regulates Translation in the Rapidly Developing Gametophyte of Marsilea vestita**. Developmental Cell, v. 24, n. 5, p. 517–529, 2013.

BOUTZ, Paul L. e BHUTKAR, Arjun e SHARP, Phillip A. **Detained introns are a novel, widespread class of post-transcriptionally spliced introns**. Genes and Development, v. 29, n. 1, p. 63–80, 2015.

BRAUNSCHWEIG, Ulrich e colab. **Widespread intron retention in mammals functionally tunes transcriptomes**. Genome Research, v. 24, n. 11, p. 1774–1786, 2014.

BUCKLEY, Peter T. e colab. **Cytoplasmic intron retention, function, splicing, and the sentinel RNA hypothesis**. Wiley Interdisciplinary Reviews: RNA, v. 5, n. 2, p. 223–230, 2014.

BUCKLEY, Peter T. e colab. **Cytoplasmic Intron Sequence-Retaining Transcripts Can Be Dendritically Targeted via ID Element Retrotransposons**. Neuron, v. 69, n. 5, p. 877–884, 2011.

BUSCH, Anke e HERTEL, Klemens J. **HEXEvent: A database of human EXon splicing Events**. Nucleic Acids Research, v. 41, n. D1, 2013.

CHEN, Mo e MANLEY, James L. **Mechanisms of alternative splicing regulation: Insights from molecular and genomics approaches**. Nature Reviews Molecular Cell Biology, v. 10, n. 11, p. 741–754, 2009.

CHIH, Long Liu e colab. **Single-nucleosome mapping of histone modifications in S. cerevisiae**. PLoS Biology, v. 3, n. 10, 2005.

CHO, V e colab. **The RNA-binding protein hnRNPLL induces a T cell alternative splicing program delineated by differential intron retention in polyadenylated RNA**. Genome Biol, v. 15, n. 1, p. R26, 2014.

CHOW, Louise T. e colab. **An amazing sequence arrangement at the 5??? ends of adenovirus 2 messenger RNA**. Cell, v. 12, n. 1, p. 1–8, 1977.

CLOSE, Pierre e colab. **Transcription Impairment and Cell Migration Defects in Elongator-Depleted Cells: Implication for Familial Dysautonomia**. Molecular Cell, v. 22, n. 4, p. 521–531, 2006.

COSSIO, María Laura T e colab. **Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks**. Nature Protocols, v. XXXIII, n. 2, p. 562, 2012.

CRAMER, Patrick. **RNA polymerase II structure: From core to functional complexes**. Current Opinion in Genetics and Development, v. 14, n. 2, p. 218–226, 2004.

CUDDAPAH, Suresh e BARSKI, Artem e ZHAO, Keji. **Epigenomics of T cell activation, differentiation, and memory**. Current Opinion in Immunology, v. 22, n. 3, p. 341–347, 2010.

DE SALORT, Jose e colab. **Ly9 (CD229) cell-surface receptor is crucial for the development of spontaneous autoantibody production to nuclear antigens**. Frontiers in Immunology, v. 4, n. JUL, p. 1–13, 2013.

DJEBALI, Sarah e colab. **Landscape of transcription in human cells**. Nature, v. 489, n. 7414, p. 101–108, 2012.

DOBIN, Alexander e colab. **STAR: Ultrafast universal RNA-seq aligner**. Bioinformatics, v. 29, n. 1, p. 15–21, 2013.

DOBIN, Alexander e GINGERAS, Thomas R. **Comment on " TopHat2 : accurate alignment of transcriptomes in the presence of insertions , deletions and gene fusions " by Kim et al . Comment on " TopHat2 : accurate alignment of transcriptomes in the presence of insertions , deletions and gene fusions** . p. 0–9, 2013.

DVINGE, Heidi e BRADLEY, Robert K. **Widespread intron retention diversifies most cancer transcriptomes.** Genome medicine, v. 7, n. 1, p. 45, 2015.

DVIR, Arik. **Promoter escape by RNA polymerase II**. Biochimica et Biophysica Acta - Gene Structure and Expression, v. 1577, n. 2, p. 208–223, 2002.

ERNST, Jason e KELLIS, Manolis. **ChromHMM: automating chromatin-state discovery and characterization.** Nature methods, v. 9, n. 3, p. 215–6, 2012.

EZQUERRA, Angel e colab. **Primary Structure of Papain-Solubilized Human Histocompatibility Antigen HLA-B27**. Biochemistry, v. 24, n. 7, p. 1733–1741, 1985.

FILIPOWICZ, Witold. **RNAi: The nuts and bolts of the RISC machine**. Cell, v. 122, n. 1, p. 17–20, 2005.

GAO, Kaiping e colab. **Human branch point consensus sequence is yUnAy**. Nucleic Acids Research, v. 36, n. 7, p. 2257–2267, 2008.

GATTINONI, L e colab. **A human memory T cell subset with stem cell-like properties**. Nat Med, v. 17, n. 10, p. 1290–1297, 2011.

GATTINONI, Luca e colab. **Adoptive immunotherapy for cancer: building on success.** Nature reviews. Immunology, v. 6, n. 5, p. 383–93, 2006.

GUEROUSSOV, Serge e colab. **An alternative splicing event amplifies evolutionary differences between vertebrates**. Science, v. 349, n. 6250, p. 868–873, 2015.

GUO, Rui e colab. **BS69/ZMYND11 reads and connects histone H3.3 lysine 36 trimethylation-decorated chromatin to regulated pre-mRNA processing**. Molecular Cell, v. 56, n. 2, p. 298–310, 2014.

HERMISTON, Michelle L e XU, Zheng e WEISS, Arthur. **CD45: A Critical Regulator of Signaling Thresholds in Immune Cells**. Annu. Rev. Immunol, v. 21, p. 107–37, 2003.

HSIN, Jing-ping e MANLEY, James L. **The RNA polymerase II CTD coordinates transcription and RNA processing**. Genes and Development, p. 2119–2137, 2012.

ILAGAN, Janine O e colab. **Rearrangements within human spliceosomes captured after exon ligation.** RNA (New York, N.Y.), v. 19, n. 3, p. 400–12, 2013.

JACOB, Aishwarya G. e SMITH, Christopher W.J. **Intron retention as a component of regulated gene expression programs**. Human Genetics, v. 136, n. 9, p. 1043–1057, 2017.

KEREN, Hadas e LEV-MAOR, Galit e AST, Gil. **Alternative splicing and evolution: diversification, exon definition and function.** Nature reviews. Genetics, v. 11, n. 5, p. 345–355, 2010.

KIM, Dongwook e colab. **Population-dependent Intron Retention and DNA Methylation in Breast Cancer**. Molecular Cancer Research, p. 1–10, 2018.

KINO, ManuscTomoshige e colab. **Noncoding RNA Gas5 Is a Growth Arrest and Starvation- Associated Repressor of the Glucocorticoid Receptor**. Science Signalling, v. 3, n. 107, p. 1–16, 2010.

KOL, Guy e LEV-MAOR, Galit e AST, Gil. **Human-mouse comparative analysis reveals that branch-site plasticity contributes to splicing regulation**. Human Molecular Genetics, v. 14, n. 11, p. 1559–1568, 2005.

KORNBERG, Rd. **Chromatin Structure : A Repeating Unit of Histones and DNA Chromatin structure is based on a repeating unit of eight**. Science, v. 184, p. 868–871, 1974.

KRISTJUHAN, Arnold e colab. **Transcriptional inhibition of genes with severe histone H3 hypoacetylation in the coding region**. Molecular Cell, v. 10, n. 4, p. 925–933, 2002.

KROGAN, Nevan J e colab. **Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II.** Molecular and cellular biology, v. 23, n. 12, p. 4207–18, 2003.

LANDER, E S e colab. **Initial sequencing and analysis of the human genome**. Nature, v. 409, n. 6822, p. 860–921, 2001.

LLORIAN, Miriam e colab. **The alternative splicing program of differentiated smooth muscle cells involves concerted non-productive splicing of post-transcriptional regulators**. Nucleic Acids Research, v. 44, n. 18, p. 8933–8950, 2016.

LOUHICHI, Amel e FOURATI, Ahmed e REBAÏ, Ahmed. **IGD: A resource for intronless genes in the human genome**. Gene, v. 488, n. 1–2, p. 35–40, 2011.

LUCO, Rf e PAN, Qun e TOMINAGA, Kaoru. **Regulation of alternative splicing by histone modifications**. Science, v. 327, n. 5968, p. 996–1000, 2010.

LUO, Weifei e BENTLEY, David. **A ribonucleolytic rat torpedoes RNA polymerase II**. Cell, v. 119, n. 7, p. 911–914, 2004.

MACIÁN, F. e LÓPEZ-RODRÍGUEZ, C. e RAO, A. **Partners in transcription: NFAT and AP-1**. Oncogene, v. 20, n. 19 REV. ISS. 2, p. 2476–2489, 2001.

MARQUEZ, Yamile e colab. **Transcriptome survey reveals increased complexity of the alternative splicing landscape in Arabidopsis**. Genome Research, v. 22, n. 6, p. 1184–1195, 2012.

MARTIN, M e colab. **Identification of Grb2 as a novel binding partner of the signaling lymphocytic activation molecule-associated protein binding receptor CD229**. Journal of immunology (Baltimore, Md.: 1950), v. 174, n. 10, p. 5977–5986, 2005.

MARTINEZ, Nicole M. e LYNCH, Kristen W. **Control of alternative splicing in immune responses: Many regulators, many predictions, much still to learn**. Immunological Reviews, v. 253, n. 1, p. 216–236, 2013.

MATLIN, Arianne J e CLARK, Francis e SMITH, Christopher W J. **Understanding alternative splicing: towards a cellular code**. Nature reviews. Molecular cell biology, v. 6, n. 5, p. 386–398, 2005.

MAUGER, Oriane e LEMOINE, Frédéric e SCHEIFFELE, Peter. **Targeted Intron Retention and Excision for Rapid Gene Regulation in Response to Neuronal Activity**. Neuron, v. 92, n. 6, p. 1266–1278, 2016.

NARO, Chiara e colab. **An Orchestrated Intron Retention Program in Meiosis Controls Timely Usage of Transcripts during Germ Cell Differentiation**. Developmental Cell, v. 41, n. 1, p. 82–93.e4, 2017.

NI, Min e colab. **Regulation of PERK signaling and leukemic cell survival by a novel cytosolic isoform of the UPR regulator GRP78/BiP**. PLoS ONE, v. 4, n. 8, 2009.

NI, Ting e colab. **Global intron retention mediated gene regulation during CD4+ T cell activation**. Nucleic Acids Research, v. 44, n. 14, p. 6817–6829, 2016.

NILSEN, Timothy W e GRAVELEY, Brenton R. **Expansion of the eukaryotic proteome by alternative splicing.** Nature, v. 463, n. 7280, p. 457–63, 2010.

NINOMIYA, Kensuke e KATAOKA, Naoyuki e HAGIWARA, Masatoshi. **Stress-responsive maturation of Clk1/4 pre-mRNAs promotes phosphorylation of SR splicing factor**. Journal of Cell Biology, v. 195, n. 1, p. 27–40, 2011.

NISHIMURA, Yasumitsu e colab. **Insufficient interleukin-2 production from splenic CD4+T cells causes impaired cell proliferation and early apoptosis in SAMP1, a strain of senescence-accelerated mouse**. Immunology, v. 107, n. 2, p. 190–198, 2002.

OHSHIMA, Y e colab. **Novel models for RNA splicing that involve a small nuclear RNA.** Proceedings of the National Academy of Sciences of the United States of America, v. 78, n. 7, p. 4471–4, 1981.

PAN, Qun e colab. **Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing**. Nature Genetics, v. 40, n. 12, p. 1413–1415, 2008.

PIMENTEL, Harold e colab. **A dynamic intron retention program enriched in RNA processing genes regulates gene expression during terminal erythropoiesis**. Nucleic Acids Research, v. 44, n. 2, p. 838–851, 2016.

POKHOLOK, Dmitry K. e colab. **Genome-wide map of nucleosome acetylation and methylation in yeast**. Cell, v. 122, n. 4, p. 517–527, 2005.

PTASHNE, M e GANN, A. **Transcriptional activation by recruitment**. Nature. [S.l: s.n.].

RAGHUNATHAN, P L e GUTHRIE, C. **RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2.** Current biology : CB, v. 8, n. 15, p. 847–55, 1998.

RAO, B. e colab. **Dimethylation of Histone H3 at Lysine 36 Demarcates Regulatory and Nonregulatory Chromatin Genome-Wide**. Molecular and Cellular Biology, v. 25, n. 21, p.

9447–9459, 2005.

REDDY, Anireddy S N e colab. **Deciphering the Plant Splicing Code: Experimental and Computational Approaches for Predicting Alternative Splicing and Splicing Regulatory Elements**. Frontiers in Plant Science, v. 3, n. February, p. 18, 2012.

ROEDER, Robert G. **Transcriptional regulation and the role of diverse coactivators in animal cells**. FEBS Letters, v. 579, n. 4 SPEC. ISS., p. 909–915, 2005.

SAINSBURY, Sarah e BERNECKY, Carrie e CRAMER, Patrick. **Structural basis of transcription initiation by RNA polymerase II.** Nature reviews. Molecular cell biology, v. 16, n. 3, p. 129–143, 2015.

SAKABE, Noboru Jo e DE SOUZA, Sandro José. **Sequence features responsible for intron retention in human.** BMC genomics, v. 8, p. 59–72, 2007.

SAMMETH, Michael e FOISSAC, Sylvain e GUIG??, Roderic. **A general definition and nomenclature for alternative splicing events**. PLoS Computational Biology, v. 4, n. 8, 2008.

SAUNDERS, Abbie e CORE, Leighton J e LIS, John T. **Breaking barriers to transcription elongation.** Nature reviews. Molecular cell biology, v. 7, n. 8, p. 557–67, 2006.

SCHWARTZ, Schraga e MESHORER, Eran e AST, Gil. **Chromatin organization marks exon-intron structure.** Nature Structural & Molecular Biology, v. 16, n. 9, p. 990–995, 2009.

SCHWARZL, Thomas e colab. **Measuring Transcription Rate Changes via Time-Course 4-Thiouridine Pulse-Labelling Improves Transcriptional Target Identification**. Journal of Molecular Biology, v. 427, n. 21, p. 3368–3374, 2015.

SCHWER, Beate e GROSS, Christian H. **Prp22, a DExH-box RNA helicase, plays two distinct roles in yeast pre-mRNA splicing**. EMBO Journal, v. 17, n. 7, p. 2086–2094, 1998.

SHALGI, Reut e colab. **Widespread Regulation of Translation by Elongation Pausing in Heat Shock**. Molecular Cell, v. 49, n. 3, p. 439–452, 2013.

SHEN, Haihong e KAN, Julie L.C. e GREEN, Michael R. **Arginine-Serine-Rich Domains Bound at Splicing Enhancers Contact the Branchpoint to Promote Prespliceosome Assembly**. Molecular Cell, v. 13, n. 3, p. 367–376, 2004.

SHEPARD, Peter J e HERTEL, Klemens J. **The SR protein family**. Genome Biology, v. 10, n. 10, p. 242, 2009.

SINTES, J. e colab. **Characterization of mouse CD229 (Ly9), a leukocyte cell surface molecule of the CD150 (SLAM) family**. Tissue Antigens, v. 70, n. 5, p. 355–362, 2007.

SMITH-GARVIN, Jennifer E e colab. **T Cell Activation**. Annual Review of Immunology, v. 27, n. 1, p. 591–619, 2009.

SPIES, Noah e colab. **Biased Chromatin Signatures around Polyadenylation Sites and Exons**. Molecular Cell, v. 36, n. 2, p. 245–254, 2009.

