

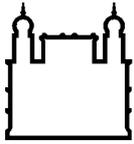
MINISTÉRIO DA SAÚDE
FUNDAÇÃO OSWALDO CRUZ
INSTITUTO OSWALDO CRUZ

Mestrado em Programa de Pós-Graduação Biologia Computacional e Sistemas

ANÁLISE COMPARATIVA DA DIVERSIDADE DE SPLICING
ALTERNATIVO NO PROTEOMA DO CÉREBRO HUMANO E MURINO

ESDRAS MATHEUS GOMES DA SILVA

Rio de Janeiro
Junho de 2018



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

Programa de Pós-Graduação em Biologia Computacional e Sistemas

ESDRAS MATHEUS GOMES DA SILVA

Análise Comparativa da Diversidade de *Splicing* Alternativo no Proteoma do Cérebro Humano e Murino

Dissertação apresentada ao Instituto Oswaldo Cruz como parte dos requisitos para obtenção do título de Mestre em Biologia Computacional e Sistemas

Orientador: Prof. Dr. Fabio Passetti

RIO DE JANEIRO

Junho de 2018

Gomes da Silva , Esdras Matheus .

Análise Comparativa da Diversidade de Splicing Alternativo no Proteoma do Cérebro Humano e Murino / Esdras Matheus Gomes da Silva . - Rio de Janeiro, 2018.

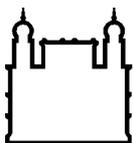
120 f.; il.

Dissertação (Mestrado) - Instituto Oswaldo Cruz, Pós-Graduação em Biologia Computacional e Sistemas, 2018.

Orientador: Fabio Passetti.

Bibliografia: f. 107-119

1. Proteogenômica . 2. Espectrometria de massas . 3. Repositório Personalizado. I. Título.



Ministério da Saúde

FIOCRUZ
Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

Programa de Pós-Graduação em Biologia Computacional e Sistemas

AUTOR: ESDRAS MATHEUS GOMES DA SILVA

ANÁLISE COMPARATIVA DA DIVERSIDADE DE *SPLICING* ALTERNATIVO NO PROTEOMA DO CÉREBRO HUMANO E MURINO

ORIENTADOR: Prof. Dr. Fabio Passeti

Aprovada em: 29 / 06 / 2018

EXAMINADORES:

Prof. Dr. Jerônimo Conceição Ruiz - *Presidente* (Fiocruz/IRR)

Prof. Dr. Antonio Basilio de Miranda (IOC/Fiocruz)

Prof. Dr. Paulo Sergio Lopes de Oliveira (LNBio)

Prof. Dr. Ana Gisele da Costa Neves Ferreira (IOC/Fiocruz)

Prof. Dr. Diogo Antonio Tschoeke (NUPEM/UFRJ.)

Rio de Janeiro, 29 de junho de 2018

Dedico esta dissertação a todos que venham ser impactados com os resultados que ela apresenta.

AGRADECIMENTOS

À minha família, por me fornecer apoio emocional e por representar um porto seguro em meio ao mar conturbado que a vida acadêmica por vezes se assemelha.

À Shayenne Olsson Freitas Silva, por me motivar durante esta etapa da minha vida acadêmica. Por acreditar em meu potencial, até mesmo nos momentos em que eu não acreditava. Por dividir todos os momentos bons e difíceis que esta jornada apresentou. Você foi parte fundamental para que eu concluísse este ciclo, nunca me esquecerei do seu papel.

Ao Fabio Freitas Silva e à Cristiana Cabreiro por me acolherem com carinho durante boa parte da minha estadia na cidade do Rio de Janeiro.

Ao meu orientador, Prof. Dr. Fabio Passetti, pelo amparo, paciência, apoio e auxílio durante todo o processo de realização deste trabalho e por garantir os recursos e infraestrutura necessários para a condução e conclusão do projeto.

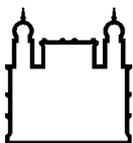
Ao Dr. Raphael Tavares, por colaborar de maneira direta para conclusão desta dissertação.

A todos do Laboratório de Genômica Funcional e Bioinformática do Instituto Oswaldo Cruz, pela ajuda durante o desenvolvimento deste projeto e também pelos momentos de descontração.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES, pelo auxílio financeiro.

A ciência é muito mais do que um corpo de conhecimento. É uma forma de pensar. Uma maneira cética de interrogar o universo com pleno entendimento da falibilidade humana.

Carl Sagan



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

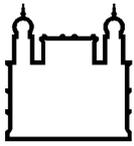
ANÁLISE COMPARATIVA DA DIVERSIDADE DE SPLICING ALTERNATIVO NO PROTEOMA DO CÉREBRO HUMANO E MURINO

RESUMO

DISSERTAÇÃO DE MESTRADO EM BIOLOGIA COMPUTACIONAL E SISTEMAS

Esdras Matheus Gomes da Silva

Os avanços obtidos em transcriptômica, em função do desenvolvimento de sequenciadores de alta vazão, e na proteômica, por meio dos modernos espectrômetros de massas (MS), resultaram em um grande volume de dados que passou a ser integrado em diversos estudos de Bioinformática, levando ao melhor entendimento sobre a fração dos RNAs mensageiros efetivamente traduzida em proteínas. A proteogenômica é a área de pesquisa que reúne estas tecnologias, atuando na interface entre a genômica e a proteômica para interpretar eventos moleculares, tais como, por exemplo, o *splicing* alternativo. Este evento molecular é capaz de gerar RNAs mensageiros diferentes a partir de um mesmo gene, podendo alterar a sequência polipeptídica e, conseqüentemente, gerar proteoformas com funções distintas. Neste sentido, atualmente alguns projetos têm realizado esta análise integrativa com o intuito de comparar os resultados de amostras de ser humano e outros mamíferos, uma vez que alguns destes animais são utilizados como organismos modelo para o estudo dos aspectos moleculares de doenças, como as neurodegenerativas. Desta forma, este projeto teve como objetivo principal analisar o perfil de expressão de variantes de *splicing* alternativo em dados de espectrometria de massas de proteínas de amostras de tecidos de cérebros saudáveis de humano e camundongo. Para tal, foram utilizados dados de mRNAs de referência (Refseq), ESTs e sequências da base de dados Uniprot/Swiss-Prot para confecção de repositórios de sequências proteicas personalizados, utilizando uma metodologia desenvolvida pelo nosso grupo de pesquisa denominada matrizes ternárias. O repositório personalizado para humano continha 20.150 sequências canônicas e 204.294 peptídeos não redundantes, totalizando 224.453 sequências. O repositório de camundongo possuía 16.888 sequências canônicas e 156.889 peptídeos não redundantes, totalizando 173.777 sequências. Estes repositórios de sequências proteicas personalizados foram empregados para a análise de dados de espectrometria de massas de três regiões distintas do cérebro de humano e camundongo (corpo caloso, nervo óptico e bulbo olfatório). A partir destas análises, nós inferimos a expressão de um total de 3.289 proteínas canônicas e 23 proteoformas de genes ortólogos entre humano e camundongo. Dentre as proteoformas, seis foram inferidas a partir de peptídeos proteotípicos idênticos identificados em dados de MS de humano e camundongo (PKM, CRMP1, PRKCB, STXBP1, CADM1 e HNRNPK). Portanto, acreditamos que a identificação de peptídeos compartilhados entre humano e camundongo, pertencentes a proteoformas de genes ortólogos, realizada neste projeto, contribuiu para o melhor conhecimento da diversidade de *splicing* do cérebro de humano e camundongo.



Ministério da Saúde

FIOCRUZ
Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

COMPARATIVE ANALYSIS OF ALTERNATIVE SPLICING DIVERSITY IN HUMAN AND MOUSE BRAIN PROTEOME

ABSTRACT

MASTER DISSERTATION IN COMPUTATIONAL BIOLOGY AND SYSTEMS

Esdras Matheus Gomes da Silva

The advances in transcriptomics, due to the development of high throughput sequencers and in proteomics, using modern mass spectrometers (MS), resulted in a large volume of data that has been integrated into several Bioinformatics studies, leading to a better understanding of messenger RNAs fraction that is effectively translated into proteins. Proteogenomics is the area of research that brings together these technologies, acting at the interface between genomics and proteomics to interpret molecular events such as alternative splicing. This molecular event can generate different messenger RNAs from the same gene, which may alter the polypeptide sequence and, consequently, generate proteoforms with different functions. Currently, some projects have performed this integrative analysis with the aim of comparing the results derived from human and other mammal samples, since some of these animals are used as model organisms for the study of molecular aspects of diseases, such as neurodegenerative diseases. Thus, the main goal of this project was to analyze the expression profile of alternative splicing variants in protein mass spectrometry data from tissue samples from healthy human and mouse brains. For this, data from reference mRNAs (Refseq), ESTs and sequences from the Uniprot/Swiss-Prot database were used to make custom protein sequence repositories using a methodology developed by our research group called ternary matrices. The human custom repository contained 20,150 canonical sequences and 204,294 non-redundant peptides, summing up to 224,453 sequences. The mouse custom repository had 16,888 canonical sequences and 156,889 non-redundant peptides, giving a total of 173,777 sequences. These custom protein sequence repositories were used for the analysis of mass spectrometric data from three distinct regions of the human and mouse brain (corpus callosum, optic nerve and olfactory bulb). From these analyzes, we inferred the expression of a total of 3,289 canonical proteins and 23 proteoforms of orthologous genes between human and mouse. Among the proteoforms, six were inferred from identical proteotypic peptides identified in human and mouse MS data (PKM, CRMP1, PRKCB, STXBP1, CADM1 and HNRNPK). Therefore, we believe that the identification of shared peptides between human and mouse belonging to proteoforms from orthologous genes carried out in this project contributed to a better understanding of the splicing diversity of human and mouse brain.

ÍNDICE

1	INTRODUÇÃO.....	16
	1.1 O <i>splicing</i> alternativo	16
	1.2 O <i>splicing</i> alternativo e o sistema nervoso	19
	1.3 O sistema nervoso de camundongo e de humano	20
	1.4 A proteômica baseada em espectrometria de massas.....	22
	1.5 A proteogenômica	24
2	OBJETIVOS	29
	Objetivo Geral	29
	Objetivos Específicos.....	29
3	MATERIAL E MÉTODOS.....	30
	3.1 Conjunto de dados	30
	3.2 Matrizes Ternárias	31
	3.3 Construção do repositório de sequências proteicas base Spliceprot.....	32
	3.4 Construção dos repositórios de sequências proteicas personalizados	34
	3.5 Análise dos dados de espectrometria de massas de proteínas	35
	3.6 Busca por proteínas expressas por genes ortólogos entre humano e camundongo	36
	3.7 Comparação das proteínas identificadas em nossas análises com as proteínas inferidas nos artigos de origem dos conjuntos de dados de MS	37
	3.8 Visualização dos eventos de <i>splicing</i> e confecção de figuras	38
4	RESULTADOS	39
	4.1 Repositório base Spliceprot e repositório personalizado de humano	39
	4.2 Proteínas Canônicas.....	39
	4.2.1 Proteínas canônicas de genes ortólogos.....	44
	4.3 Proteoformas	45
	4.3.1 Proteoformas de genes ortólogos.....	47
	4.3.1.1 Proteoformas exclusivas do nervo óptico	53
	4.3.1.2 Proteoformas exclusivas de bulbo olfatório	56
	4.3.1.3 Proteoformas comuns ao nervo óptico e bulbo olfatório.....	79

4.3.1.4	Proteoformas comuns ao nervo óptico e corpo caloso	86
4.3.1.5	Proteoformas comuns as três regiões do cérebro	89
4.3.1.6	Proteoformas de humano e camundongo inferidas por peptídeos proteotípicos idênticos	88
5	DISCUSSÃO.....	96
6	CONCLUSÃO.....	101
7	REFERÊNCIAS BIBLIOGRÁFICAS	102

ÍNDICE DE FIGURAS

Legenda	Página
Figura 1.1 – Representação esquemática dos principais padrões de splicing alternativo.	17
Figura 1.2 – Representação esquemática de um fluxo de análises de proteômica <i>shotgun</i> .	24
Figura 1.3 – Esquema representativo de um fluxo de análises geral para construção de um repositório de sequências proteicas personalizado.	26
Figura 3.1 – Construção de uma matriz ternária.	32
Figura 3.2 – Esquema representativo do fluxo de análises para construção do repositório de sequências proteicas personalizado Spliceprot.	33
Figura 3.3 – Esquema para a construção do repositório proteico personalizado.	34
Figura 3.4 – Esquema do fluxo de análises para identificação de peptídeos e proteínas de genes ortólogos.	37
Figura 4.1 – Diagramas de Venn demonstrando a distribuição dos genes, peptídeos e proteínas entre os tecidos corpo caloso, nervo óptico e bulbo olfatório de humano.	40
Figura 4.2 – Diagramas de Venn demonstrando a distribuição dos genes, peptídeos e proteínas entre os tecidos corpo caloso, nervo óptico e bulbo olfatório de camundongo.	41
Figura 4.3.2 – Diagramas de Venn demonstrando a distribuição das proteínas identificadas pelos autores dos artigos de origem dos dados de MS e das proteínas identificadas em nossas análises em corpo caloso, nervo óptico e bulbo olfatório de camundongo.	43
Figura 4.4 – Diagrama de Venn demonstrando os genes ortólogos de proteínas inferidas nos experimentos de proteômica <i>shotgun</i> dos três tecidos do cérebro de humano e camundongo analisados neste projeto: corpo caloso, nervo óptico e bulbo olfatório.	44
Figura 4.5 – Diagrama de Venn demonstrando os genes ortólogos de proteoformas inferidas nos experimentos de proteômica <i>shotgun</i> das três regiões do cérebro de humano e camundongo analisadas neste projeto: corpo caloso, nervo óptico e bulbo olfatório.	52
Figura 4.6.1 – Esquema representativo das diferenças no padrão de <i>splicing</i> da proteína canônica e proteoforma do gene MBP de humano.	53
Figura 4.6.2 – Esquema representativo das diferenças no padrão de <i>splicing</i> da proteína canônica e proteoforma do gene Mbp de camundongo.	54
Figura 4.7.1 – Esquema representativo das diferenças no padrão de <i>splicing</i> da proteína canônica e proteoforma de <i>splicing</i> do gene PLEC de humano.	55
Figura 4.7.2 – Esquema representativo das diferenças no padrão de <i>splicing</i> da proteína canônica e proteoformas do gene Plec de camundongo.	56
Figura 4.8.1 – Esquema representativo das diferenças no padrão de <i>splicing</i> da proteína canônica e proteoforma do gene TPM2 de humano.	57
Figura 4.8.2 – Esquema representativo das diferenças no padrão de <i>splicing</i> da proteína canônica e proteoforma de <i>splicing</i> do gene Tpm2 de camundongo.	57
Figura 4.9.1 – Esquema representativo das diferenças no padrão de <i>splicing</i> da proteína canônica e proteoforma do gene NDRG4 de humano.	58
Figura 4.9.2 – Esquema representativo das diferenças no padrão de <i>splicing</i> da proteína	

canônica e proteoforma de <i>splicing</i> do gene Ndr4 de camundongo.	59
Figura 4.10 – Esquema representativo das diferenças no padrão de <i>splicing</i> da proteína canônica e proteoforma de <i>splicing</i> do gene HNRNPK de humano e camundongo.	60
Figura 4.11 – Esquema representativo das diferenças no padrão de <i>splicing</i> da proteína canônica e proteoforma de <i>splicing</i> do gene CRMP1 de humano e seu ortólogo em camundongo (Crmp1).	61
Figura 4.12.1 – Esquema representativo das diferenças no padrão de <i>splicing</i> da proteína canônica e proteoformas do gene MACF1 de humano.	62
Figura 4.12.2 – Esquema representativo das diferenças no padrão de <i>splicing</i> da proteína canônica e proteoformas do gene Macf1 de camundongo.	63
Figura 4.13.1 – Esquema representativo das diferenças no padrão de <i>splicing</i> da proteína canônica e proteoformas do gene SYT7 de humano.	64
Figura 4.13.2 – Esquema representativo das diferenças no padrão de <i>splicing</i> da proteína canônica e proteoforma de <i>splicing</i> do gene Syt7 de camundongo.	65
Figura 4.14.1 – Esquema representativo das diferenças no padrão de <i>splicing</i> da proteína canônica e proteoformas do gene MAPT de humano.	66
Figura 4.14.2 – Esquema representativo das diferenças no padrão de <i>splicing</i> da proteína canônica e proteoformas do gene Mapt de camundongo.	67
Figura 4.15.1 – Esquema representativo das diferenças no padrão de <i>splicing</i> da proteína canônica e proteoforma de <i>splicing</i> do gene RTN3 de humano.	68
Figura 4.15.2 – Esquema representativo das diferenças no padrão de <i>splicing</i> da proteína canônica e proteoforma de <i>splicing</i> do gene Rtn3 de camundongo.	68
Figura 4.16.1 – Esquema representativo das diferenças no padrão de <i>splicing</i> da proteína canônica e proteoformas do gene CASK de humano.	69
Figura 4.16.2 – Esquema representativo das diferenças no padrão de <i>splicing</i> da proteína canônica e proteoforma de <i>splicing</i> do gene Cask de camundongo.	70
Figura 4.17 – Esquema representativo das diferenças no padrão de <i>splicing</i> da proteína canônica e proteoformas do gene PRKCB de humano e camundongo.	71
Figura 4.18 – Esquema representativo das diferenças no padrão de <i>splicing</i> da proteína canônica e proteoformas do gene STXBP1 de humano e camundongo.	72
Figura 4.19 – Esquema representativo das diferenças no padrão de <i>splicing</i> da proteína canônica e proteoformas do gene CADM1 de humano e camundongo.	73
Figura 4.20.1 – Esquema representativo das diferenças no padrão de <i>splicing</i> da proteína canônica e proteoformas do gene NFASC de humano.	74
Figura 4.20.2 – Esquema representativo das diferenças no padrão de <i>splicing</i> da proteína canônica e proteoformas do gene Nfasc de camundongo.	75
Figura 4.21.1 – Esquema representativo das diferenças no padrão de <i>splicing</i> da proteína canônica e proteoforma de <i>splicing</i> do gene MAP4 de humano.	76
Figura 4.21.2 – Esquema representativo das diferenças no padrão de <i>splicing</i> da proteína canônica e proteoforma de <i>splicing</i> do gene Map4 de camundongo.	77
Figura 4.22.1 – Esquema representativo das diferenças no padrão de <i>splicing</i> da proteína	

canônica e proteoformas do gene EPB41L3 de humano.	78
Figura 4.22.2 – Esquema representativo das diferenças no padrão de <i>splicing</i> da proteína canônica e proteoformas do gene Pb41l3 de camundongo.	78
Figura 4.23 – Esquema representativo das diferenças no padrão de <i>splicing</i> da proteína canônica e proteoformas do gene DMN1 humano e camundongo.	79
Figura 4.24.1 – Esquema representativo das diferenças no padrão de <i>splicing</i> da proteína canônica e proteoforma de <i>splicing</i> do gene SPTAN1 de humano.	80
Figura 4.24.2 – Esquema representativo das diferenças no padrão de <i>splicing</i> da proteína canônica e proteoformas do gene Sptan1 de camundongo.	81
Figura 4.25.1 – Esquema representativo das diferenças no padrão de <i>splicing</i> da proteína canônica e proteoforma de <i>splicing</i> do gene PLEKHB1 de humano.	82
Figura 4.25.2 – Esquema representativo das diferenças no padrão de <i>splicing</i> da proteína canônica e proteoforma de <i>splicing</i> do gene Plekhh1 de humano.	82
Figura 4.26 – Esquema representativo das diferenças no padrão de <i>splicing</i> da proteína canônica e proteoforma de <i>splicing</i> do gene PKM de humano e camundongo.	83
Figura 4.27.1 – Esquema representativo das diferenças no padrão de <i>splicing</i> da proteína canônica e proteoformas do gene TPM1 de humano.	85
Figura 4.27.2 – Esquema representativo das diferenças no padrão de <i>splicing</i> da proteína canônica e proteoformas do gene Tpm1 de camundongo.	86
Figura 4.28.1 – Esquema representativo das diferenças no padrão de <i>splicing</i> da proteína canônica e proteoformas do gene TPM3 de humano.	87
Figura 4.28.2 – Esquema representativo das diferenças no padrão de <i>splicing</i> da proteína canônica e proteoformas do gene Tpm3 de camundongo.	88
Figura 4.29 – Alinhamento entre a proteoforma P05771-2 (<i>Query</i>) do gene PRKCB de humano e a proteoforma P68404-2 (<i>Sbjct</i>) do gene Prkcb de camundongo.	89
Figura 4.30 – Alinhamento entre a proteoforma P61764-2 (<i>Query</i>) do gene STXBP1 de humano e a proteoforma O08599-2 do gene (<i>Sbjct</i>) Stxbp1 de camundongo.	90
Figura 4.31 – Alinhamento entre a proteoforma Q9BY67- 5 (<i>Query</i>) do gene CADM1 de humano e a proteoforma Q8R5M8-4 (<i>Sbjct</i>) do gene Cadm1 de camundongo.	91
Figura 4.32 – Alinhamento entre a proteoforma P61978-3 (<i>Query</i>) do gene HNRNPK de humano e a proteoforma P61979-3 (<i>Sbjct</i>) do gene Hnrnpk de camundongo.	92
Figura 4.33 – Alinhamento entre a proteoforma P14618-2 (<i>Query</i>) do gene PKM de humano e a proteoforma P52480-2 (<i>Sbjct</i>) do gene Pkm de camundongo.	93
Figura 4.34 – Alinhamento entre a proteoforma Q14194-2 (<i>Query</i>) do gene CRMP1 de humano e a proteoforma NP_001129530 (<i>Sbjct</i>) do gene Crmp1 de camundongo.	94