STAIGER, Dorothee e SIMPSON, Gordon G. **Enter exitrons**. Genome Biology, v. 16, n. 1, p. 15–17, 2015.

STERNER, D A e CARLO, T e BERGET, S M. **Architectural limits on split genes.** Proceedings of the National Academy of Sciences of the United States of America, v. 93, n. 26, p. 15081–5, 1996.

TARN, Woan Yuh e STEITZ, Joan A. **A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class (AT-AC) intron in vitro**. Cell, v. 84, n. 5, p. 801–811, 1996.

TRAVELLA, S. **RNA Interference-Based Gene Silencing as an Efficient Tool for Functional Genomics in Hexaploid Bread Wheat**. Plant Physiology, v. 142, n. 1, p. 6–20, 2006.

TURUNEN, Janne J. e colab. **The significant other: Splicing by the minor spliceosome**. Wiley Interdisciplinary Reviews: RNA, v. 4, n. 1, p. 61–76, 2013.

VALCÁRCEL, Juan e GREEN, Michael R. **The SR protein family: Pleiotropic functions in pre-mRNA splicing**. Trends in Biochemical Sciences, v. 21, n. 8, p. 296–301, 1996.

VAN BERGEIJK, Petra e HOOGENRAAD, Casper C. e KAPITEIN, Lukas C. **Right Time, Right Place: Probing the Functions of Organelle Positioning**. Trends in Cell Biology, v. 26, n. 2, p. 121–134, 2016.

VAN HAAFTEN-VISSER, Désirée Y. e colab. **Ankyrin repeat and zinc-finger domain-containing 1 mutations are associated with infantile-onset inflammatory bowel disease**. Journal of Biological Chemistry, v. 292, n. 19, p. 7904–7920, 2017.

VANICHKINA, Darya P. e colab. **Challenges in defining the role of intron retention in**

normal biology and disease. Seminars in Cell and Developmental Biology, 2017.

VITTING-SEERUP, Kristoffer e colab. **spliceR: an R package for classification of alternative splicing and prediction of coding potential from RNA-seq data.** BMC bioinformatics, v. 15, p. 81, 2014.

WAHL, Markus C. e WILL, Cindy L. e LÜHRMANN, Reinhard. **The Spliceosome: Design Principles of a Dynamic RNP Machine**. Cell, v. 136, n. 4, p. 701–718, 2009.

WANG, Eric T e colab. **Alternative isoform regulation in human tissue transcriptomes.** Nature, v. 456, n. 7221, p. 470–6, 2008.

WANG, Kai e colab. **MapSplice: Accurate mapping of RNA-seq reads for splice junction discovery**. Nucleic Acids Research, v. 38, n. 18, p. 1–14, 2010.

WANG, W e CAREY, M e GRALLA, JD. **Polymerase II promoter activation: closed complex formation and ATP-driven start-site opening**. Science, v. 255, n. 5043, p. 450–453, 1992.

WANG, Zhibin e colab. **Combinatorial patterns of histone acetylations and methylations in the human genome.** Nature genetics, v. 40, n. 7, p. 897–903, 2008.

WEAKE, V e WORKMAN, J. **Inducible gene expression: diverse regulatory mechanisms**. Nat Rev Genet, v. 11, n. 6, p. 426–437, 2010.

WONG, Justin J.L. e colab. **Intron retention is regulated by altered MeCP2-mediated splicing factor recruitment**. Nature Communications, v. 8, n. May, p. 1–13, 2017.

WONG, Justin J L e colab. **Orchestrated intron retention regulates normal granulocyte differentiation**. Cell, v. 154, n. 3, p. 583–595, 2013.

XIAO, Tiaojiang e colab. **Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast**. Genes and Development, v. 17, n. 5, p. 654–663, 2003.

YAP, Karen e colab. **Coordinated regulation of neuronal mRNA steady-state levels through developmentally controlled intron retention**. Genes and Development, v. 26, n. 11, p. 1209–1223, 2012.

YEO, G e colab. **Variation in alternative splicing across human tissues**. Genome Biol, v. 5, n. 10, p. R74, 2004.

ZHANG, Yi e colab. **Host-reactive CD8+ memory stem cells in graft-versus-host disease.**

Nature medicine, v. 11, n. 12, p. 1299–1305, 2005.

ZHANG, Zhiqiang e WU, Chwen Huey e GILMOUR, David S. **Analysis of polymerase II elongation complexes by native gel electrophoresis: Evidence for a novel carboxyl-terminal domain-mediated termination mechanism**. Journal of Biological Chemistry, v. 279, n. 22, p. 23223–23228, 2004.

9. ANEXOS

9.1 ANEXO A

Tabela 7 – Recorte da lista de transcritos com retenção intrônica exclusivos da célula T CD4+ em repouso

RPKM	Gene	Isoforma-ID	CTP	RPKM total do gene	% da isoforma no gene
6,06	MRFAP1	TCONS_00223923	-	79,59	7,61
6,08	-	TCONS_00044663	-	107,05	5,68
6,30	BSDC1	TCONS_00023434	SIM	24,88	25,35
6,36	UBA7	TCONS_00210830	-	73,08	8,71
6,64	MTF2	TCONS_00008675	SIM	21,03	31,57
6,68	-	TCONS_00044685	NÃO	50,01	13,37
6,82	RP4-781K5.2	TCONS_00018409	-	6,82	100
7,27	-	TCONS_00131865	NÃO	702,20	1,03
7,66	EIF4A2	TCONS_00205701	NÃO	1021,09	0,75
7,81	CCDC66	TCONS_00198808	TRUE	78,35	9,96
7,87	CAPG	TCONS_00165388	NÃO	22,01	35,74
7,97	RPSA	TCONS_00196574	NÃO	442,95	1,79
8,02	BCLAF1	TCONS_00280585	-	250,07	3,20
8,07	SNHG8	TCONS_00228668	-	426,70	1,89
8,33	CD28	TCONS_00156040	NÃO	140,80	5,92
8,99	-	TCONS_00069601	-	38,26	23,5
10,31	EIF4A2	TCONS_00205705	-	1021,09	1,01
10,37	LY9	TCONS_00013304	NÃO	32,04	32,37
10,94	MTCH1	TCONS_00275848	SIM	51,82	21,11
10,95	RPS7	TCONS_00144084	NÃO	166,18	6,59
12,24	-	TCONS_00068212	NÃO	67,05	18,26
13,06	LDLRAP1	TCONS_00002887	SIM	55,02	23,74
14,13	PIK3R1	TCONS_00244844	NÃO	231,85	6,09
14,20	UBXN11	TCONS_00022609	SIM	28,50	49,83
14,42	ANKZF1	TCONS_00157297	SIM	77,81	18,53
33,83	EIF4A2	TCONS_00205685	SIM	1021,09	3,31

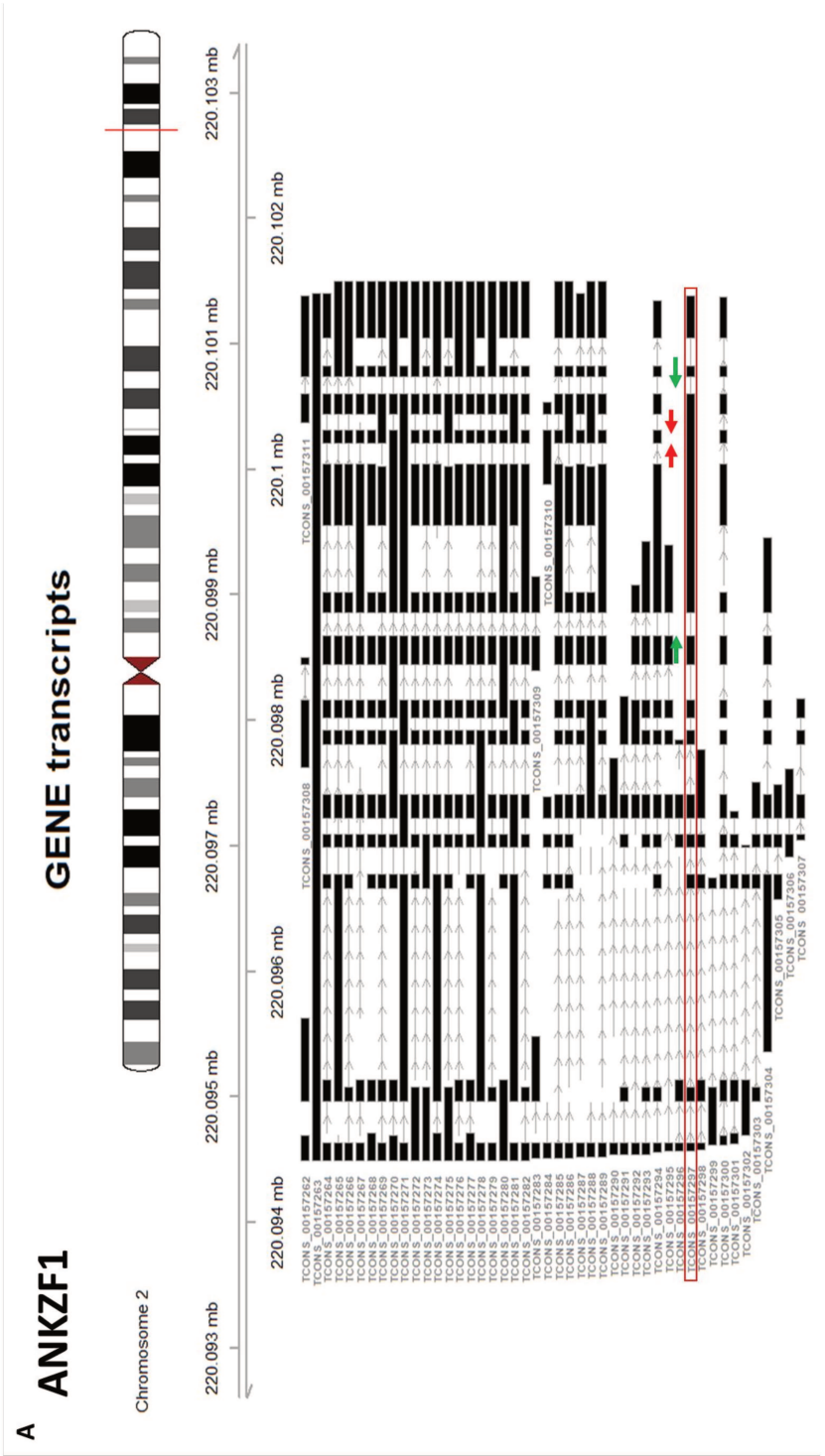
Tabela 8 – Recorte da lista de transcritos com retenção intrônica exclusivos da célula T CD4+ ativada

RPKM	Gene	Isoforma-ID	CTP	RPKM total do gene	% da isoforma no gene
23,16	INO80B	TCONS_00148616	NÃO	70,00	33,09
23,47	RBPJ	TCONS_00224787	NÃO	488,63	4,80
23,53	NOP58	TCONS_00155841	-	688,77	3,41
24,00	EIF4A2	TCONS_00205691	NÃO	1724,77	1,39
24,09	EIF4A2	TCONS_00205687	NÃO	1724,77	1,39
24,93	RBPJ	TCONS_00224771	NÃO	488,63	5,10
26,55	TIMMDC1	TCONS_00201029	NÃO	127,71	20,78
28,34	DAP3	TCONS_00012612	-	518,49	5,46
29,15	TNIP1	TCONS_00260070	NÃO	379,28	7,68
33,71	EIF4G1	TCONS_00205352	NÃO	374,68	8,99
37,40	PNPT1	TCONS_00163368	SIM	136,85	27,33
38,65	NCBP2	TCONS_00220148	NÃO	294,17	13,13

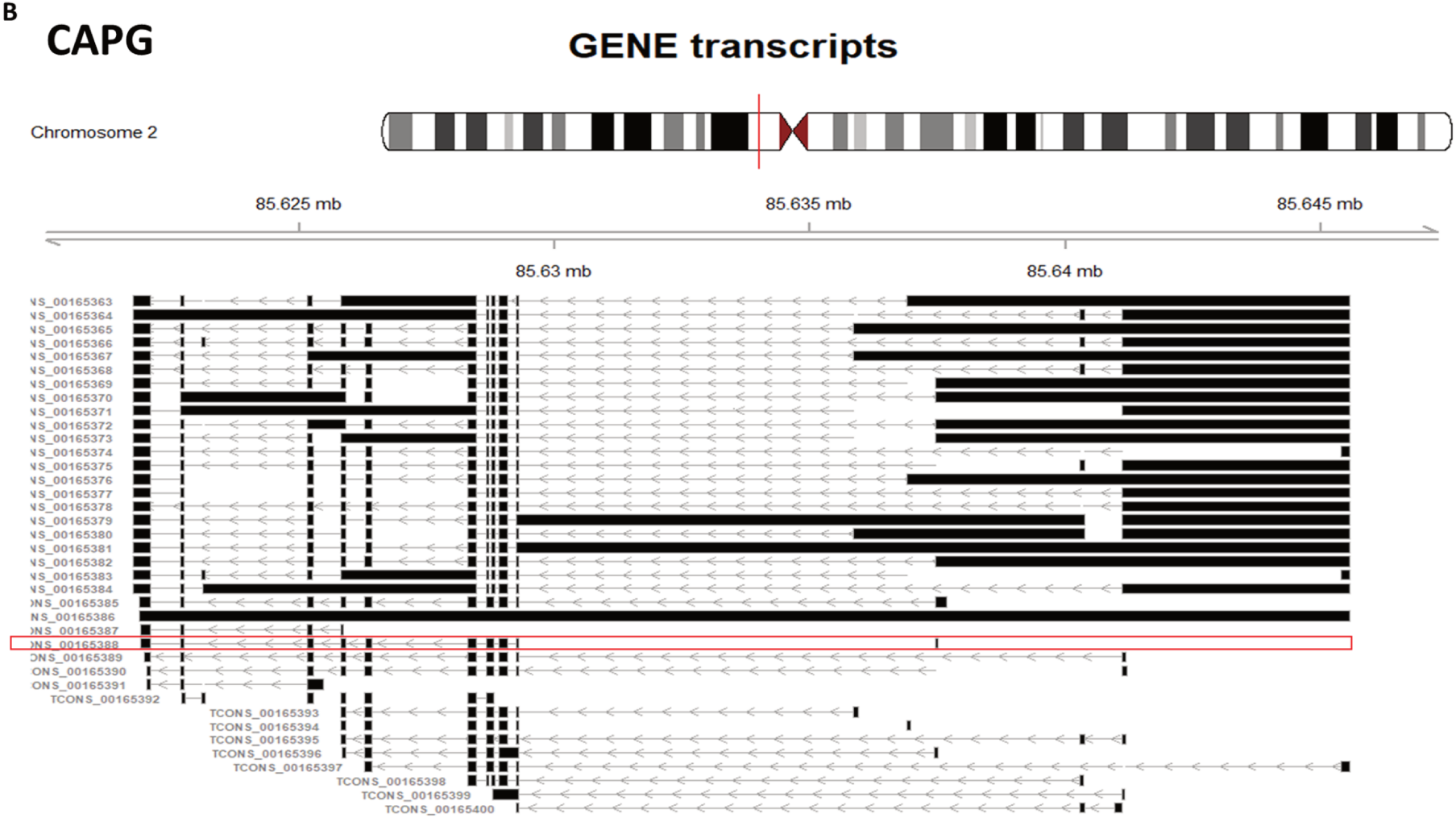
46,73	CCDC58	TCONS_00214609	NÃO	135,59	34,46
47,35	NOP16	TCONS_00261231	NÃO	450,08	10,52
48,25	IMPDH2	TCONS_00210465	-	534,92	9,02
76,83	CLDND1	TCONS_00213169	SIM	281,84	27,26
90,03	MBNL1	TCONS_00203574	NÃO	291,58	30,87
90,36	RPS14	TCONS_00259923	NÃO	1604,24	5,63
102,39	GAS5	TCONS_00034588	-	593,47	17,25
208,21	GNB2L1	TCONS_00262323	SIM	5244,05	3,97

9.2 ANEXO B

Neste anexo B, mostramos a representações que fizemos de todos os transcritos de cada gene a partir do nosso programa algorítmico, com suas respectivas tabelas de expressão. Usamos essas representações para identificar os transcritos com retenção intrônica e desenhar os *primers*. As caixas pretas são cassetes de éxons ou íntrons em um mRNA já processado e as flechas são partes do transcrito que foram excluídos durante o *splicing* alternativo. Para facilitar a visualização desses dados, colocamos uma caixa vermelha para evidenciar o transcrito alvo que analisámos nessa dissertação. Nos transcritos analisados experimentalmente, desenhamos setas em verde para retratar em qual cassete colocamos os *primers* da primeira reação de nested-PCR e as setas em vermelho para a segunda reação do *nested*-PCR.