LISTA DE TABELAS

Legenda	Página
Tabela 3.1 – Informações dos dados de proteômica <i>shotgun</i> .	31
Tabela 3.2 – Parâmetros utilizados para busca dos espectros de massas contra o repositório de sequências proteicas personalizado.	35
Tabela 3.3 – Lista de tabelas suplementares contendo as proteínas inferidas nos artigos de origem dos dados de MS.	37
Tabela 4.1 – Quantidade de genes, proteínas canônicas inferidas e peptídeos identificados nos dados de proteômica <i>shotgun</i> de corpo caloso, nervo óptico e bulbo olfatório do cérebro de humano e camundongo.	40
Tabela 4.2 – Quantidade de genes ortólogos, proteínas canônicas inferidas e peptídeos de genes ortólogos identificados nos dados de proteômica <i>shotgun</i> de corpo caloso, nervo óptico e bulbo olfatório do cérebro de humano e camundongo.	44
Tabela 4.3 – Quantidade de genes, proteoformas e peptídeos inferidos e identificados nos dados de proteômica <i>shotgun</i> de corpo caloso, nervo óptico e bulbo olfatório do cérebro de humano e camundongo.	45
Tabela 4.4 – Peptídeos proteotípicos de proteoformas identificados nos dados de MS de corpo caloso.	45
Tabela 4.5 – Peptídeos proteotípicos de proteoformas de genes ortólogos identificados nos dados de MS de nervo óptico.	46
Tabela 4.6 – Peptídeos proteotípicos de proteoforma identificadas nos dados de MS de bulbo olfatório.	47
Tabela 4.7 – Genes ortólogos entre humano e camundongo que expressam proteoformas inferidas nos dados de espectrometria de massas de corpo caloso, nervo óptico e bulbo olfatório.	48
Tabela 4.8 – Peptídeos identificados nos experimentos de MS de humano e camundongo e suas respectivas contagens de espectros de acordo com as três regiões do cérebro análises.	49-51
Tabela 5.1 – Informações sobre as proteoformas de genes ortólogos entre humano e camundongo inferidas nos dados de MS a partir de peptídeos proteotípicos idênticos.	96

LISTA DE SIGLAS E ABREVIATURAS

DNA – Ácido desoxirribonucleico (*deoxyribonucleic acid*)

RNA – Ácido ribonucleico (*ribonucleic acid*)

pré-mRNA – RNA mensageiro recém-sintetizado (*pre-messenger RNA*)

mRNA – RNA mensageiro (*messenger RNA*)

Perl – *Practical Extration and Report Language*

ESTs – Etiquetas de sequências transcritas (*Expressed Sequence Tag*)

EMBOSS – *European Molecular Biology Open Software Suite*

ncRNAs – RNAs não codificantes (*non-coding RNA*)

cDNA – DNA complementar (*complementary DNA*)

RefSeq – *Reference Sequence*

Uniprot – *The Universal Protein Resource*

PMS – *Peptide Spectrum Match*

FTP – *File Transfer Protocol*

PRIDE – *PRoteomics IDentifications database*

BLAST – *Basic Local Alignment Search Tool*

BLAT – *BLAST-like alignment tool*

SAV – *single amino acid variant*

NCBI – *National Center for Biotechnology Information*

PostgreSQL – programa utilizado como sistema gerenciador de banco de dados

UCSC – *University of California, Santa Cruz*

snRNA - pequenas ribonucleoproteínas nucleares (Small Nuclear RNA)

FDR – *False Discovery Rate*

PEP – *Posterior Error Probability*

MS – Espectrometria de massas (*Mass Spectrometry*)

RT-qPCR – *Real time quantitative PCR*

1 INTRODUÇÃO

O dogma central da biologia molecular, proposto por Francis Crick em 1970, corresponde, de maneira simplificada, a ideia de que uma porção da molécula de DNA, conhecida como gene, pode ser transcrita em RNA e traduzida em proteína (1,2). Com o aprimoramento tecnológico das plataformas desenvolvidas para se determinar a sequência dos blocos moleculares que compõem esses três tipos de moléculas, volumes de dados cada vez maiores passaram a ser produzidos (3–5). Com o intuito de se extrair informações com maior eficiência, tem-se realizado a integração dos dados destas diferentes plataformas tecnológicas, auxiliando o entendimento de eventos moleculares, como o *splicing* alternativo (6,7). Neste sentido, atualmente alguns projetos tem realizado esta análise integrativa com o intuito de comparar os resultados de amostras de ser humano e outros mamíferos (8–10), uma vez que alguns destes animais são utilizados como organismos modelo para o estudo dos aspectos moleculares de doenças como as neurodegenerativas (11–14).

1.1 O *splicing* alternativo

Em 1978 Walter Gilbert propôs a ideia de que os transcritos primários (pré-mRNAs) possuem regiões internas que são excisadas (íntrons) durante o processo de maturação do RNA mensageiro (mRNA), permanecendo apenas as regiões que serão traduzidas em proteínas (éxons) (15). Este evento molecular denominado *splicing* do pré-mRNA é realizado por um complexo de proteínas e RNAs não codificadores de proteínas (ncRNAs), conhecido como spliceossomo, e ocorre em conjunto a uma série de outros eventos de processamento [revisito por (16)].

O *splicing* alternativo foi descrito pela primeira vez em 1980 com a descoberta de um gene capaz de produzir dois mRNAs diferentes através da inclusão ou exclusão de éxons durante o processamento do pré-mRNA (17). Estes diferentes mRNAs ao serem traduzidos dão origem a proteínas distintas, comumente referidas como proteoformas (18,19). Este evento molecular é considerado um dos principais mecanismos responsáveis pela criação e perda de éxons em eucariotos (20) e foi de suma importância para a diversificação da vida na terra, sendo correlacionado à complexidade dos organismos (21).

Este processamento alternativo do pré-mRNA segue determinados padrões que podem ser apresentados de cinco maneiras distintas: (i) Uso alternativo de éxon, no qual um éxon ora se encontra presente no mRNA ora não se encontra; (ii) retenção de íntrons, em que um íntron é mantido no mRNA maduro; (iii) uso alternativo do sítio de *splicing* 5' e (iv) 3', em casos em que existem sítios de *splicing* alternativos nos íntrons; (v) éxons mutuamente exclusivos, no qual um éxon apenas é observado em um mRNA maduro na ausência de outro éxon (Figura 1.1) [revisado por (22)].

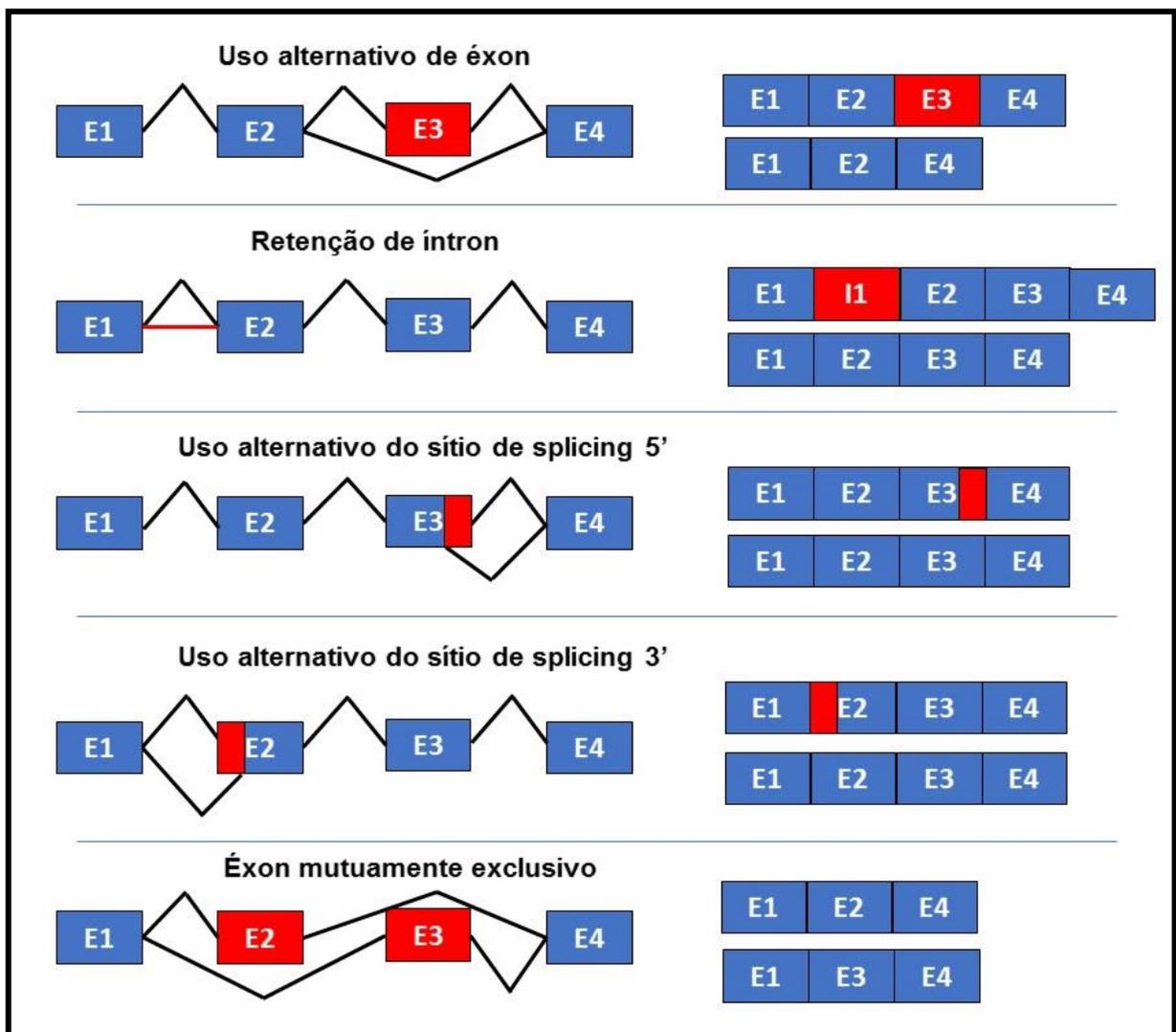


Figura 1.1 – Representação esquemática dos principais padrões de *splicing* alternativo. Os éxons estão representados pela letra “E” e íntrons pela letra I e pelas linhas. As caixas e linha em vermelho representam éxons e íntrons usados de forma alternativa. Adaptado de Wang e colaboradores (2015) (22).

O padrão de *splicing* alternativo mais prevalente em vertebrados é o uso alternativo de éxon. Entretanto, em metazoários menos complexos, o padrão mais comum é a retenção de íntron (23). O uso alternativo de éxon é o padrão de *splicing* mais passível de modificar domínios conservados entre espécies, como humano e camundongo, contribuindo como uma fonte adicional de complexidade e diferenças espécies-específicas entre os mamíferos (24).

As distintas conformações tridimensionais que as proteoformas produzidas por um gene assumem impactam suas funções. No verme *Caenorhabditis elegans*, o *splicing* alternativo do pré-mRNA do gene *egl-15* produz duas proteoformas de um fator de crescimento de fibroblasto, as quais se ligam a receptores celulares diferentes, desempenhando duas funções fisiológicas distintas (25). Este mesmo organismo produz duas proteoformas de uma proteína de membrana sintetizada pelo gene *fb1-1*. Estas proteoformas desempenham papéis distintos, porém, complementares na formação e organização de tecidos (26).

As diversas proteoformas produzidas em um tecido podem também exercer papéis específicos no comportamento de um organismo. Um exemplo particularmente interessante neste sentido é a produção de proteoformas sexo específicas do gene *fru* no sistema nervoso central de *Drosophila melanogaster*. Estas proteoformas são responsáveis pelo comportamento de acasalamento e orientação sexual das moscas machos e fêmeas. A indução experimental dos padrões de *splicing* masculinos em moscas fêmeas fez com que estas tentassem copular com outras fêmeas (27).

Estudos utilizando dados de sequenciamento de alta vazão demonstraram que o *splicing* alternativo ocorre em aproximadamente 95% dos genes de humano e é controlado de forma específica nos diversos tecidos (28,29). Este evento molecular, além de aumentar a diversidade funcional das proteínas, também participa na regulação da expressão gênica (30). Além disso, o *splicing* do pré-mRNA possui grande importância médica, uma vez que o seu mau funcionamento pode causar a super-expressão de algumas proteoformas ou gerar proteoformas aberrantes, estando associadas a algumas doenças do sistema nervoso central (31), tais como Alzheimer (32), Parkinson (33) e Esquizofrenia (34).

1.2 O *splicing* alternativo e o sistema nervoso central

No sistema nervoso central, milhares de isoformas de mRNAs são formadas através do mecanismo de *splicing* alternativo. Um dos exemplos mais ilustrativos da capacidade que o *splicing* alternativo possui em gerar múltiplas isoformas no sistema nervoso central de uma espécie é o gene *Dscam*, codificador de um receptor de orientação dos axônios da mosca *Drosophila melanogaster*. Este gene é capaz de expressar 38.016 mRNAs diferentes, produzidos através do *splicing* alternativo do pré-mRNA que, ao serem traduzidos, geram proteoformas que atuam em funções neuronais específicas (35).

Os processos celulares do sistema nervoso central de mamíferos também são coordenados pelo *splicing* alternativo através da produção de proteoformas tecido específicas (36). Por exemplo, as proteínas neuroliginas e neuroxinas atuam no processo de sinapse dos neurônios como moléculas de adesão celular (37,38). Ambas apresentam uma ampla diversidade de proteoformas (39,40). Foi observado que determinadas proteoformas de neuroligina 1 apenas se ligam a proteoformas específicas de alfa-neurexinas, estimulando a formação das sinapses, enquanto outras se ligam tanto às proteoformas de alfa-neurexinas quanto às proteoformas de beta-neurexinas, promovendo a expansão das sinapses (41).

Os eventos de *splicing* alternativo são dirigidos pelo recrutamento de proteínas reguladoras às junções específicas entre éxons e íntrons. Uma das famílias de proteínas que realizam este tipo de regulação no sistema nervoso de vertebrados são as proteínas ricas em Ser/Arg (proteínas SR), em especial a SR100 (42). Esta proteína reguladora é de extrema importância para que o *splicing* dos pré-mRNAs de diversos genes expressos no sistema nervoso central ocorra de maneira adequada. Foi demonstrado em estudo *in vivo* que a deleção do gene codificador da SR100 prejudica o desenvolvimento do sistema nervoso central e periférico de mamíferos através da interrupção do crescimento dos neurônios e da orientação dos axônios em regiões importantes do cérebro, como o corpo caloso, responsável por realizar a conexão de um hemisfério cerebral com o outro (43).

A PTBP1 (*polypyrimidine tract binding protein 1*) e PTBP2 (*polypyrimidine tract binding protein 2*) são duas outras proteínas reguladoras do *splicing* específicas do sistema nervoso central (44). A PTBP1 funciona como repressora do programa de *splicing* específico dos neurônios e é amplamente expressa em células de outros tecidos (45). Durante o processo de neurogênese, a redução na expressão de PTBP1 permite a expressão de PTBP2, causando uma mudança no padrão de *splicing* dos genes expressos nos neurônios em formação (46). A influência da mudança de expressão destes reguladores do *splicing* alternativo sobre a formação dos neurônios foi demonstrada em um experimento *in vitro* de transdiferenciação de fibroblastos em neurônios utilizando um sistema de cultura celular de fibroblastos com depleção de PTBP1 e super-expressão de PTBP2 (47).

Em um humano adulto o cérebro é um dos órgãos em que ocorre o maior nível de expressão de proteoformas (48) as quais participam de diversos processos neuronais, tais como a neurogênese (49), mudança de atividade de receptores neuronais, controle da função sináptica (50), formação da grande diversidade dos canais de íons de potássio (51) e a despolarização dos neurônios (52). Os padrões de *splicing* alternativo tendem a ser mais conservados no sistema nervoso de vertebrados (53). Entretanto, o *splicing* alternativo também é capaz de gerar proteoformas espécie-específicas, sendo um dos eventos moleculares responsáveis pelas diferenças entre o cérebro humano e de outros mamíferos como o camundongo (54) e primatas não-humanos (55).

1.3 O sistema nervoso de camundongo e de humano

O cérebro de camundongo é considerado similar ao de humano em diversos aspectos. Por exemplo: a sequência de eventos chave na maturação do cérebro de humanos e roedores é equivalente (56); os mecanismos moleculares e celulares responsáveis pela formação do corpo caloso são similares em humano e camundongo (57); o fato de que certas mutações observadas em pacientes com doenças neurológicas podem causar fenótipos similares em ambas as espécies (58,59); a elevada conservação das sequências nucleotídicas, tamanho e quantidade de éxons (60) e nível de expressão gênica (61) que estas espécies apresentam; a semelhança da distribuição dos neurônios através da área visual e não visual do córtex cerebral de camundongo e humano (62).

Entretanto, existem algumas diferenças neurais estruturais e funcionais entre os seres humanos e o camundongo. Por exemplo, os lobos cerebrais humanos são definidos por sulcos, giros e marcações externas, enquanto o cérebro de camundongo não apresenta essas características; humanos possuem um lobo olfatório pequeno, com bastante substância branca, enquanto em camundongos essa região é relativamente espessa; o nervo óptico de camundongo possui dez vezes menos axônios que o de humano; o hipocampo do cérebro de camundongos possui um aspecto dorsal em relação ao cérebro, porém, em humanos, essa região apresenta um aspecto ventral; as meninges dos camundongos são finas, entretanto, as de humanos são grossas e bem desenvolvidas (63); o cérebro humano pesa 1,5 quilogramas e possui 86 bilhões de neurônios; contudo, o cérebro de camundongo apresenta 0,5 grama e contém 70 milhões de neurônios (64); humanos possuem astrócitos maiores, mais rápidos e muito mais complexos que os dos camundongos (65); camundongos e humanos possuem sequências envolvidas no estado/organização da cromatina e regulação transcricional bem distintas (66).

O camundongo tem sido utilizado como modelo biológico para o estudo de doenças neurodegenerativas (11–14) e para a caracterização de diferentes regiões do sistema nervoso central, como o corpo caloso (67), nervo óptico (68) e bulbo olfatório (69). Este roedor tem sido utilizado como modelo biológico por ser um mamífero pequeno, o que torna mais fácil mantê-lo em instalações de pesquisa, e por ser fácil de criar e de treinar para manipulação em análises *in vivo* (70). Os estudos realizados em modelos murinos se baseiam na extrapolação dos dados obtidos nestes modelos para o ser humano. Entretanto, o melhor meio de garantir que esta extrapolação seja realizada de forma racional é através do entendimento das diferenças e similaridades entre as diversas estruturas do cérebro do ser humano e do camundongo (63).