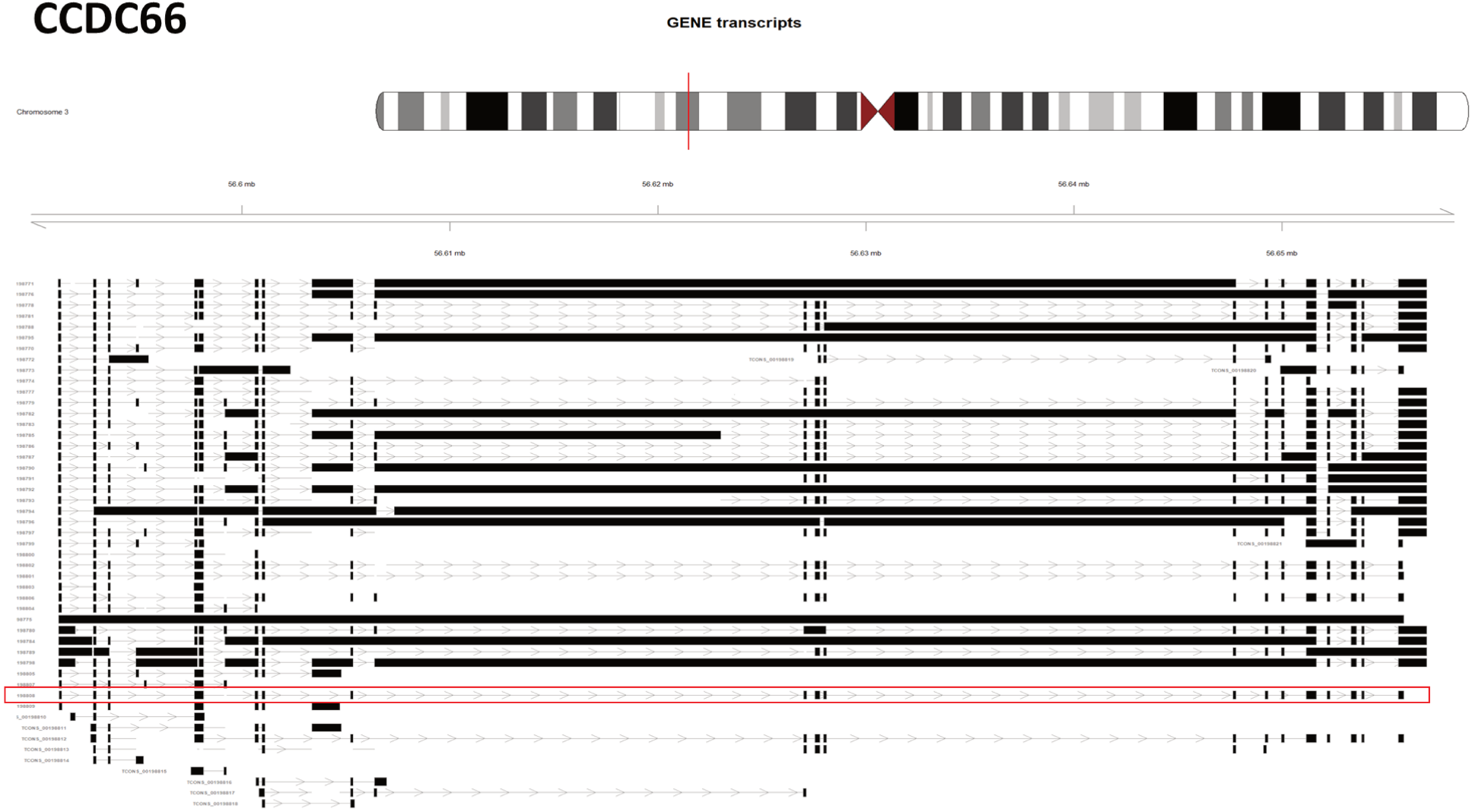

A
ANKZF1
GENE transcripts
Chromosome 2
220.093 mb
220.095 mb
220.097 mb
220.099 mb
220.101 mb
220.103 mb
220.094 mb
220.096 mb
220.098 mb
220.1 mb
220.102 mb
TCONS_00157262
TCONS_00157263
TCONS_00157264
TCONS_00157265
TCONS_00157266
TCONS_00157267
TCONS_00157268
TCONS_00157269
TCONS_00157270
TCONS_00157271
TCONS_00157272
TCONS_00157273
TCONS_00157274
TCONS_00157275
TCONS_00157276
TCONS_00157277
TCONS_00157278
TCONS_00157279
TCONS_00157280
TCONS_00157281
TCONS_00157282
TCONS_00157283
TCONS_00157284
TCONS_00157285
TCONS_00157286
TCONS_00157287
TCONS_00157288
TCONS_00157289
TCONS_00157290
TCONS_00157291
TCONS_00157292
TCONS_00157293
TCONS_00157294
TCONS_00157295
TCONS_00157296
TCONS_00157297
TCONS_00157298
TCONS_00157299
TCONS_00157300
TCONS_00157301
TCONS_00157302
TCONS_00157303
TCONS_00157304
TCONS_00157305
TCONS_00157306
TCONS_00157307
TCONS_00157308
TCONS_00157309
TCONS_00157310
TCONS_00157311

B


CAPG
GENE transcripts
Chromosome 2
85.625 mb
85.635 mb
85.645 mb
85.63 mb
85.64 mb
NS_00165363
NS_00165364
NS_00165365
NS_00165366
NS_00165367
NS_00165368
NS_00165369
NS_00165370
NS_00165371
NS_00165372
NS_00165373
NS_00165374
NS_00165375
NS_00165376
NS_00165377
NS_00165378
NS_00165379
NS_00165380
NS_00165381
NS_00165382
NS_00165383
NS_00165384
NS_00165385
NS_00165386
NS_00165387
NS_00165388
NS_00165389
NS_00165390
NS_00165391
TCONS_00165392
TCONS_00165393
TCONS_00165394
TCONS_00165395
TCONS_00165396
TCONS_00165397
TCONS_00165398
TCONS_00165399
TCONS_00165400

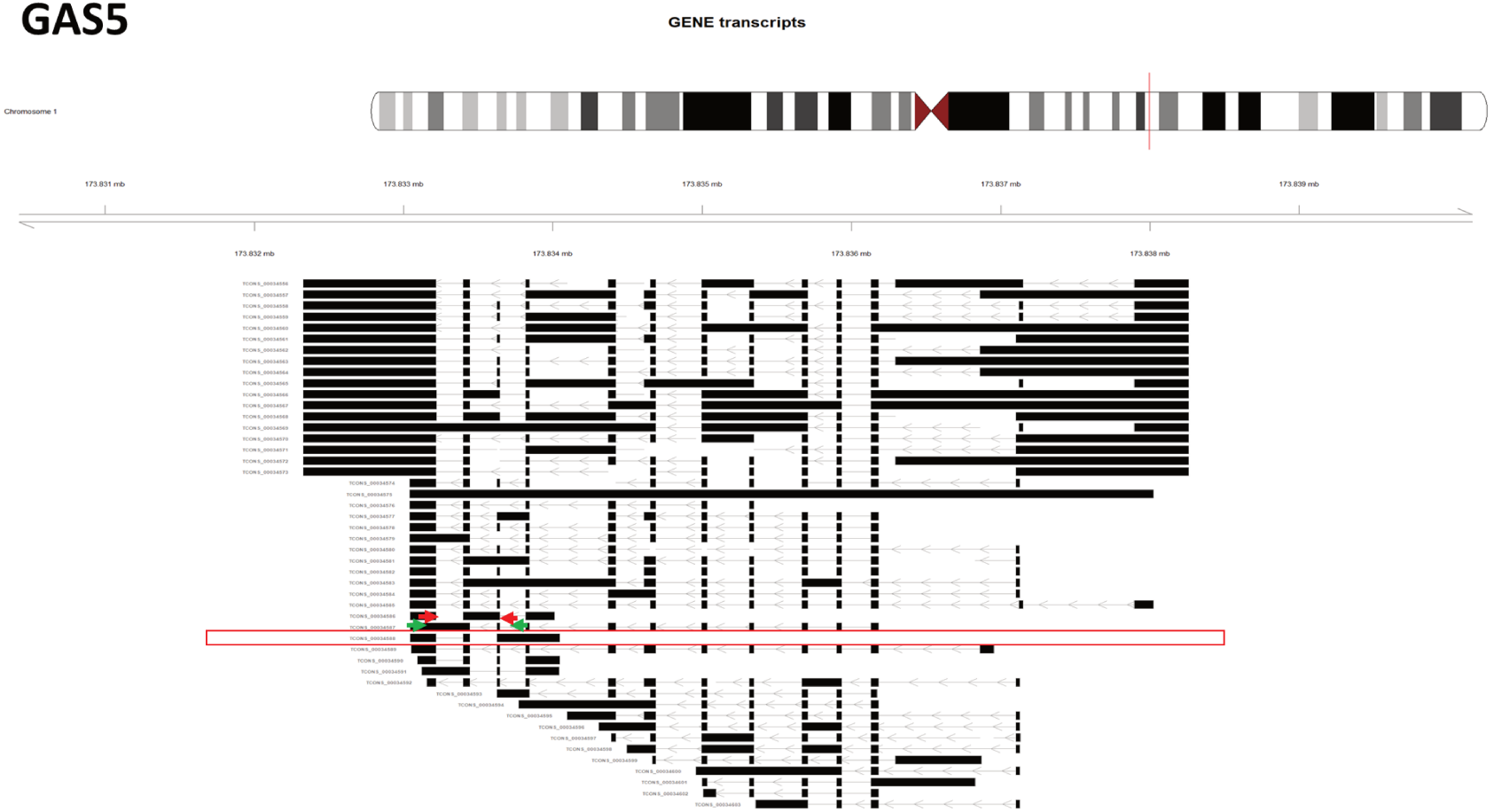
c

CCDC66


GENE transcripts
Chromosome 3
56.6 mb
56.62 mb
56.64 mb
56.61 mb
56.63 mb
56.65 mb