Os dados presentes na literatura indicam a ocorrência de processos regulatórios após a formação do mRNA, fazendo com que nem todos os mRNAs sejam traduzidos em proteínas (71). Portanto, alguns projetos têm sido desenvolvidos voltados para a caracterização do conteúdo proteico do cérebro de camundongo (72) e de suas diferentes regiões (73,74) em diferentes estágios de desenvolvimento (75). Desta maneira, a elucidação das diferenças e similaridades dos componentes proteicos do cérebro de humano e de camundongo pode ser

considerada uma maneira de se contribuir para a melhor compreensão da real aplicabilidade dos modelos biológicos murinos.

1.4 A proteômica baseada em espectrometria de massas

O termo proteoma foi introduzido na literatura em 1995 (76) e se refere ao complexo proteico total expresso por uma célula ou tecido em uma determinada condição ou momento específico (77). Neste sentido, o estudo do proteoma passou a ser conhecido como proteômica (76). Atualmente, a estratégia mais utilizada em estudos de proteômica é a espectrometria de massas (MS, do inglês *mass spectrometry*), por meio da abordagem *shotgun* (78). Esta metodologia tem sido utilizada no âmbito das neurociências para a identificação de proteínas diferencialmente expressas em pacientes com doenças neurodegenerativas (79), desordens psicológicas (80) e lesões cerebrais (81). Além disso, também tem sido utilizada na identificação de biomarcadores em diversos tipos de câncer (82–84).

Esta estratégia consiste na digestão enzimática das proteínas em peptídeos, normalmente por tripsina, os quais são separados em uma coluna de cromatografia líquida de fase reversa hifenada com o espectrômetro de massas (85) (Figura 1.2A). Os peptídeos fracionados são ionizados por *electrospray* e, em seguida, analisados sob ação de campos elétricos/magnéticos em um sistema a vácuo. Isto resulta na separação destes íons de peptídeos de acordo com suas relações massa/carga, gerando os espectros de MS1 (86) (Figura 1.2B). A cada espectro de MS1, íons peptídicos precursores são selecionados sequencialmente e submetidos individualmente à fragmentação, comumente realizada por meio da colisão com um gás inerte. A fragmentação ocorre preferencialmente nas ligações peptídicas, gerando íons-fragmento das séries y (contendo o C-terminal do precursor) e/ou b (contendo o N-terminal do precursor), que serão separados, também segundo sua relação massa/carga, gerando o espectro de MS2 ou MS/MS (87) (Figura 1.2C).

A identificação da sequência de aminoácidos normalmente é feita através da comparação dos espectros de MS2 experimentais com espectros de fragmentação teóricos, gerados a partir de um repositório de sequências proteicas, uma técnica denominada *Peptide Spectrum Matching* (PSM) (88) (Figura 1.2D). Os principais repositórios de sequências proteicas utilizados em PSM são o Uniprot (89), Refseq

(90) e Gencode (91). O Uniprot tem sido um dos repositórios mais utilizados, pois possui sequências curadas manualmente (UniprotKB/Swiss-Prot), com validação experimental e anotação funcional. Entretanto, o Refseq é baseado em traduções hipotéticas do transcriptoma curadas manualmente, enquanto o Gencode contém sequências provenientes da tradução hipotética de genomas, sem validação experimental no proteoma.

Os peptídeos identificados passam por uma etapa de seleção segundo critérios estatísticos. Nesta etapa, existem duas métricas que podem ser aplicadas: a taxa de falsos positivos (do inglês, *False Discovery Rate*; FDR), utilizada como uma medida global de falsos positivos identificados (92,93) e a probabilidade de erro posterior (do inglês, *Posterior Error Probability*; PEP), definida como a probabilidade de um dado PSM ser incorreto e é vista como uma versão local da FDR (94,95) (Figura 1.2E). Um dos principais desafios da proteômica *shotgun* é a identificação de peptídeos proteotípicos, ou seja, peptídeos exclusivos que possam garantir, de forma inequívoca, a identificação de uma única proteína do repositório de sequências proteicas utilizado (96).

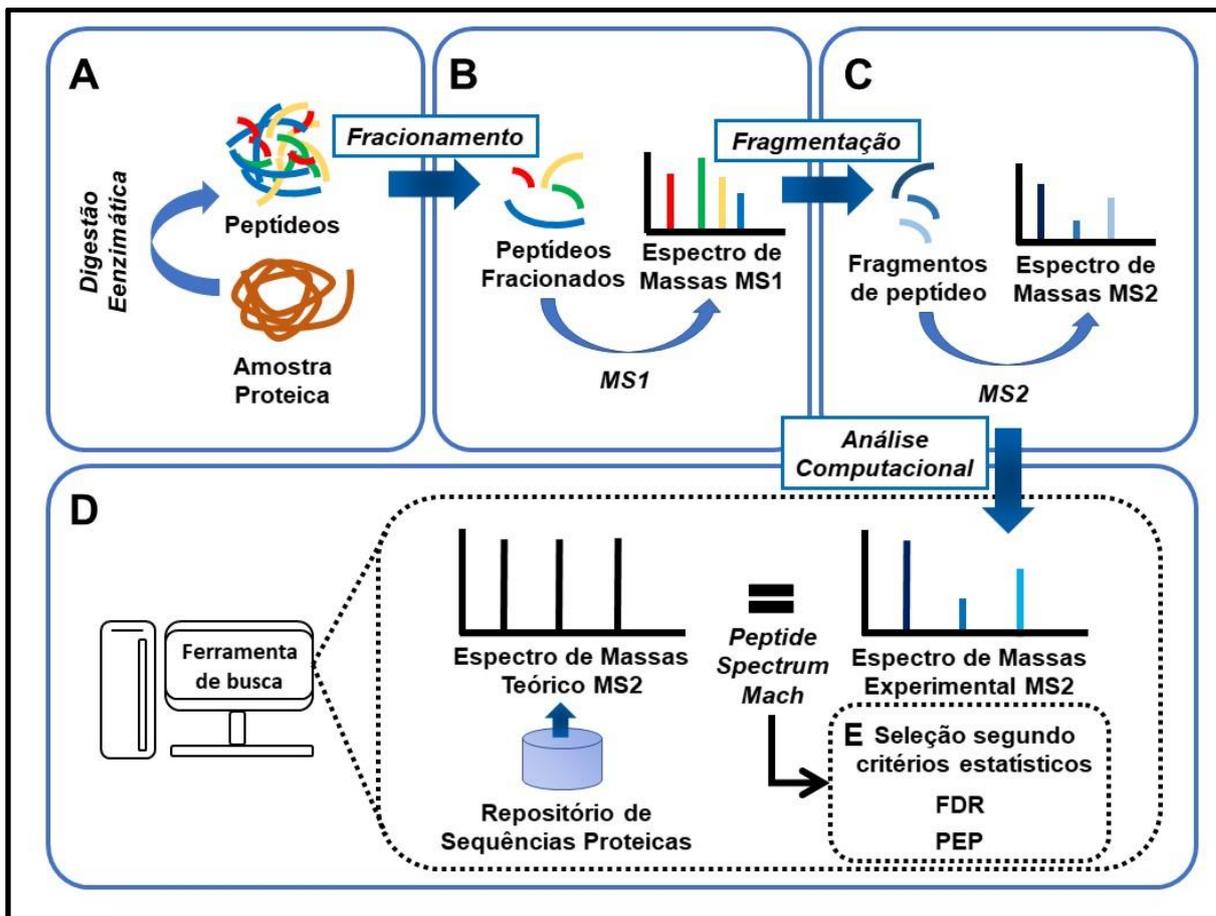


Figura 1.2 – Representação esquemática de um fluxo de análises de proteômica *shotgun*. (A) As proteínas de uma amostra proteica são digeridas em peptídeos por uma enzima específica. (B) Os peptídeos são fracionados por uma metodologia de cromatografia e em seguida são obtidos os espectros de massa MS1 correspondentes à razão massa/carga destes peptídeos. (C) Os peptídeos precursoros fracionados são então fragmentados e têm suas razões massa/carga medidas, dando origem aos espectros de massa MS2. (D) Os espectros de massas experimentais MS2 são comparados aos espectros de massas MS2 teóricos, gerados computacionalmente, a partir de um repositório de sequências proteicas. (E) Caso coincidam, é realizada a seleção segundo critérios estatísticos da identificação do peptídeo.

1.5 A proteogenômica

Ao se realizar um experimento de proteômica *shotgun*, é frequente a ausência de associação de espectros de massas de alta qualidade com uma sequência peptídica conhecida (97). Portanto, alguns desses espectros podem ser correspondentes a peptídeos de proteoformas que não estão contemplados em repositórios de sequências proteicas convencionais (7). Uma das abordagens mais frequentes para a identificação desses peptídeos é a proteogenômica (98). Este termo foi introduzido na literatura em 2004 (99) e pode ser definido como o uso de dados de sequenciamento do genoma ou transcriptoma com a finalidade de se criar

repositórios de sequências proteicas personalizados para aperfeiçoar a detecção de variantes proteicas (100). Portanto, a proteogenômica é parte da bioinformática integrativa, visando à integração de dados de diferentes fontes com diferentes formatos (101).

A construção de repositórios de sequências proteicas personalizados normalmente envolve duas etapas principais. A primeira etapa consiste na obtenção de sequências nucleotídicas. Para algumas espécies, estas sequências estão disponíveis publicamente em bancos de dados como Refseq (102) e dbEST (103). Entretanto, estes bancos de dados não possuem sequências específicas para determinadas condições celulares ou para alguns organismos pouco estudados. Desta forma, as sequências nucleotídicas podem ser obtidas por meio do sequenciamento do genoma (104), do exoma (105), do transcriptoma (106) ou do transloma (107). Na segunda etapa, é realizada a tradução hipotética dessas sequências nucleotídicas em aminoácidos, gerando o repositório de sequências proteicas personalizado. A tradução hipotética pode ser realizada nos três quadros de leitura (quando se sabe a orientação do transcrito) ou nos seis quadros de leitura, utilizando um programa como o transeq do pacote EMBOSS (108) ou através de funções de módulos de linguagens de programação como Bioperl (109), Biopython (110) e Biostrings (Bioconductor). Estes repositórios podem ser específicos para identificação de variantes de aminoácido único (do inglês, *Single Aminoacid Variant*; SAV) (111), variantes de *splicing* (112,113) ou RNAs mensageiros quiméricos (114) (Figura 1.3).