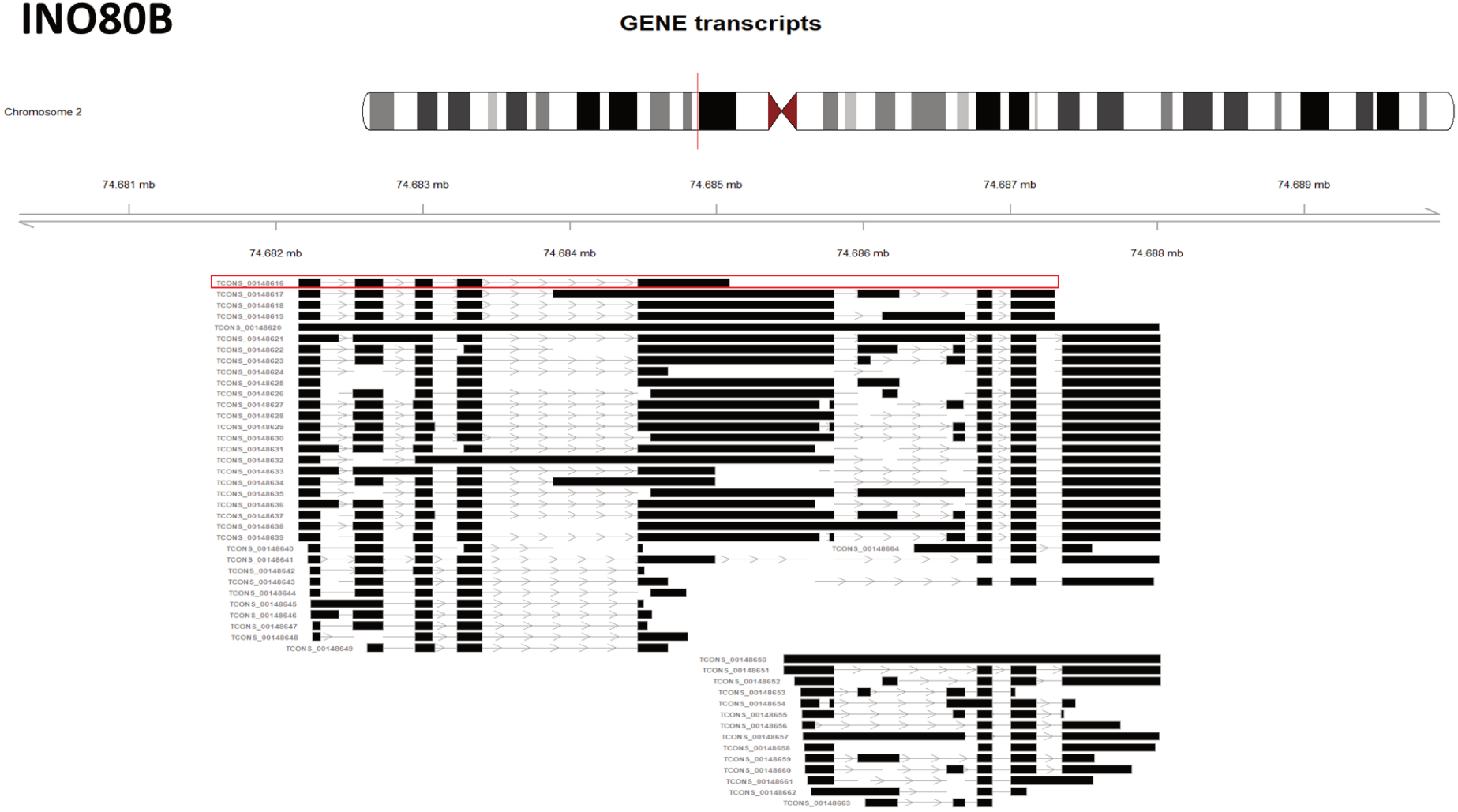
D

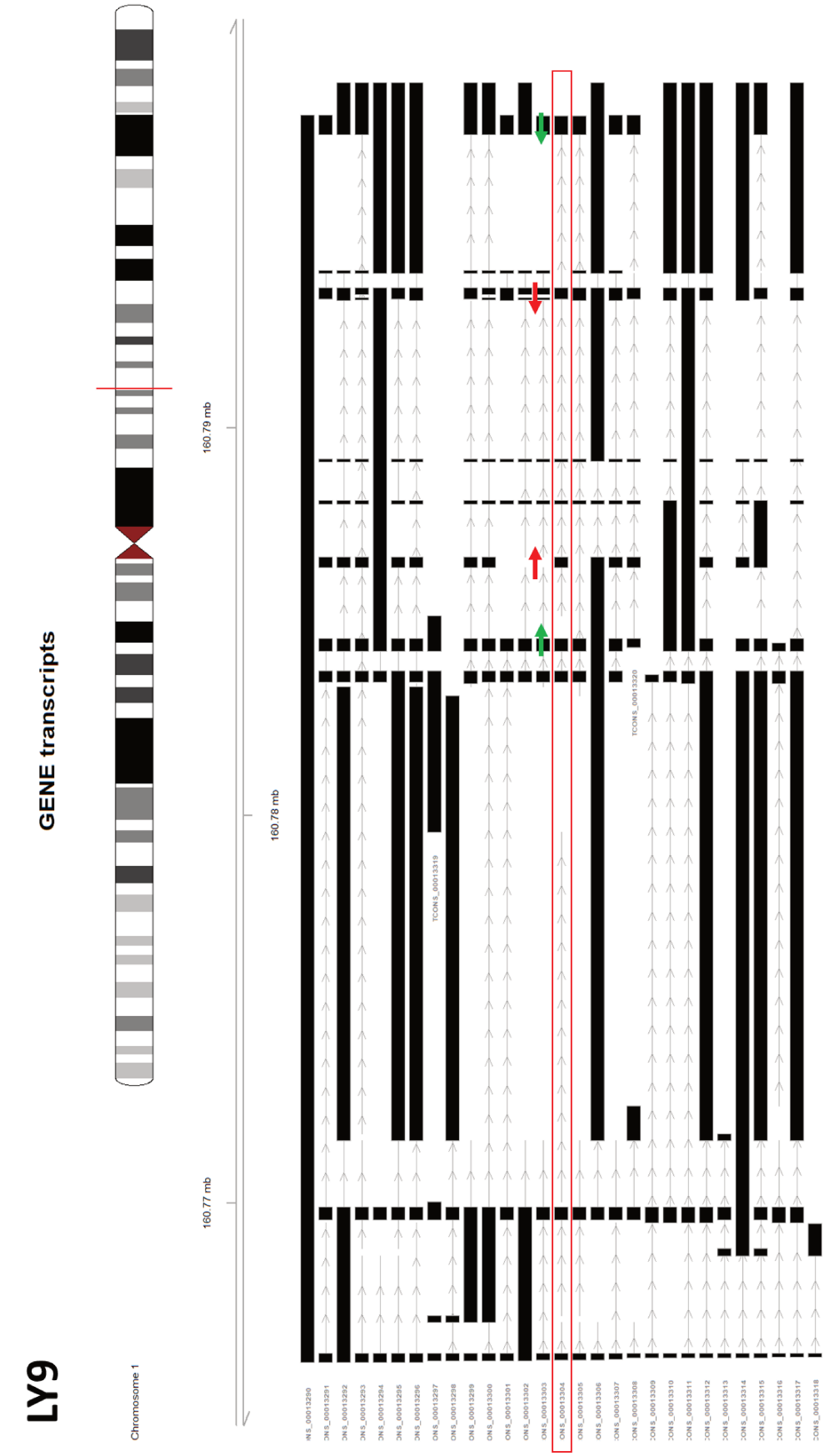
GAS5


GENE transcripts
Chromosome 1
173.831 mb
173.833 mb
173.835 mb
173.837 mb
173.839 mb
173.832 mb
173.834 mb
173.836 mb
173.838 mb

E

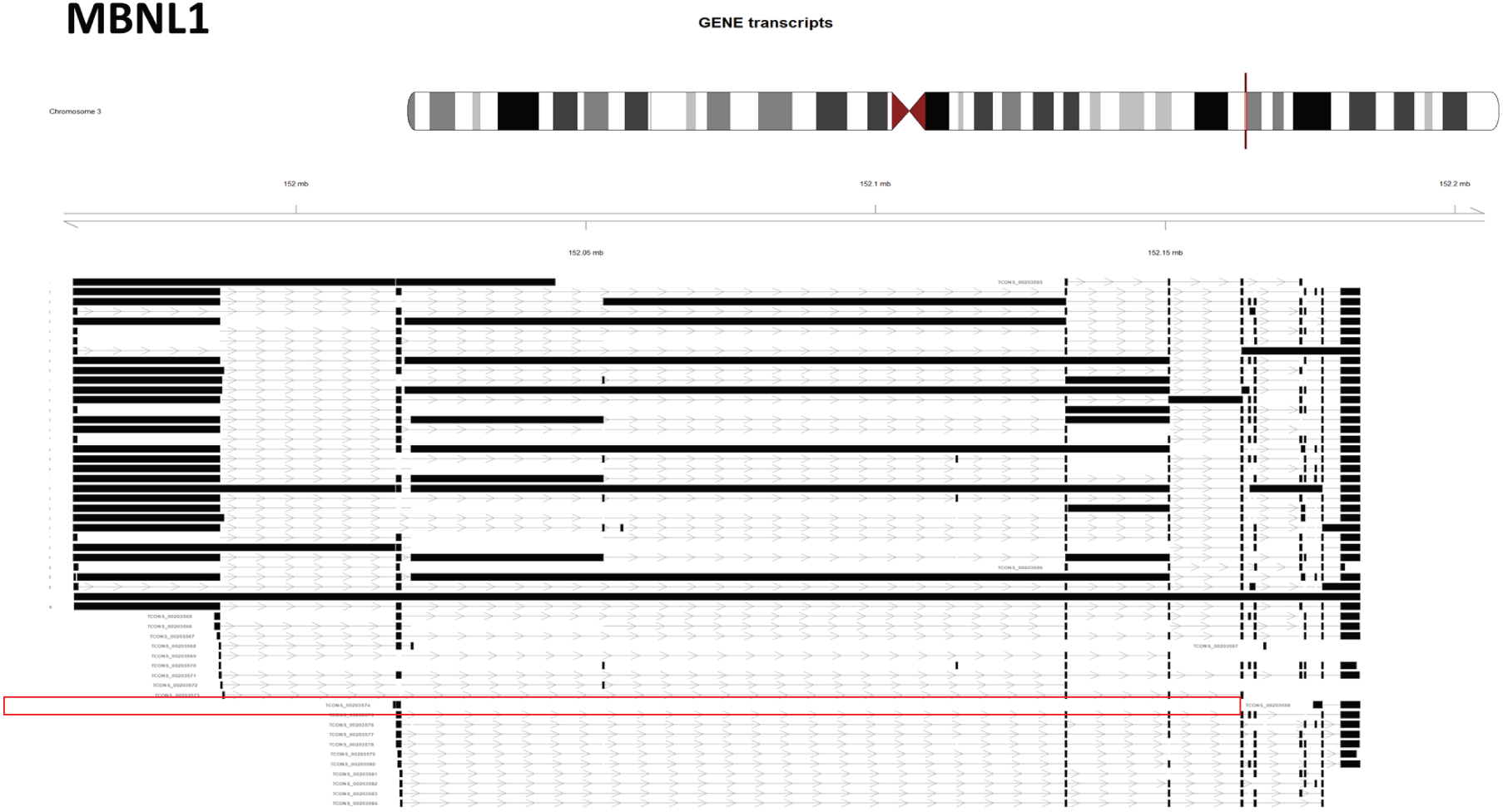
INO80B


GENE transcripts
Chromosome 2
74.681 mb
74.683 mb
74.685 mb
74.687 mb
74.689 mb
74.682 mb
74.684 mb
74.686 mb
74.688 mb
TCONS_00148616
TCONS_00148617
TCONS_00148618
TCONS_00148619
TCONS_00148620
TCONS_00148621
TCONS_00148622
TCONS_00148623
TCONS_00148624
TCONS_00148625
TCONS_00148626
TCONS_00148627
TCONS_00148628
TCONS_00148629
TCONS_00148630
TCONS_00148631
TCONS_00148632
TCONS_00148633
TCONS_00148634
TCONS_00148635
TCONS_00148636
TCONS_00148637
TCONS_00148638
TCONS_00148639
TCONS_00148640
TCONS_00148641
TCONS_00148642
TCONS_00148643
TCONS_00148644
TCONS_00148645
TCONS_00148646
TCONS_00148647
TCONS_00148648
TCONS_00148649
TCONS_00148664
TCONS_00148650
TCONS_00148651
TCONS_00148652
TCONS_00148653
TCONS_00148654
TCONS_00148655
TCONS_00148656
TCONS_00148657
TCONS_00148658
TCONS_00148659
TCONS_00148660
TCONS_00148661
TCONS_00148662
TCONS_00148663


F
LY9
GENE transcripts
Chromosome 1
160.77 mb
160.78 mb
160.79 mb
TCONS_00013319
TCONS_00013320

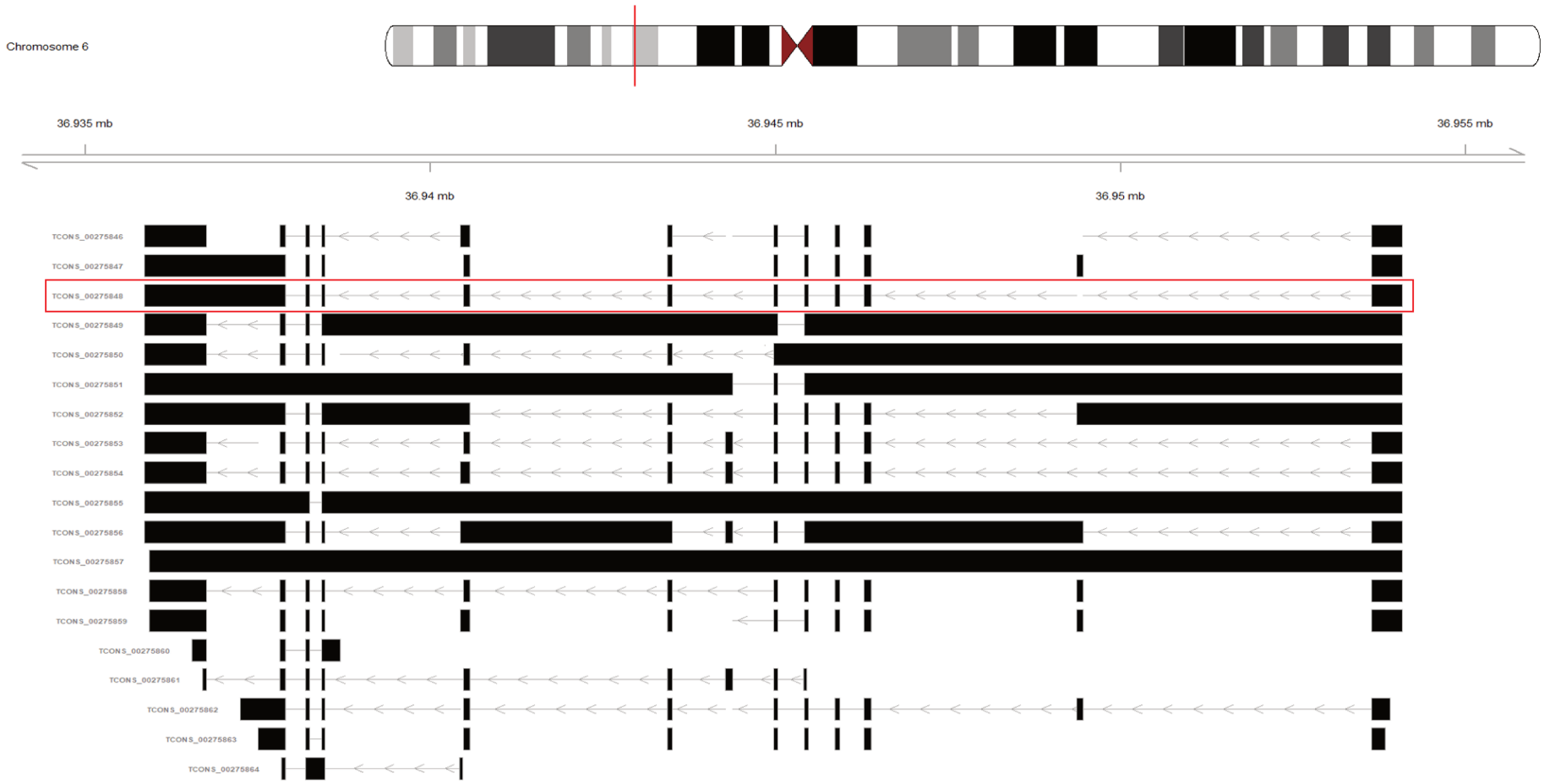
G

MBNL1


GENE transcripts
Chromosome 3
152 mb
152.1 mb
152.2 mb
152.05 mb
152.15 mb

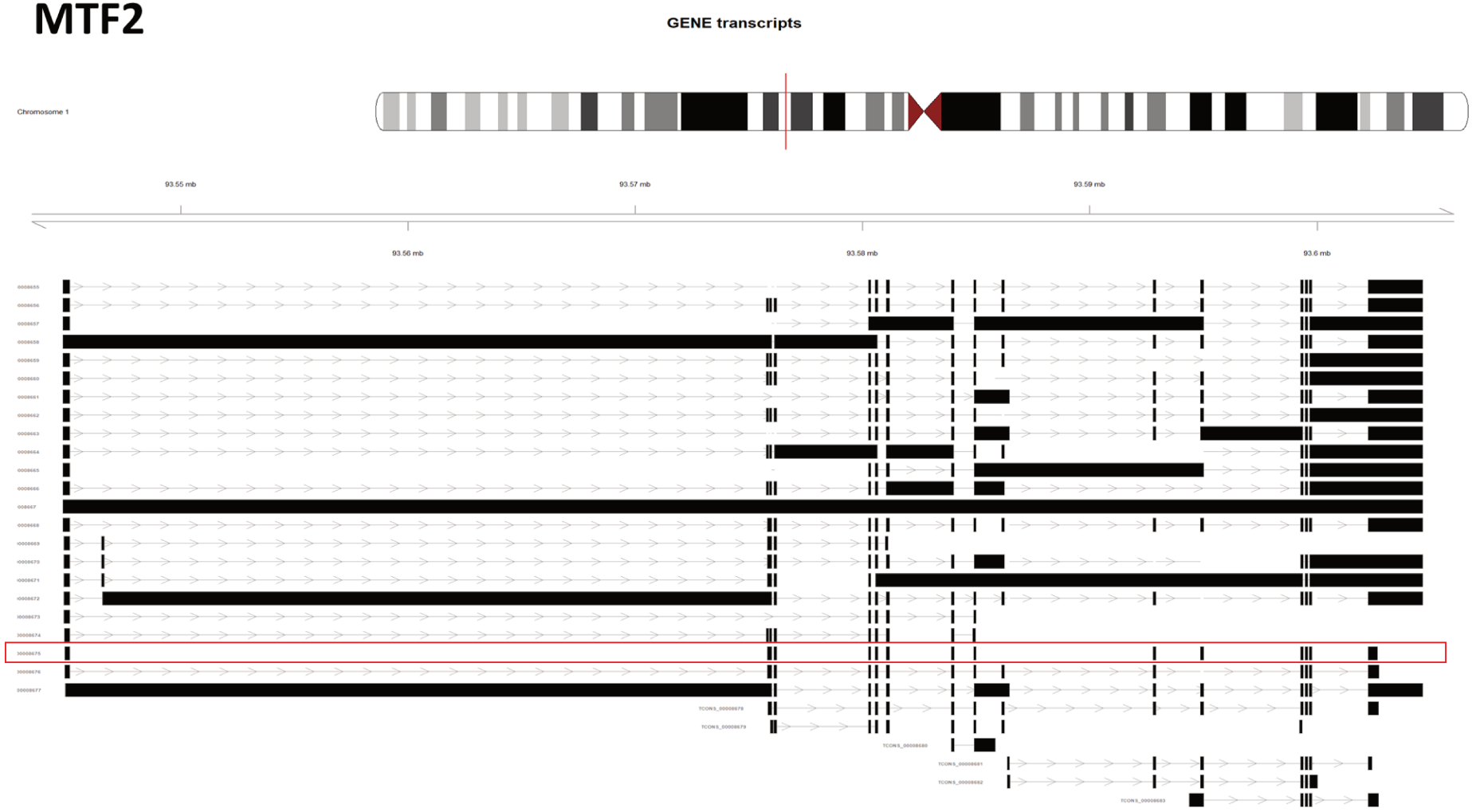
H

MTCH1


GENE transcripts
Chromosome 6
36.935 mb
36.945 mb
36.955 mb
36.94 mb
36.95 mb
TCONS_00275846
TCONS_00275847
TCONS_00275848
TCONS_00275849
TCONS_00275850
TCONS_00275851
TCONS_00275852
TCONS_00275853
TCONS_00275854
TCONS_00275855
TCONS_00275856
TCONS_00275857
TCONS_00275858
TCONS_00275859
TCONS_00275860
TCONS_00275861
TCONS_00275862
TCONS_00275863
TCONS_00275864

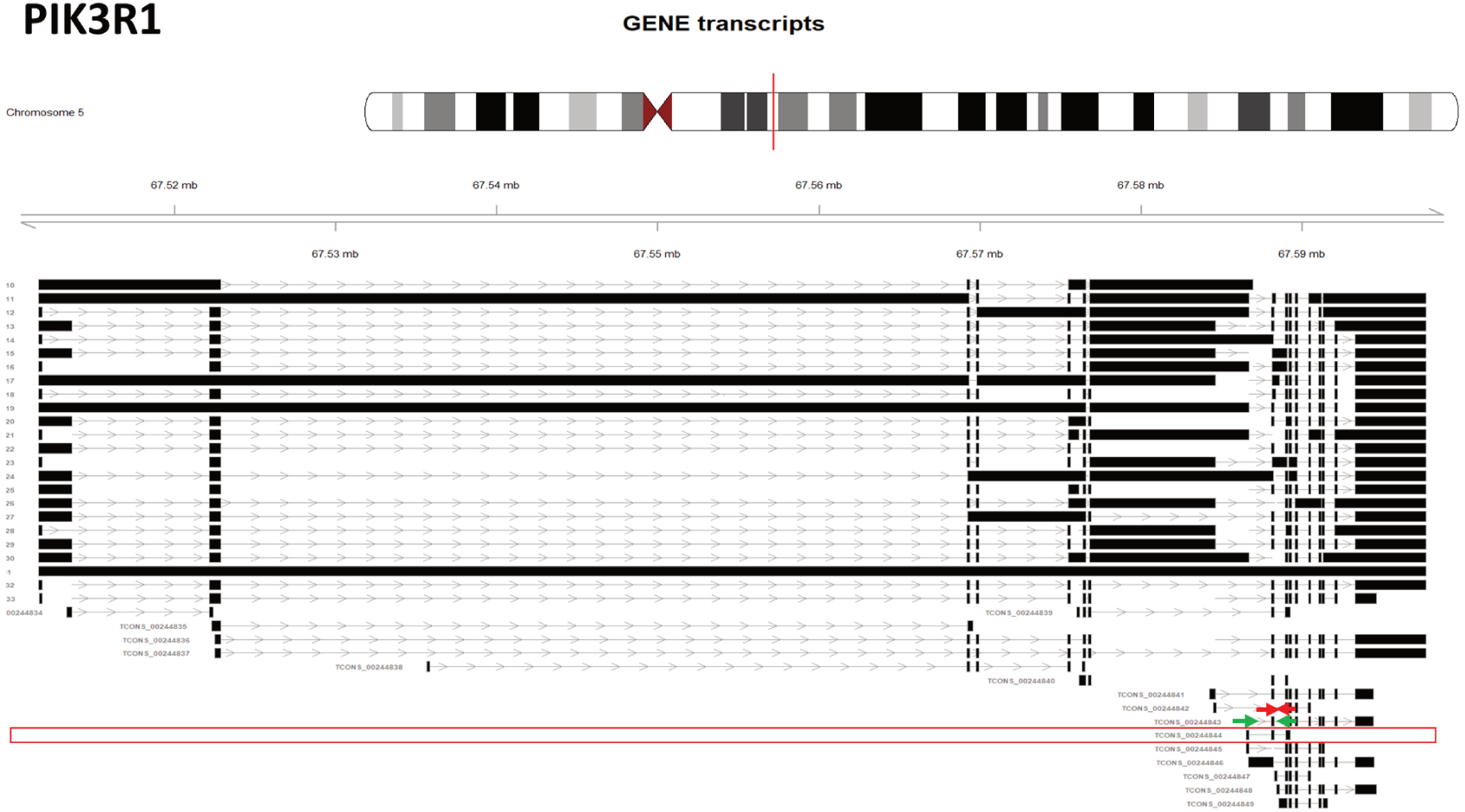
I

MTF2


GENE transcripts
Chromosome 1
93.55 mb
93.57 mb
93.59 mb
93.56 mb
93.58 mb
93.6 mb

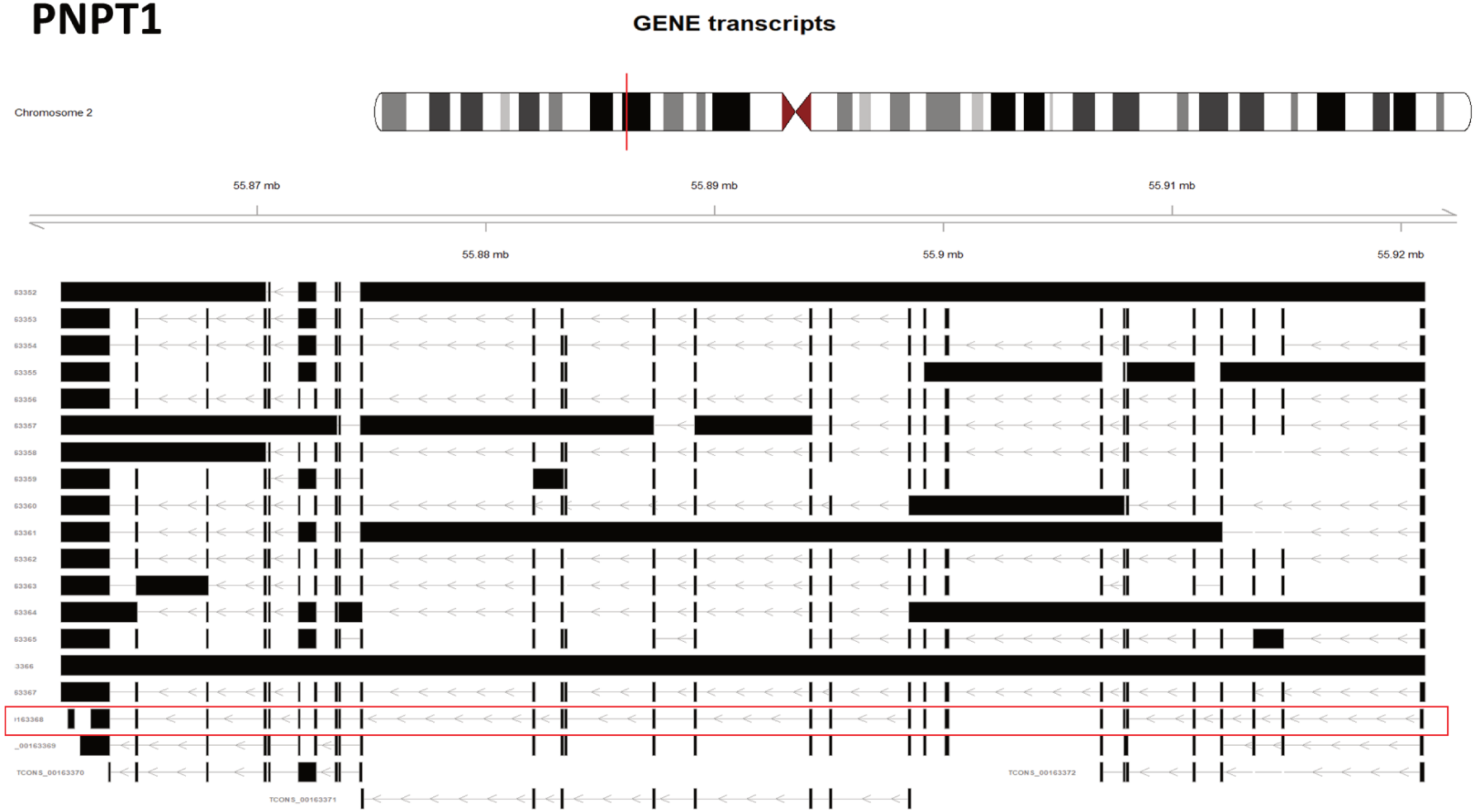
J

PIK3R1


GENE transcripts
Chromosome 5
67.52 mb
67.54 mb
67.56 mb
67.58 mb
67.53 mb
67.55 mb
67.57 mb
67.59 mb
TCONS_00244834
TCONS_00244835
TCONS_00244836
TCONS_00244837
TCONS_00244838
TCONS_00244839
TCONS_00244840
TCONS_00244841
TCONS_00244842
TCONS_00244843
TCONS_00244844
TCONS_00244845
TCONS_00244846
TCONS_00244847
TCONS_00244848
TCONS_00244849

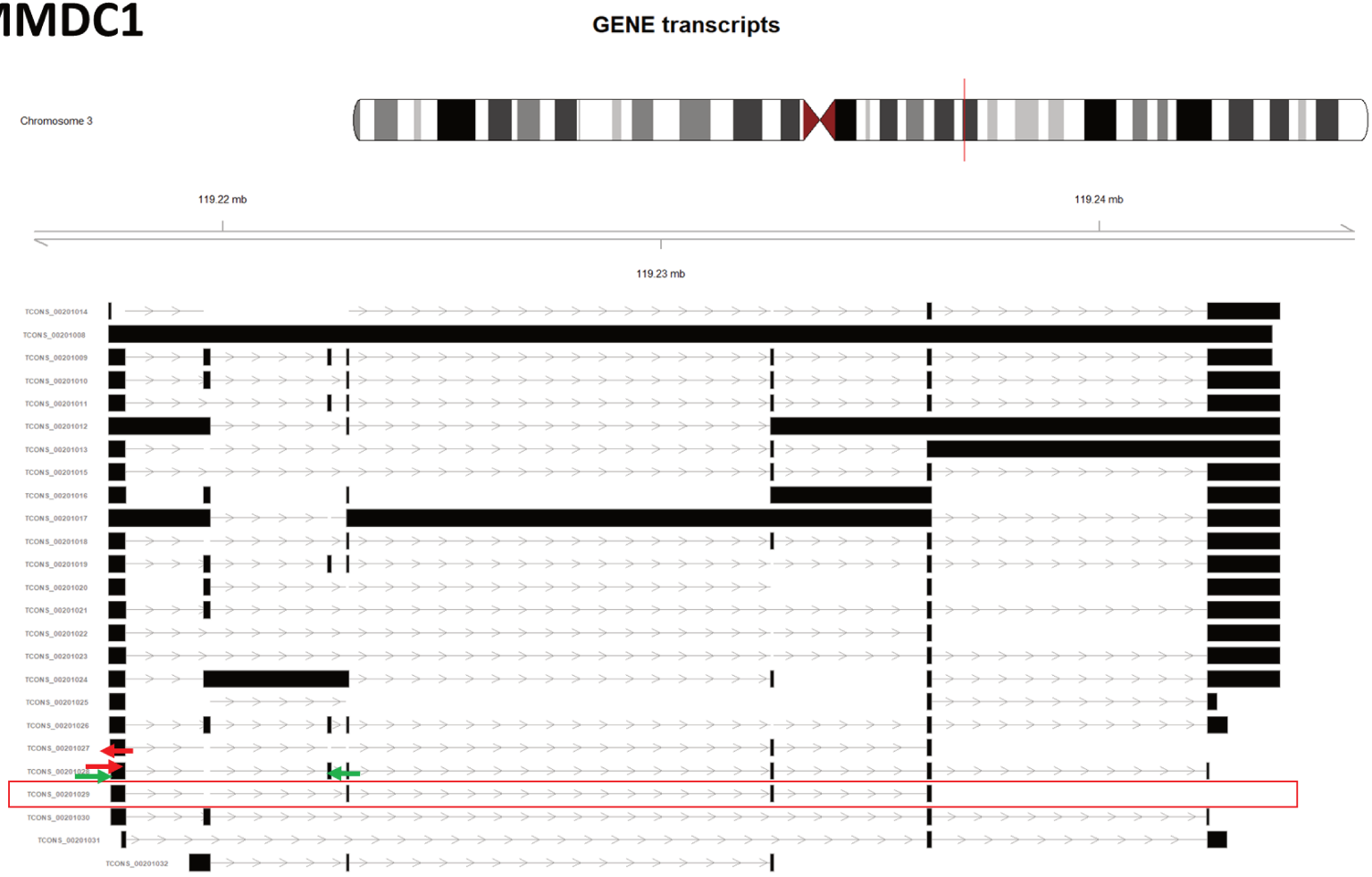
K

PNPT1


GENE transcripts
Chromosome 2
55.87 mb
55.89 mb
55.91 mb
55.88 mb
55.9 mb
55.92 mb
63352
63353
63354
63355
63356
63357
63358
63359
63360
63361
63362
63363
63364
63365
3366
63367
163368
_00163369
TCONS_00163370
TCONS_00163371
TCONS_00163372