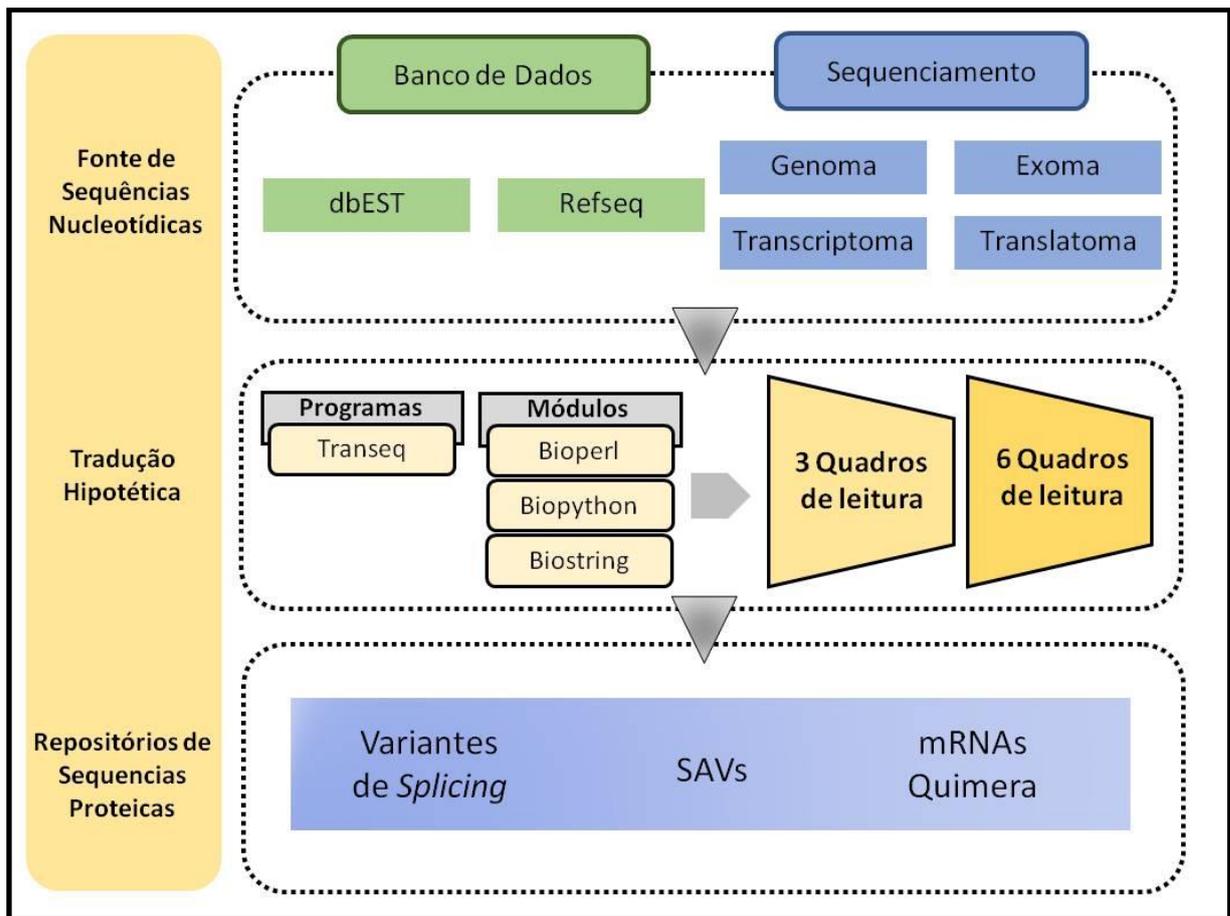


Figura 1.3 – Esquema representativo de um fluxo de análises geral para construção de um repositório de sequências proteicas personalizado.

Neste sentido, muitas estratégias têm sido desenvolvidas para a construção de repositórios de sequências proteicas personalizados e análise integrativa dos dados de genoma, transcriptoma e proteoma. Por exemplo, Evans e colaboradores (2012) (115) utilizaram dados de sequenciamento de RNA mensageiro para a construção de um repositório de sequências proteicas personalizado, com o intuito de identificar genes e proteínas de espécies que não possuem um repositório de sequências proteicas disponível publicamente. Para realização da prova de conceito, os autores utilizaram um modelo de cultura celular humana infectada com adenovírus. Esta abordagem possibilitou a identificação de mais de 99% do total de 3.700 proteínas identificadas pela análise com repositórios de sequências proteicas convencionais.

Castellana e colaboradores (2010) (116) desenvolveram uma abordagem híbrida entre o PMS e a identificação de peptídeos *de novo*, isto é, a identificação de peptídeos sem a utilização de um repositório de sequências proteicas. Nesta metodologia, inicialmente ocorre a busca dos espectros de massa MS2 contra um repositório de sequências proteicas, os peptídeos identificados são chamados de âncora. Em seguida, os espectros de massa MS2 não identificados no repositório de sequências proteicas são utilizados para o sequenciamento *de novo*. Os peptídeos âncora têm suas sequências estendidas a partir da sobreposição de peptídeos que foram identificados pelo sequenciamento *de novo*, possibilitando a reconstrução da sequência completa da proteína. Neste trabalho, os autores utilizaram dados de espectrometria de massas para reconstruir a sequência completa de um anticorpo e obtiveram 98% de acurácia. Uma das desvantagens desta metodologia é que ela apenas pode ser utilizada para proteômica dirigida, isto é, para o sequenciamento de proteínas previamente estabelecidas, neste caso, um anticorpo.

Helmy e colaboradores (2012) (117) desenvolveram uma estratégia de proteogenômica chamada subtração sequencial de espectros de massa (do inglês, *Mass Spectrum Sequential Subtraction*; MSSS). Os autores realizaram a busca dos espectros de massas de proteínas de culturas de células de arroz contra diversos repositórios de sequências proteicas de maneira sequencial, de forma que os espectros que identificassem peptídeos não fossem utilizados na busca contra o próximo repositório. Esta abordagem computacional reduziu o tempo de busca em 75% e não teve efeito sobre a FDR e sobre a identificação de novos peptídeos. Além disso, esta metodologia possibilitou a identificação de 74 peptídeos novos, que ao serem mapeados ao genoma revelaram 47 características genômicas novas. Entretanto, Ivanov e colaboradores (2017) (118) demonstraram que a busca dos espectros de massa contra um repositório de sequências proteicas personalizado único proporciona uma maior identificação de variantes de peptídeos do que a busca contra diversos repositórios de sequências proteicas de maneira sequencial.

Nosso grupo de pesquisa também tem contribuído com o estado da arte da proteogenômica através do desenvolvimento de um repositório de sequências proteicas personalizado com o intuito de identificar proteoformas em dados de proteômica *shotgun* de humano. Foram utilizadas sequências de referência do transcriptoma (Refseq), ESTs e *reads* de sequenciamento de alta vazão de RNA mensageiro para a identificação dos transcritos variantes de *splicing*, através de uma metodologia desenvolvida pelo grupo de pesquisa chamada matrizes ternárias. As sequências identificadas como variantes de *splicing* foram traduzidas *in silico* através do programa *transeq*, gerando assim o repositório de sequências proteicas personalizado chamado SpliceProt. Foram identificados 54 peptídeos de proteoformas que não estavam presentes no UniprotKB/TrEMBL através da análise de dados de espectrometria de massas de proteínas oriundos de bancos de dados públicos (113).

A proteogenômica tem sido aplicada no melhoramento da anotação do genoma humano (119), no estudo de proteoformas alvo para o desenvolvimento de fármacos [revisão por (120)], na identificação de variantes proteicas implicadas em determinados tipos de câncer (82) e na detecção de proteoformas pertencentes a tipos celulares específicos (121). Entretanto, esta metodologia também tem sido aplicada em dados de diversos organismos (122–125), o que possibilita o seu emprego no âmbito da biologia comparada. Devido à vasta utilização do camundongo como modelo biológico para o estudo de doenças neurodegenerativas e de processos biológicos cerebrais, a análise comparativa da diversidade destas proteoformas em distintas regiões do cérebro de humano e camundongo pode gerar resultados que sirvam como ponto de partida para futuros estudos que visem avaliar as limitações do modelo murino.

2 OBJETIVOS

Objetivo Geral

O presente projeto tem como objetivo analisar o perfil de expressão de variantes por *splicing* alternativo em dados de espectrometria de massas de proteínas de amostras de tecidos de cérebros saudáveis de humano e camundongo em busca de proteoformas de genes ortólogos entre as duas espécies.

Objetivos Específicos

- 1) Construção do banco de dados contendo as variantes de *splicing* de camundongo, através da metodologia das matrizes ternárias.
- 2) Construção do repositório SpliceProt e repositório de sequências proteicas personalizado de camundongo.
- 3) Análise dos dados de espectrometria de massas de proteínas do cérebro saudável de humano e camundongo, utilizando o repositório de humano previamente desenvolvido pelo grupo de pesquisa ao qual o aluno encontra-se inserido e o repositório personalizado de camundongo construído em 2).
- 4) Comparação do perfil de *splicing* alternativo de humano e camundongo.

3 MATERIAL E MÉTODOS

Todas as análises realizadas neste projeto de pesquisa utilizaram a infraestrutura do Laboratório de Genômica Funcional e Bioinformática do Instituto Oswaldo Cruz da Fundação Oswaldo Cruz (IOC/Fiocruz) e da Plataforma de Bioinformática da Fiocruz RPT04A/RJ. Todos os programas desenvolvidos neste projeto foram escritos na linguagem de programação Perl e as informações processadas foram armazenadas e recuperadas em uma base de dados utilizando o sistema gerenciador de banco de dados (SGBD) PostgreSQL (versão 8.4.7). Os diagramas de Venn apresentados nos resultados foram confeccionados utilizando o programa *powerpoint* (versão 2016).