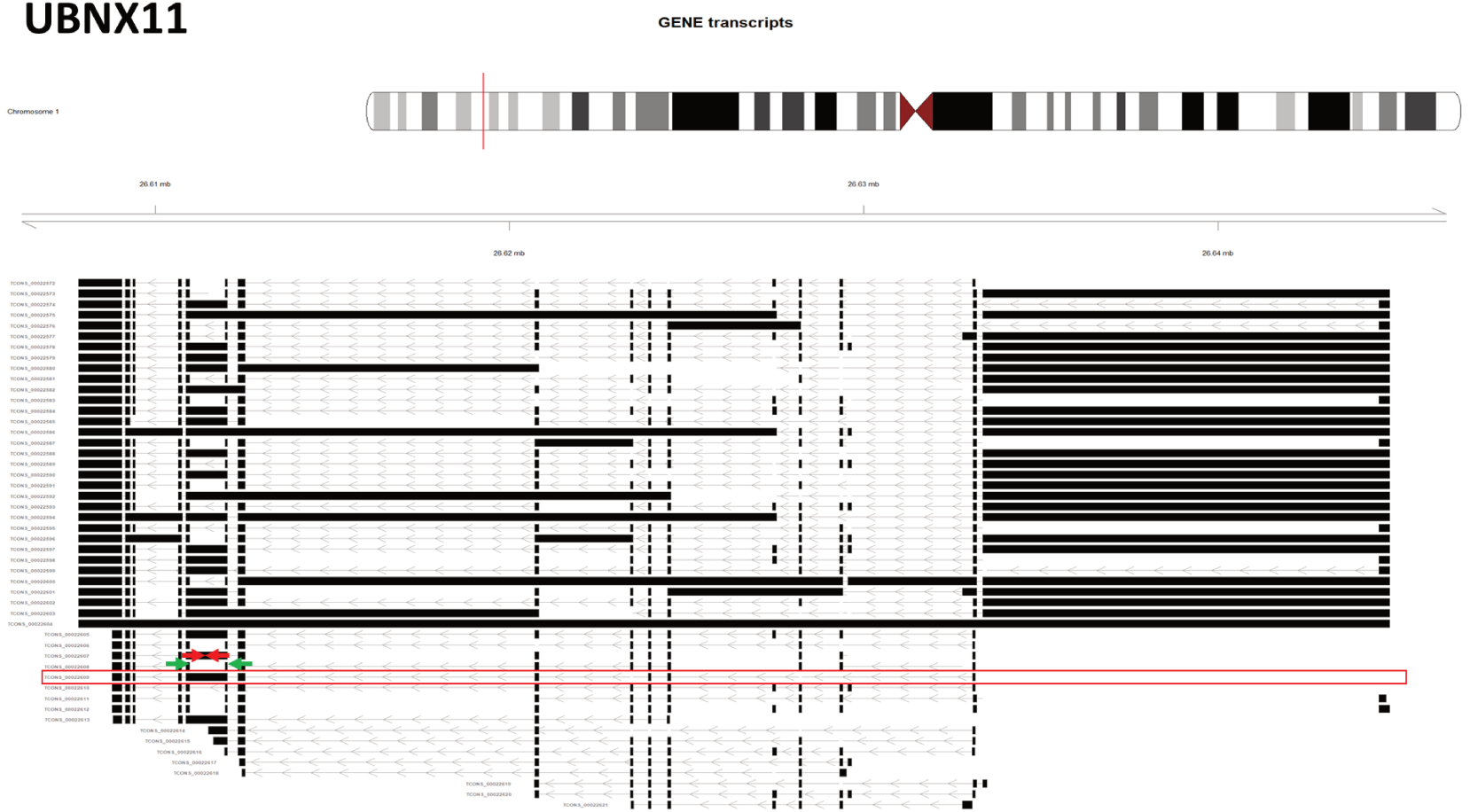
L

TIMMDC1

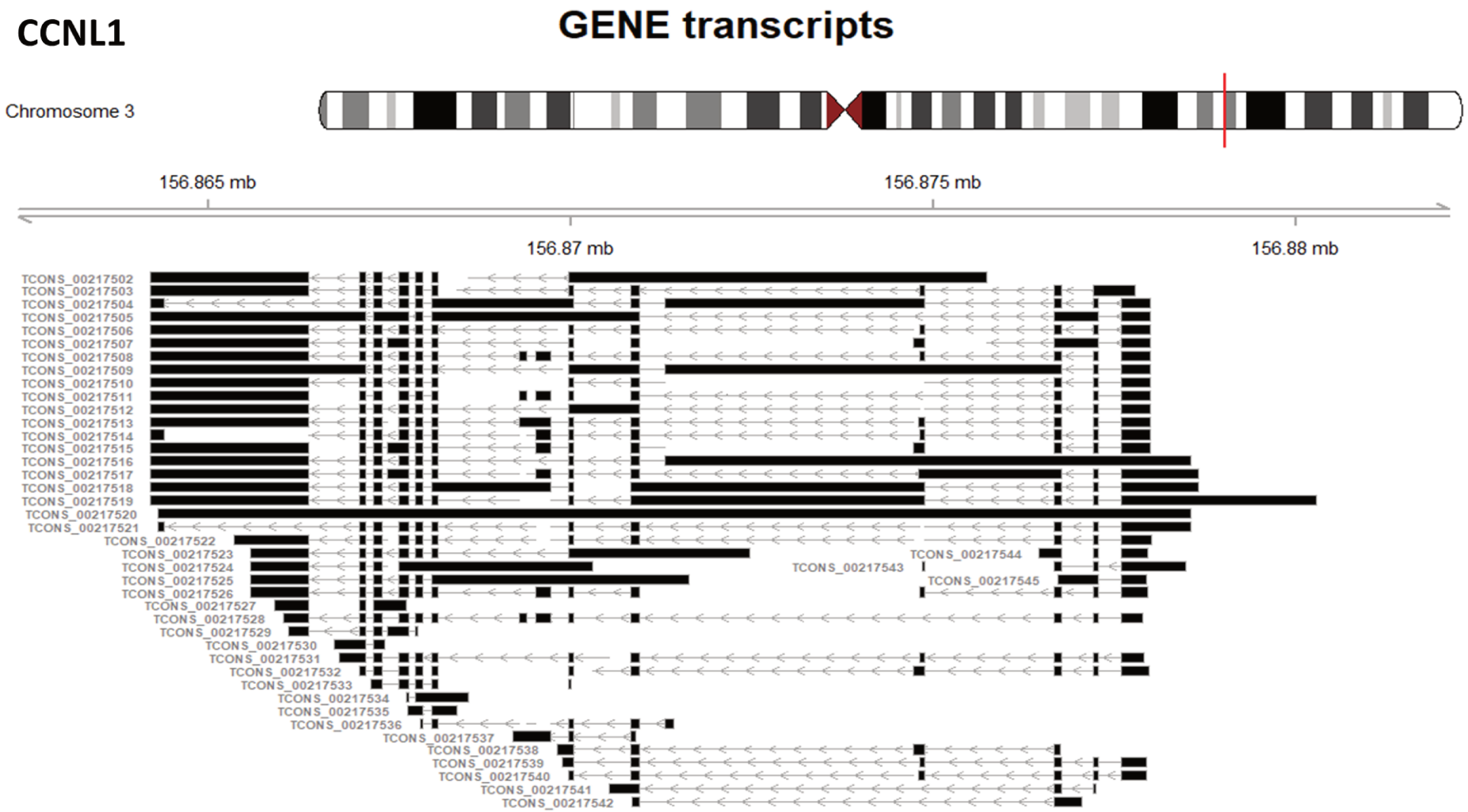

GENE transcripts
Chromosome 3
119.22 mb
119.24 mb
119.23 mb
TCONS_00201014
TCONS_00201008
TCONS_00201009
TCONS_00201010
TCONS_00201011
TCONS_00201012
TCONS_00201013
TCONS_00201015
TCONS_00201016
TCONS_00201017
TCONS_00201018
TCONS_00201019
TCONS_00201020
TCONS_00201021
TCONS_00201022
TCONS_00201023
TCONS_00201024
TCONS_00201025
TCONS_00201026
TCONS_00201027
TCONS_00201029
TCONS_00201030
TCONS_00201031
TCONS_00201032

M

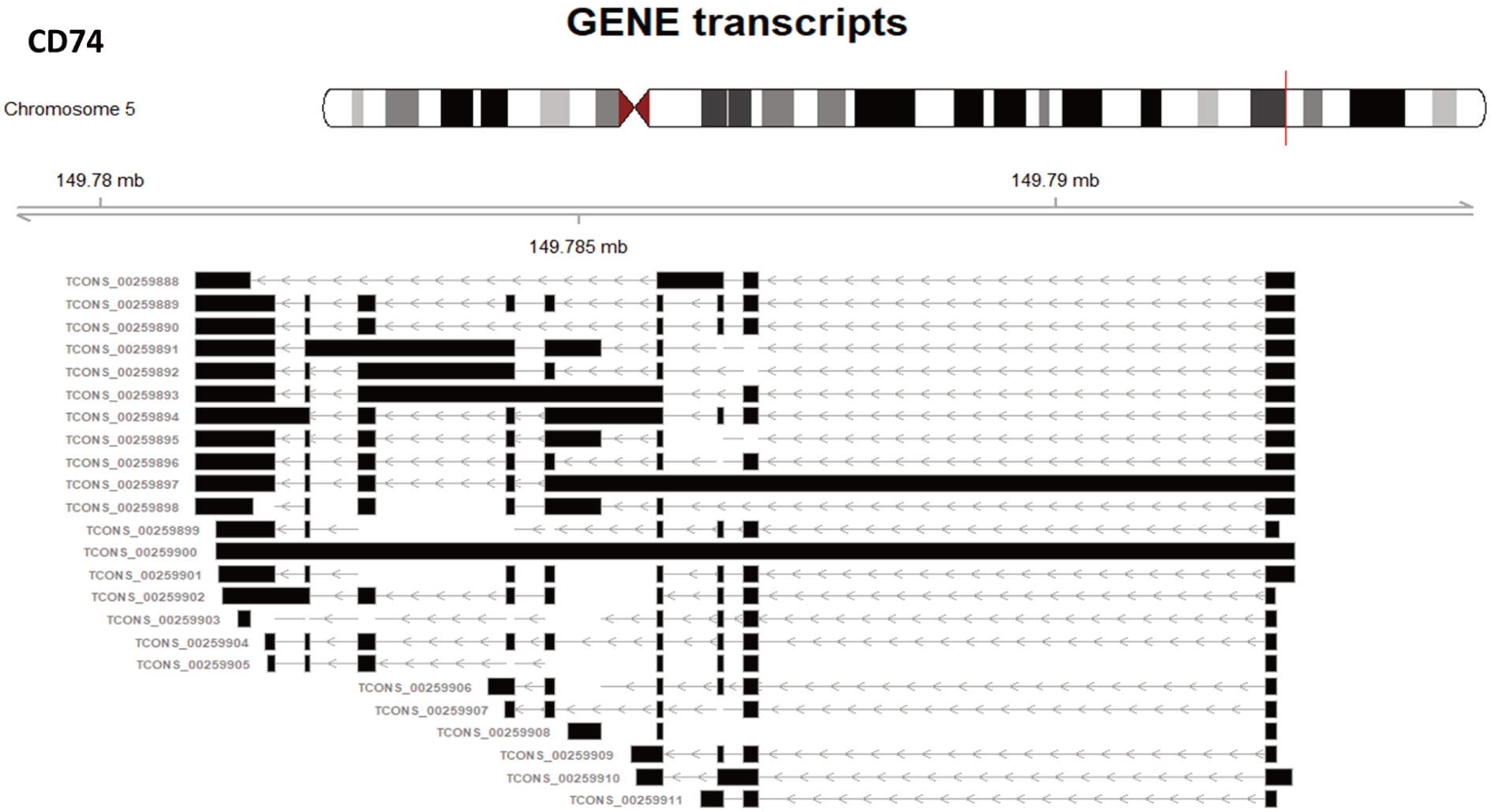
UBNX11


GENE transcripts
Chromosome 1
26.61 mb
26.63 mb
26.62 mb
26.64 mb

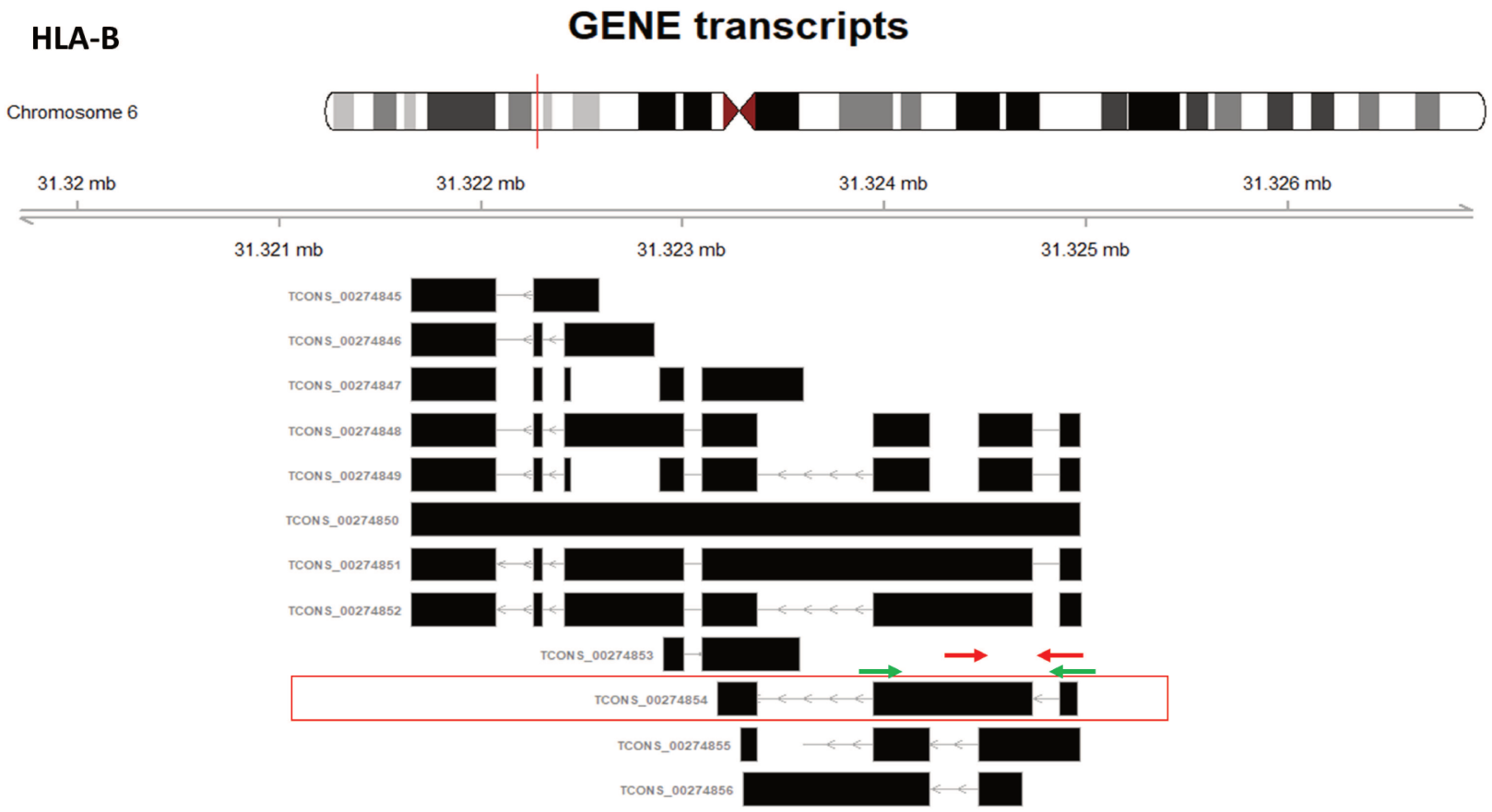
N


CCNL1
GENE transcripts
Chromosome 3
156.865 mb
156.875 mb
156.87 mb
156.88 mb
TCONS_00217502
TCONS_00217503
TCONS_00217504
TCONS_00217505
TCONS_00217506
TCONS_00217507
TCONS_00217508
TCONS_00217509
TCONS_00217510
TCONS_00217511
TCONS_00217512
TCONS_00217513
TCONS_00217514
TCONS_00217515
TCONS_00217516
TCONS_00217517
TCONS_00217518
TCONS_00217519
TCONS_00217520
TCONS_00217521
TCONS_00217522
TCONS_00217523
TCONS_00217524
TCONS_00217525
TCONS_00217526
TCONS_00217527
TCONS_00217528
TCONS_00217529
TCONS_00217530
TCONS_00217531
TCONS_00217532
TCONS_00217533
TCONS_00217534
TCONS_00217535
TCONS_00217536
TCONS_00217537
TCONS_00217538
TCONS_00217539
TCONS_00217540
TCONS_00217541
TCONS_00217542
TCONS_00217543
TCONS_00217544
TCONS_00217545

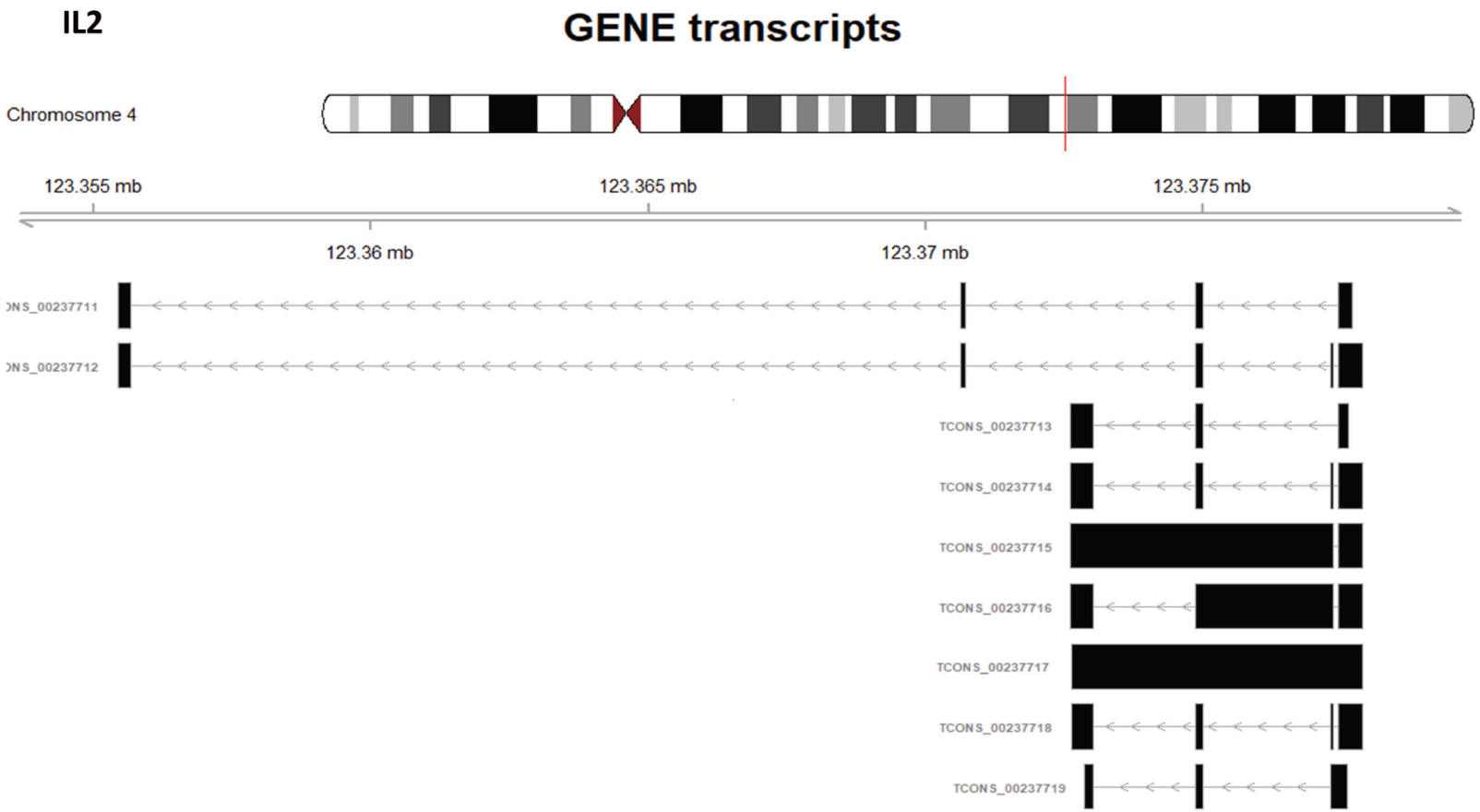
o


CD74
GENE transcripts
Chromosome 5
149.78 mb
149.79 mb
149.785 mb
TCONS_00259888
TCONS_00259889
TCONS_00259890
TCONS_00259891
TCONS_00259892
TCONS_00259893
TCONS_00259894
TCONS_00259895
TCONS_00259896
TCONS_00259897
TCONS_00259898
TCONS_00259899
TCONS_00259900
TCONS_00259901
TCONS_00259902
TCONS_00259903
TCONS_00259904
TCONS_00259905
TCONS_00259906
TCONS_00259907
TCONS_00259908
TCONS_00259909
TCONS_00259910
TCONS_00259911

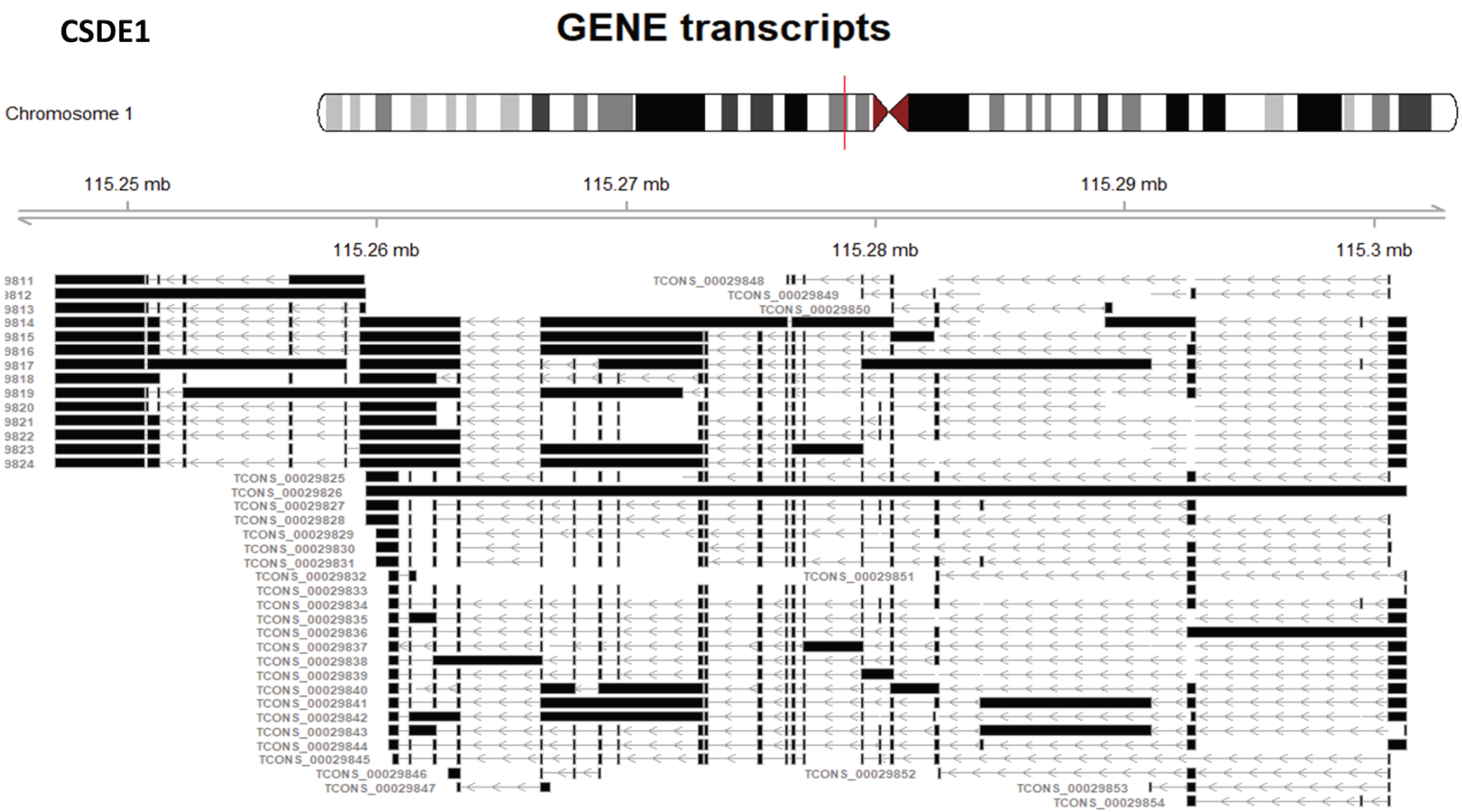
P


HLA-B
GENE transcripts
Chromosome 6
31.32 mb
31.322 mb
31.324 mb
31.326 mb
31.321 mb
31.323 mb
31.325 mb
TCONS_00274845
TCONS_00274846
TCONS_00274847
TCONS_00274848
TCONS_00274849
TCONS_00274850
TCONS_00274851
TCONS_00274852
TCONS_00274853
TCONS_00274854
TCONS_00274855
TCONS_00274856

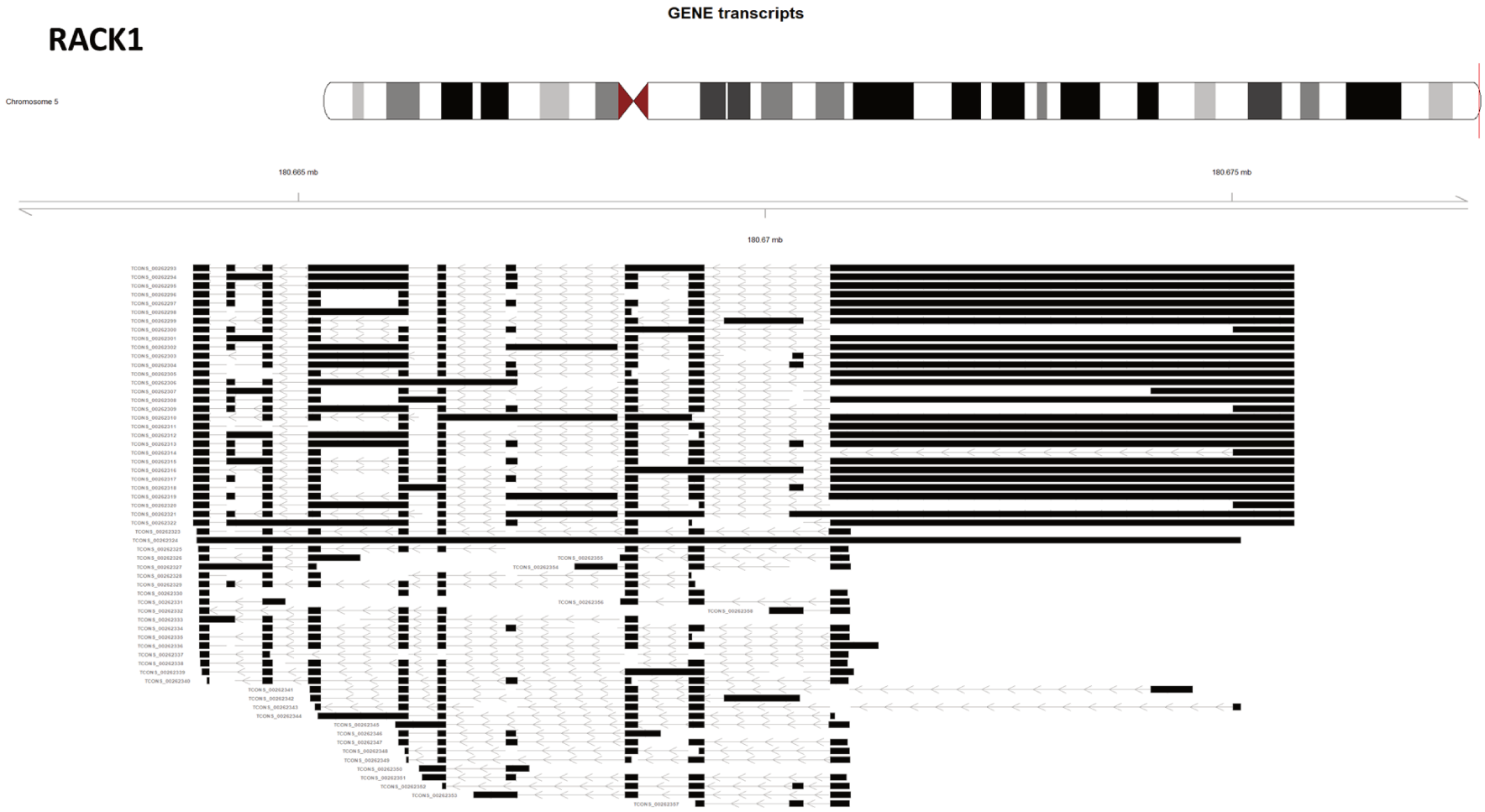
Q


IL2
GENE transcripts
Chromosome 4
123.355 mb
123.365 mb
123.375 mb
123.36 mb
123.37 mb
ONS_00237711
ONS_00237712
TCONS_00237713
TCONS_00237714
TCONS_00237715
TCONS_00237716
TCONS_00237717
TCONS_00237718
TCONS_00237719

R

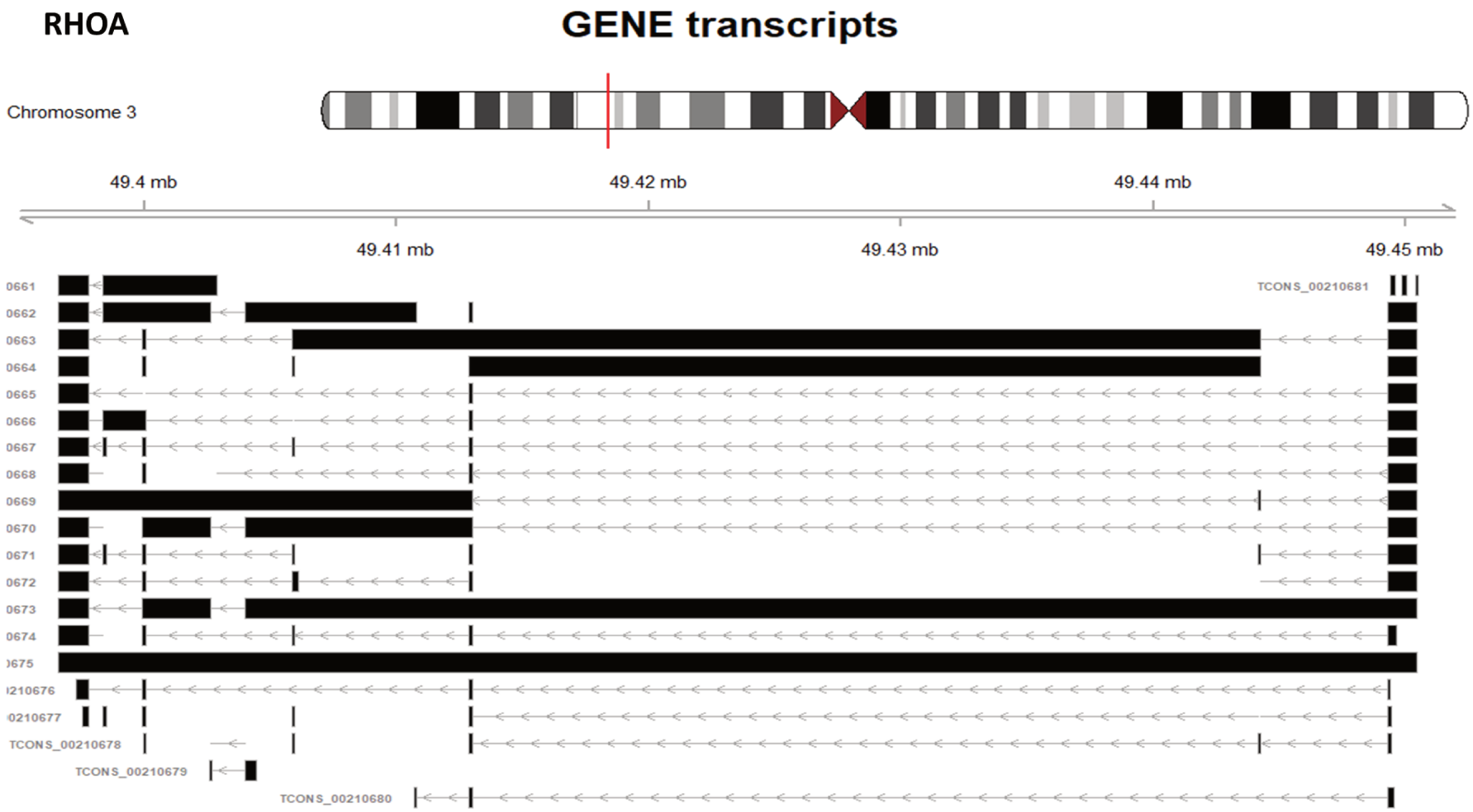

CSDE1
GENE transcripts
Chromosome 1
115.25 mb
115.27 mb
115.29 mb
115.26 mb
115.28 mb
115.3 mb
9811
9812
9813
9814
9815
9816
9817
9818
9819
9820
9821
9822
9823
9824
TCONS_00029825
TCONS_00029826
TCONS_00029827
TCONS_00029828
TCONS_00029829
TCONS_00029830
TCONS_00029831
TCONS_00029832
TCONS_00029833
TCONS_00029834
TCONS_00029835
TCONS_00029836
TCONS_00029837
TCONS_00029838
TCONS_00029839
TCONS_00029840
TCONS_00029841
TCONS_00029842
TCONS_00029843
TCONS_00029844
TCONS_00029845
TCONS_00029846
TCONS_00029847
TCONS_00029848
TCONS_00029849
TCONS_00029850
TCONS_00029851
TCONS_00029852
TCONS_00029853
TCONS_00029854