3.1 Conjunto de dados

Os genomas de humano (versão hg-19) e de camundongo (versão mm9) foram obtidos através do sítio de FTP da Universidade da Califórnia, campus Santa Cruz (UCSC) (126). Os dados referentes aos transcritos foram obtidos através do projeto Unigene (127) de humano (versão 235) e de camundongo (versão 194), nos quais foram adquiridas sequências de referência (prefixos NM) e ESTs (do inglês, *Expressed Sequence Tags*). As sequências proteicas curadas manualmente de humano e camundongo foram obtidas no banco de dados Uniprot/Swissprot (89) em novembro de 2017.

Nós utilizamos dados de espectrometria de massas de proteínas de três regiões do cérebro sadio de humano e camundongo: corpo caloso, bulbo olfatório e nervo óptico. Estas regiões do cérebro foram escolhidas devido a relevância que apresentam no contexto cognitivo, além de já possuírem dados disponíveis publicamente. Os dados de espectrometria de massas no formato *RAW* foram adquiridos a partir do repositório PRIDE do consórcio *ProteomExchange* (128), os quais foram oriundos de quatro projetos diferentes (3 de humano e 1 de camundongo) representados pelos seus respectivos identificadores (Tabela 3.1).

Tabela 3.1 – Informações dos dados de proteômica *shotgun*.

Organismo	Humano			Camundongo		
	Região Cerebral	Corpo Caloso	Nervo Óptico	Bulbo Olfatório	Corpo Caloso	Nervo Óptico
Identificador	PXD000547	PXD001581	PXD005629	PXD001250		
Referência	(129)	(130)	(131)	(74)		
Espectrômetro de Massas	LTQ-Orbitrap XL	Orbitrap Elite	Orbitrap Fusion Tribrid	Q Exactive		
Quantidade de Amostras	2	5	3	3	3	4
Tamanho dos datasets (Giga Bytes)	9.8	15	47	13.3	13.6	23.8

3.2 Matrizes Ternárias

As matrizes ternárias consistem na construção de matrizes de tamanho $N \times M$, onde N é o número total de quaisquer transcritos ou proteínas mapeadas em uma região específica de um dado genoma, e M é o número total de éxons e íntrons encontrados na mesma região (132). Nesta metodologia, os transcritos são mapeados às suas respectivas unidades transcricionais para verificação do padrão de *splicing* (Figura 3.1A). Cada célula da matriz recebe o caractere “1” para a presença de um éxon (ou parte dele) mapeado, “0” na ausência de mapeamento e “|” para delimitar os limites de éxons e também os limites do primeiro e do último éxon do gene (Figura 3.1B). Uma variante de *splicing* é selecionada quando são encontrados um ou mais caracteres “0” flanqueados por “1”. Os éxons (ou parte deles) não mapeados e que não são flanqueados por “1” são preenchidos por caracteres “1” de forma a eliminar possíveis redundâncias, uma vez que as sequências de ESTs são incompletas e a ausência de porções nas suas extremidades podem evidenciar uma ausência de sequenciamento daquela região e não um novo evento de *splicing* (Figura 3.1C).

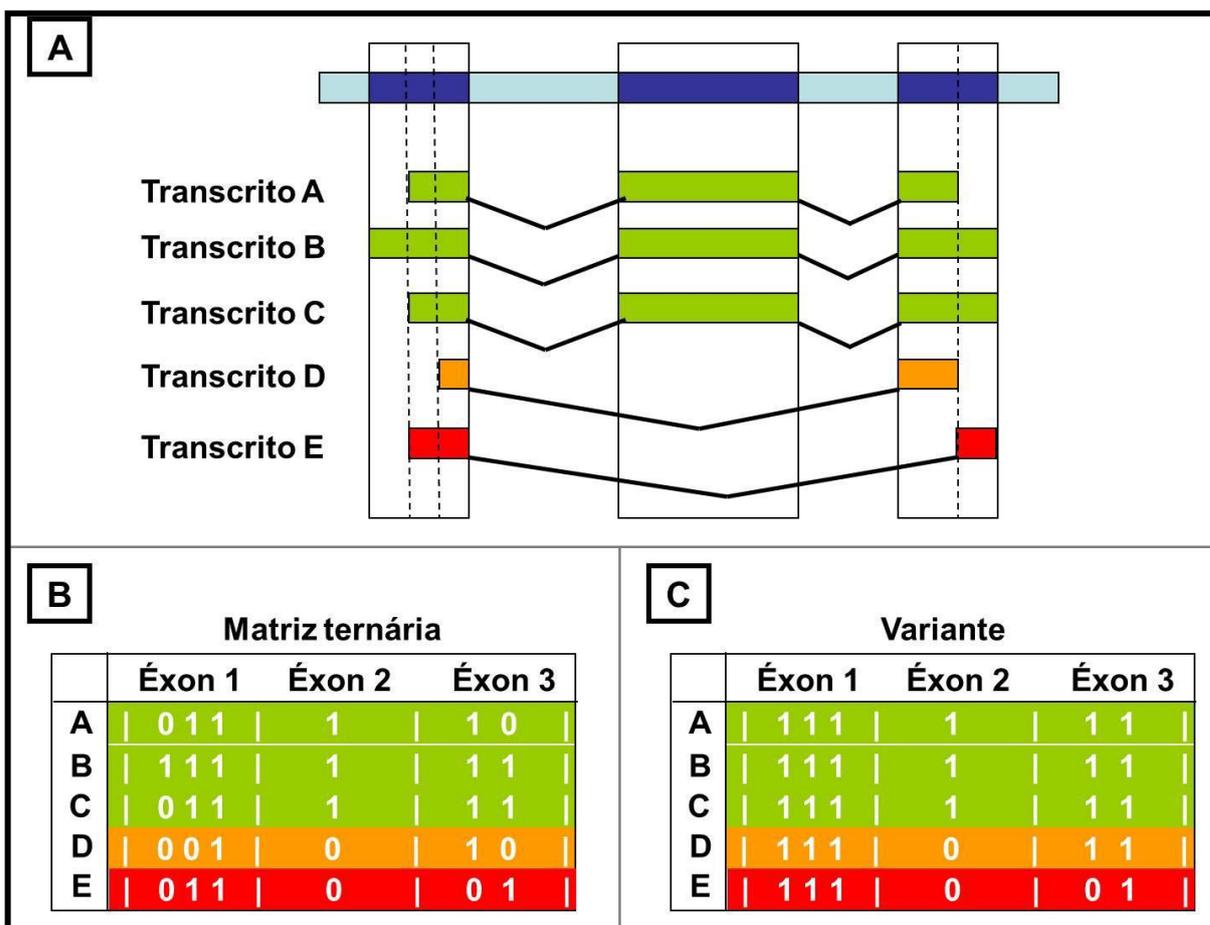


Figura 3.1 - Construção de uma matriz ternária. (A) Todos os transcritos pertencentes a uma determinada unidade transcricional são mapeados para identificação das regiões genômicas expressas. (B) Após a delimitação das coordenadas de cada éxon, a presença ou ausência de cada região expressa é classificada como “1” ou “0”, respectivamente. (C) Cada variante de *splicing* é identificada quando um ou mais “0” são flanqueados por “1”. No exemplo, podemos observar três variantes: os transcritos A, B e C seriam uma mesma variante (não apresentando qualquer evento de *splicing*), o transcrito D (apresentando uso alternativo do éxon 2) e o transcrito E (apresentando uso alternativo do éxon 2 e uso alternativo do sítio de *splice* 3’ do éxon 2). Figura de Tavares, 2016 (133).

3.3 Construção do repositório de sequências proteicas Spliceprot

Para a realização da construção do repositório Spliceprot de camundongo, nós utilizamos um conjunto de programas escrito pelo Dr. Raphael Tavares durante seu período de mestrado. A primeira etapa consistiu em alinhar os dados de Refseq e EST contra os genomas de ambos os organismos para a determinação das coordenadas gênicas, utilizando o programa BLAT (134) com os seguintes parâmetros: “-minIdentity=97 -out=sim4”. Transcritos contendo apenas um éxon foram desconsiderados. Em seguida, cada transcrito foi analisado de acordo com seu respectivo mapeamento às unidades transcricionais do genoma de referência,

agrupado em variantes de *splicing* através da metodologia de matrizes ternárias, descrita no tópico anterior, e armazenado em um banco de dados SQL.

Subsequentemente, através deste conjunto de programas nós recuperamos do banco de dados as sequências nucleotídicas dos transcritos variantes de *splicing* identificados pelas matrizes ternárias e direcionamos para o programa *transeq* do pacote EMBOSS (versão 6.3.1) (135). Este programa foi configurado para realizar a tradução *in silico* nas 3 fases de leitura do transcrito (parâmetro: “-frame=3”), de acordo com a sua respectiva orientação gênica. Após esta etapa, a seleção da fase de leitura correta ocorreu utilizando o programa *clustalw* (versão 2.1) ao compararmos as sequências traduzidas das 3 fases de leitura para aqueles genes que possuem uma proteína conhecida. Foi escolhida assim a sequência traduzida da fase de leitura com maior identidade e cobertura em relação à sequência da proteína conhecida. Para aqueles genes cuja proteína não é conhecida, foi selecionada a sequência mais longa dentre as traduzidas a partir das 3 fases de leitura (Figura 3.2).