T

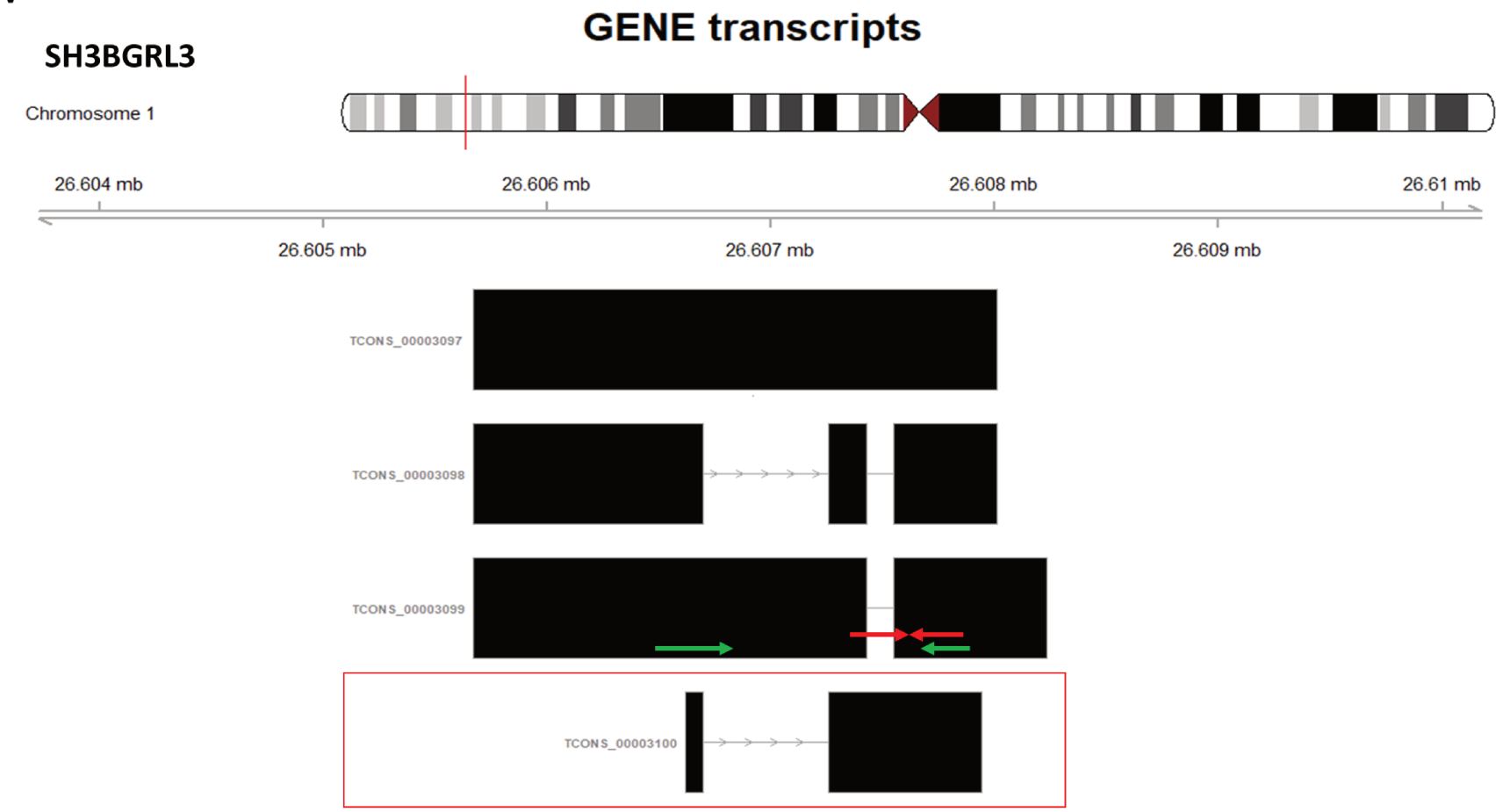

RACK1
GENE transcripts
Chromosome 5
180.665 mb
180.675 mb
180.67 mb

U

RHOA


GENE transcripts
Chromosome 3
49.4 mb
49.42 mb
49.44 mb
49.41 mb
49.43 mb
49.45 mb
TCONS_00210681
TCONS_00210678
TCONS_00210679
TCONS_00210680

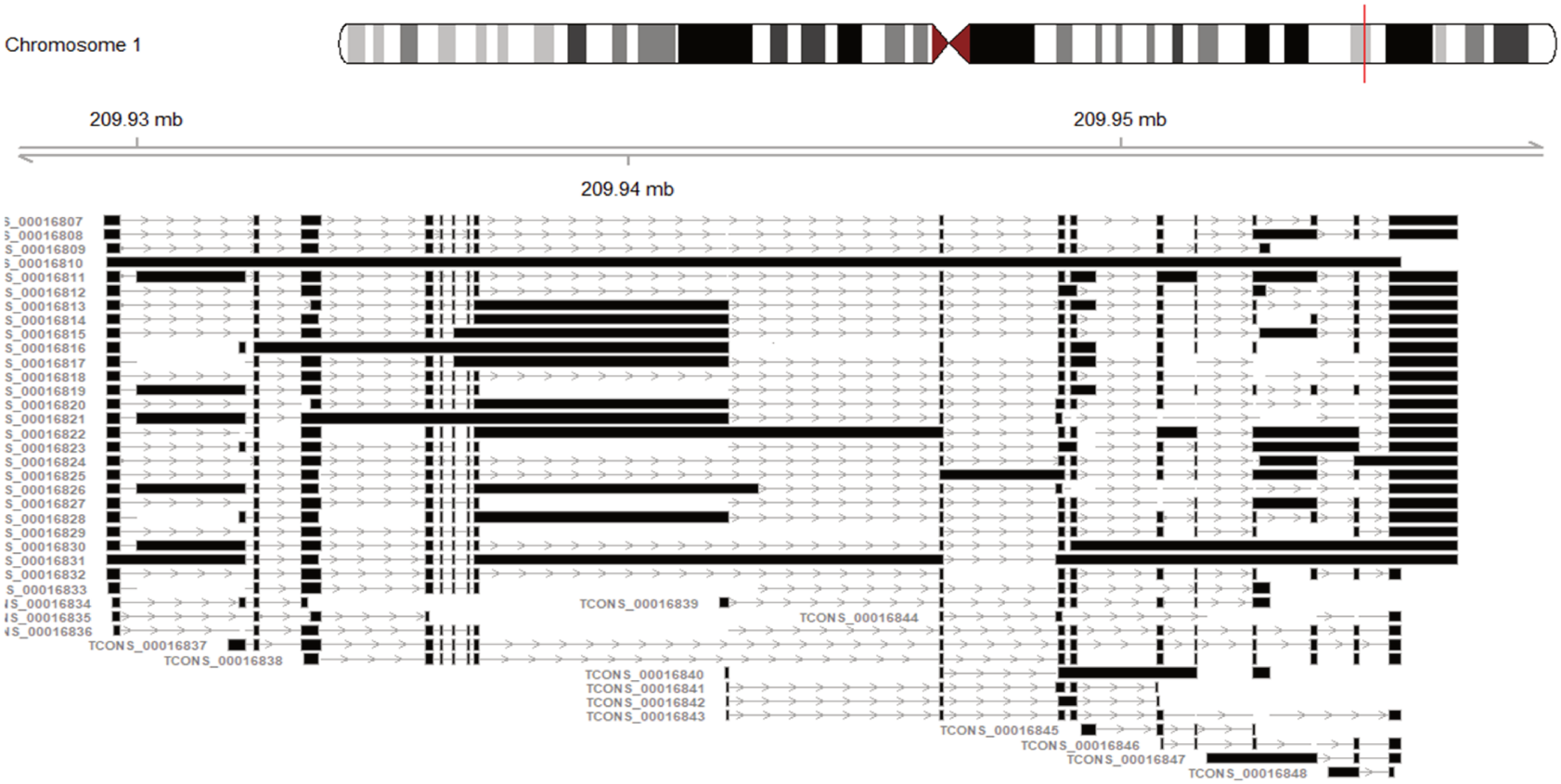
V


SH3BGRL3
GENE transcripts
Chromosome 1
26.604 mb
26.606 mb
26.608 mb
26.61 mb
26.605 mb
26.607 mb
26.609 mb
TCONS_00003097
TCONS_00003098
TCONS_00003099
TCONS_00003100

X

TRAF3IP3

GENE transcripts



9.3 ANEXO C

CENTRO INFANTIL DE INVESTIGAÇÕES HEMATOLÓGICAS DR.

Plataforma Brasil

PARECER CONSUBSTANCIADO DO CEP

DADOS DA EMENDA

Título da Pesquisa: Estudo pré-clinico "in vitro" e em modelos murinos de estratégias de imunoengenharia possuidoras de potencial para o futuro desenvolvimento de tratamentos para a leucemia linfóide aguda da criança

Pesquisador: Pedro Otavio de Campos Lima

Área Temática:

Versão: 4

CAAE: 36599414.9.0000.5376

Instituição Proponente: Centro Infantil de Investigações Hematológicas Dr.Domingos A Boldrini

Patrocinador Principal: FUNDACAO DE AMPARO A PESQUISA DO ESTADO DE SAO PAULO

DADOS DO PARECER

Número do Parecer: 2.553.710

Apresentação do Projeto:

Inclusão de subprojeto ao projeto CAAE:36599414.9.0000.5376 já em andamento. Serão utilizadas células derivadas das mesmas amostras dos 24 doadores saudáveis a serem empregadas no projeto original, para simultaneamente estudar o processo de ativação dos linfócitos. Será utilizado o mesmo tipo de célula, a saber, linfócitos do sangue de doadores normais. O volume das amostras sanguíneas, os riscos, os potenciais benefícios e o número de doadores mantêm-se os mesmos.

Objetivo da Pesquisa:

O subprojeto estenderá e complementará o trabalho desenvolvido no contexto do projeto original que desenvolve um método de imunoterapia para leucemia linfóide aguda, baseado em um receptor quimérico anti-CD19, bem como na programação metabólica de linfócitos autólogos e na modulação da sinalização desses linfócitos.

) objetivo específico do subprojeto de que se trata o adendo é melhor entender o processo de ativação do linfócito.

Peço, portanto, a inclusão do subprojeto "Estudo da Ativação do Linfócito: Papel Biológico da Retenção

Endereço: Rua Dr. Gabriel Porto, 1270 Cidade Universitária
Bairro: Barão Geraldo **CEP:** 13.083-210
UF: SP **Município:** CAMPINAS
Telefone: (19)3787-5001 **Fax:** (19)3289-3571 **E-mail:** cep@boldrini.org.br

CENTRO INFANTIL DE INVESTIGAÇÕES HEMATOLÓGICAS DR.

Intrônica como Splicing Alternativo", referente ao trabalho experimental da tese de mestrado do estudante Pedro
Faria Zeni, como adendo ao projeto CAAE: 36599414.9.0000.5376, já aprovado pelo CEP-Boldrini. Pede-se igualmente a dispensa de novo TCLE, uma vez que os doadores serão os mesmos e os projetos têm a mesma
finalidade e uso de amostras.

Avaliação dos Riscos e Benefícios:
Não há qualquer risco adicional ao acrescentar-se o subprojeto como adendo ao projeto original e integra-se no objetivo global do projeto principal, uma vez que o conhecimento sobre o funcionamento do linfócito auxilia no desenvolvimento de munoterapias celulares.

Comentários e Considerações sobre a Pesquisa:
Subprojeto relevante para auxiliar a compreensão de munoterapias celulares do projeto original.

Considerações sobre os Termos de apresentação obrigatória:
O TCLE para os doadores saudáveis será o mesmo já utilizado e é plenamente adequado.

Conclusões ou Pendências e Lista de Inadequações:
Projeto complementar aprovado para compor o projeto original.

Considerações Finais a critério do CEP:

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_646431_E1.pdf	23/02/2018 15:12:55		Aceito
Outros	SUBPROJETO_ADENDO_ZENI_CEP.pdf	23/02/2018 14:23:51	Pedro Otavio de Campos Lima	Aceito
Outros	CARTA_AO_CEP_ADENDO_ZENI_v2.pdf	23/02/2018 14:16:01	Pedro Otavio de Campos Lima	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE-Pais-Responsaveis dos pacientes com LLA-REVISADO.pdf	10/11/2014 18:28:57		Aceito
Outros	EMENDA-CARTA RESPOSTA AO CEP-ProjFAPESP v4.pdf	10/11/2014 18:25:28		Aceito

Endereço: Rua Dr. Gabriel Porto, 1270 Cidade Universitária
Bairro: Barão Geraldo **CEP:** 13.083-210
UF: SP **Município:** CAMPINAS
Telefone: (19)3787-5001 **Fax:** (19)3289-3571 **E-mail:** cep@boldrini.org.br

CENTRO INFANTIL DE INVESTIGAÇÕES HEMATOLÓGICAS DR.

Plataforma Brasil

Continuação do Parecer: 2.553.710

Outros	CARTA DE RESPOSTA AO CEP-Projeto FAPESP v2.pdf	16/10/2014 20:01:26		Aceito
Outros	Termo de Assentimento-FAPESP.pdf	16/10/2014 20:00:02		Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE-Pais-Responsaveis dos pacientes com LLA.pdf	16/10/2014 19:58:38		Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE-FAPESP-DOADOR SAUDAVEL-v2.pdf	16/10/2014 19:58:11		Aceito
Projeto Detalhado / Brochura Investigador	PROJETO FAPESP-MET62ce.pdf	16/10/2014 19:56:53		Aceito
Declaração de Pesquisadores	Documento comprobatorio do credenciamento do serviço que abrigara animais.pdf	16/10/2014 19:55:23		Aceito
Outros	Despacho de avaliacao cientifica-FAPESP.pdf	23/09/2014 20:52:46		Aceito
Folha de Rosto	FOLHA DE ROSTO-PODCL.pdf	23/09/2014 19:18:01		Aceito

Situação do Parecer:
Aprovado

Necessita Apreciação da CONEP:
Não

CAMPINAS, 21 de Março de 2018

Assinado por:
Maristela Amaral Palazzi
(Coordenador)

Endereço: Rua Dr. Gabriel Porto, 1270 Cidade Universitária
Bairro: Barão Geraldo CEP: 13.083-210
UF: SP Município: CAMPINAS
Telefone: (19)3787-5001 Fax: (19)3289-3571 E-mail: cep@boldrini.org.br

9.4 ANEXO D

Declaração

As cópias de artigos de minha autoria ou de minha co-autoria, já publicados ou submetidos para publicação em revistas científicas ou anais de congressos sujeitos a arbitragem, que constam da minha Dissertação/Tese de Mestrado/Doutorado, intitulada **Estudo da ativação do linfócito: papel da retenção intrônica como splicing alternativo**, não infringem os dispositivos da Lei n.° 9.610/98, nem o direito autoral de qualquer editora.

Campinas, 15 de junho de 2018

Assinatura : Pedro Faria Zeni
Nome do(a) autor(a): **Pedro Faria Zeni**
RG n.° 3.218.860-9

Assinatura :
Nome do(a) orientador(a): **Pedro Otavio de Campos Lima**
RG n.° 2.109.195