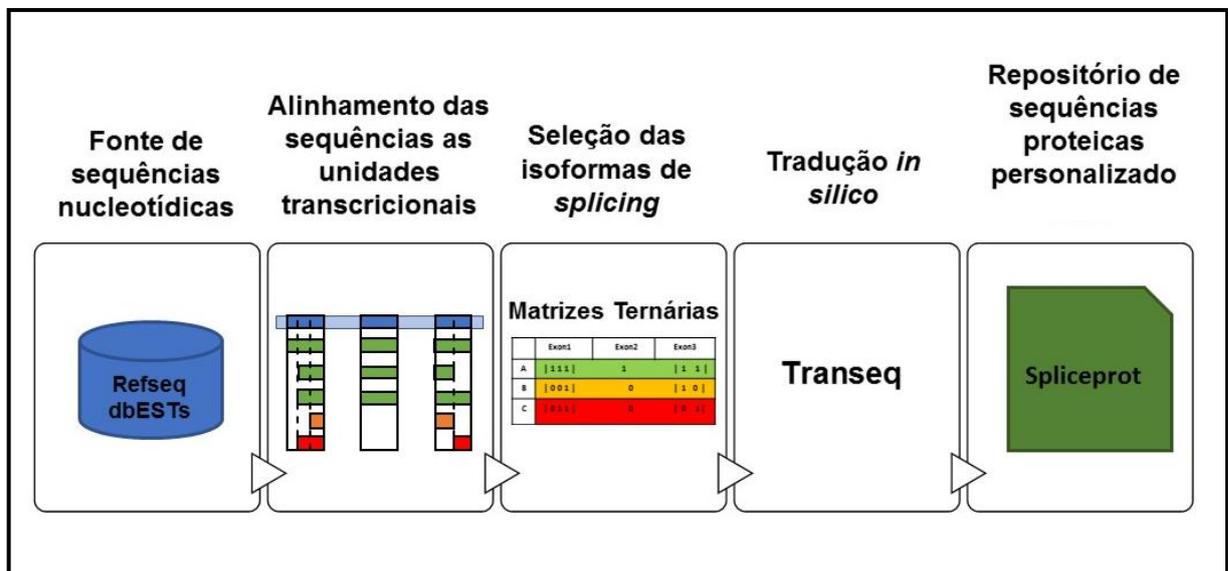


Figura 3.2 – Esquema representativo do fluxo de análises para construção do repositório de sequências proteicas personalizado (Spliceprot).

3.4 Construção dos repositórios de seqüências proteicas personalizados

Para a construção do repositório proteico personalizado foi desenvolvido um conjunto de programas de acordo com o desenho experimental de Tavares e colaboradores (2016) (121). Para tal, em primeiro lugar, seqüências canônicas do repositório Uniprot/Swiss-Prot tiveram suas redundâncias removidas e em seguida, foram copiadas para um arquivo temporário. As seqüências das proteoformas foram acrescentadas a um arquivo a parte junto com as seqüências do SpliceProt e, após a remoção das seqüências redundantes, foi efetuada a digestão computacional pelo programa Digest (pacote EMBOSS versão 6.3.1). Este programa foi previamente modificado para que a digestão computacional hipotética por tripsina fosse realizada somente na porção c-terminal dos aminoácidos lisina (K) e arginina (R), exceto se sucedidos por uma prolina (P). Os peptídeos produzidos por esta digestão computacional foram selecionados de forma que não pudessem estar presentes em qualquer seqüência canônica. Por fim, esses peptídeos selecionados tiveram também a suas redundâncias removidas e, em seguida, foram acrescentados ao repositório proteico-base que continha as seqüências canônicas do Uniprot, finalizando o repositório personalizado (Figura 3.3).

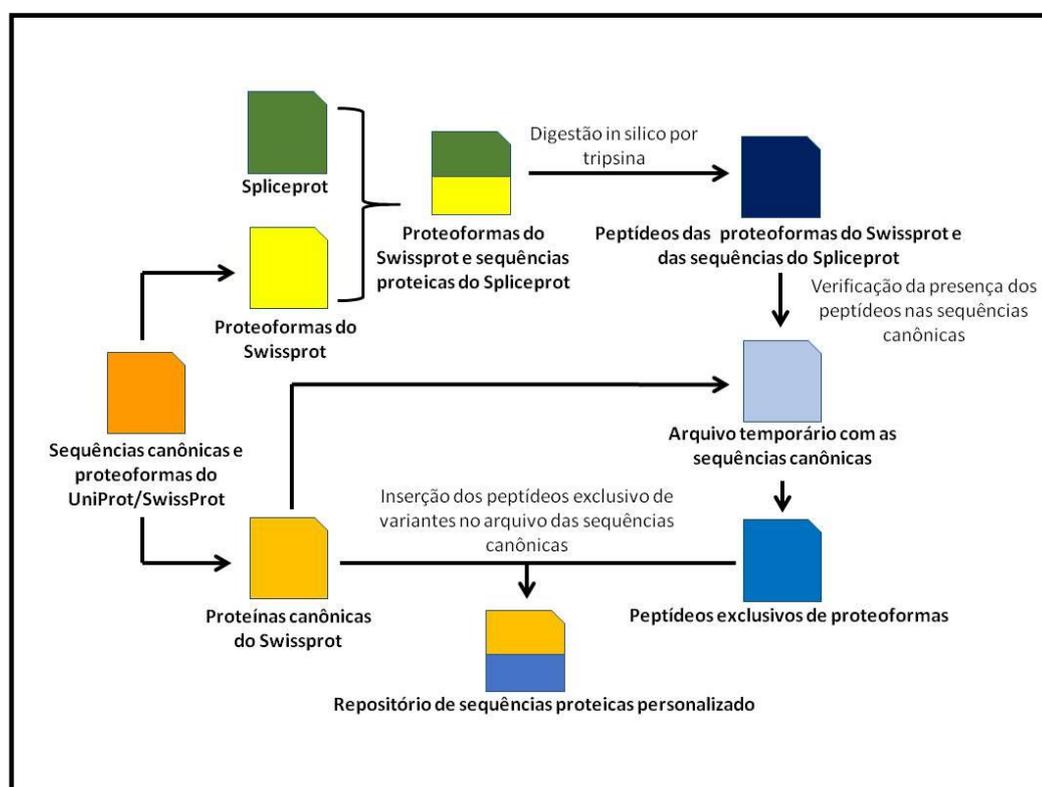


Figura 3.3 – Esquema para a construção do repositório proteico personalizado. Figura adaptada de Tavares e colaboradores (2017) (121).

3.5 Análise dos dados de espectrometria de massas de proteínas

Os dados de espectrometria de massas das diferentes regiões do cérebro de humano e camundongo foram analisados por meio do programa *PatternLab for Proteomics* (versão 4.0.0.74) (136). Nós utilizamos este programa porque foi demonstrado que ele apresenta a menor taxa de falsos positivos para identificação de peptídeos dentre os programas comumente utilizados para análise de dados de proteômica *shotgun*, por meio da abordagem PSM (137).

Na primeira etapa, foi gerado um banco *decoy* do tipo Target-Reverse contendo as sequências do repositório de sequências proteicas personalizado no sentido correto (Target) e invertidas (Reverse). Em seguida, foi realizada a busca pelos espectros de massas neste repositório. Esta etapa foi conduzida pela ferramenta de busca *Comet* (138), a qual se encontra implementada no programa *PatternLab for Proteomics*. Nós utilizamos os parâmetros recomendados por Eng e colaboradores (2015) (139), de acordo com as informações de resolução dos espectrômetros de massas e protocolos utilizados para gerar os dados de humano e camundongo em cada projeto (Tabela 3.2).

Tabela 3.2 – Parâmetros utilizados para busca dos espectros de massas contra o repositório de sequências proteicas personalizado.

Parâmetros	Humano	Camundongo
<i>Enzyme</i>	<i>TrypsinNotFollowedByP</i>	<i>TrypsinNotFollowedByP</i>
<i>Enzyme Specificity</i>	<i>semi-specific</i>	<i>semi-specific</i>
<i>Fix Modifications</i>	<i>Carbamidomethylation of Cysteine</i>	<i>Carbamidomethylation of Cysteine</i>
<i>Variable Modifications</i>	<i>Oxidation of Methionine</i>	<i>Oxidation of Methionine</i>
<i>Precursor Mass Tolerance</i>	40	40
<i>Fragment Bin Tolerance</i>	1.0005	0.2
<i>Fragment Bin Offset</i>	0.04	0
<i>Missed Cleavages</i>	2	2
<i>Max variable mods per peptide</i>	2	2
<i>Theoretical frag ions</i>	<i>M peak only</i>	<i>default peak shape</i>
Acurácia dos fragmentos (MS1-MS2)	<i>High-Low</i>	<i>High-High</i>

Por fim, foi realizado o controle de confiança estatística dos PSMs, baseado na taxa de falsos positivos FDR através do programa *SEpro* (140), implementado no programa *PatternLab for Proteomics*. Para todos os conjuntos de dados, apenas foi admitida a identificação do conjunto de peptídeos e proteínas com um valor de FDR menor ou igual a 1%.

3.6 Busca por proteínas de genes ortólogos entre humano e camundongo

Nós desenvolvemos um conjunto de programas para o processamento dos peptídeos identificados e posterior busca das proteínas e peptídeos de genes ortólogos entre humano e camundongo. Os arquivos *gene2refseq.txt*, *gene2unigene.txt*, *gene_refseq_uniprotkb_collab.txt*, utilizados para a conversão dos códigos de proteínas (*Uniprot ID*, *Refseq ID* e *EST ID*), e o arquivo utilizado para identificação dos genes ortólogos (*gene_group.txt*) foram obtidos através do sítio FTP do *National Center for Biotechnology Information* (NCBI) (<https://ftp.ncbi.nlm.nih.gov/gene/DATA/>).

Na primeira etapa, os peptídeos foram separados de acordo com sua origem, isto é, por serem de proteínas canônicas ou exclusivos de proteoformas (Figura 3.4A). As sequências do Uniprot/Swiss-Prot mais prevalentes, com maior similaridade entre sequências ortólogas de outras espécies e com maior comprimento, são definidas como canônicas e possuem um código Uniprot específico (141). Na segunda etapa, tanto os peptídeos de proteínas canônicas, quanto os peptídeos de proteoformas passaram por um filtro, no qual foram removidos os oriundos de sequências decoy e contaminantes e selecionados os peptídeos proteotípicos com contagem expectral superior a dois. Com exceção das proteoformas, neste caso também foram aceitos peptídeos compartilhados entre proteoformas (Figura 3.4B). Na terceira etapa, os códigos das sequências proteicas (*Uniprot ID*, *Refseq ID* e *EST ID*) dos peptídeos foram convertidos para código de gene (*Gene ID*), utilizando os arquivos *gene_refseq_uniprotkb_collab.txt*, *gene2refseq.txt* e *gene2unigene.txt* (Figura 3.4C). Na quarta e última etapa, os *Gene IDs* de humano e camundongo, resultantes da etapa anterior, foram comparados através do quarto arquivo (*gene_group.txt*) para a identificação dos genes ortólogos identificados. Este arquivo contém todos os genes ortólogos entre humano e camundongo conhecidos na literatura (Figura 3.4D).

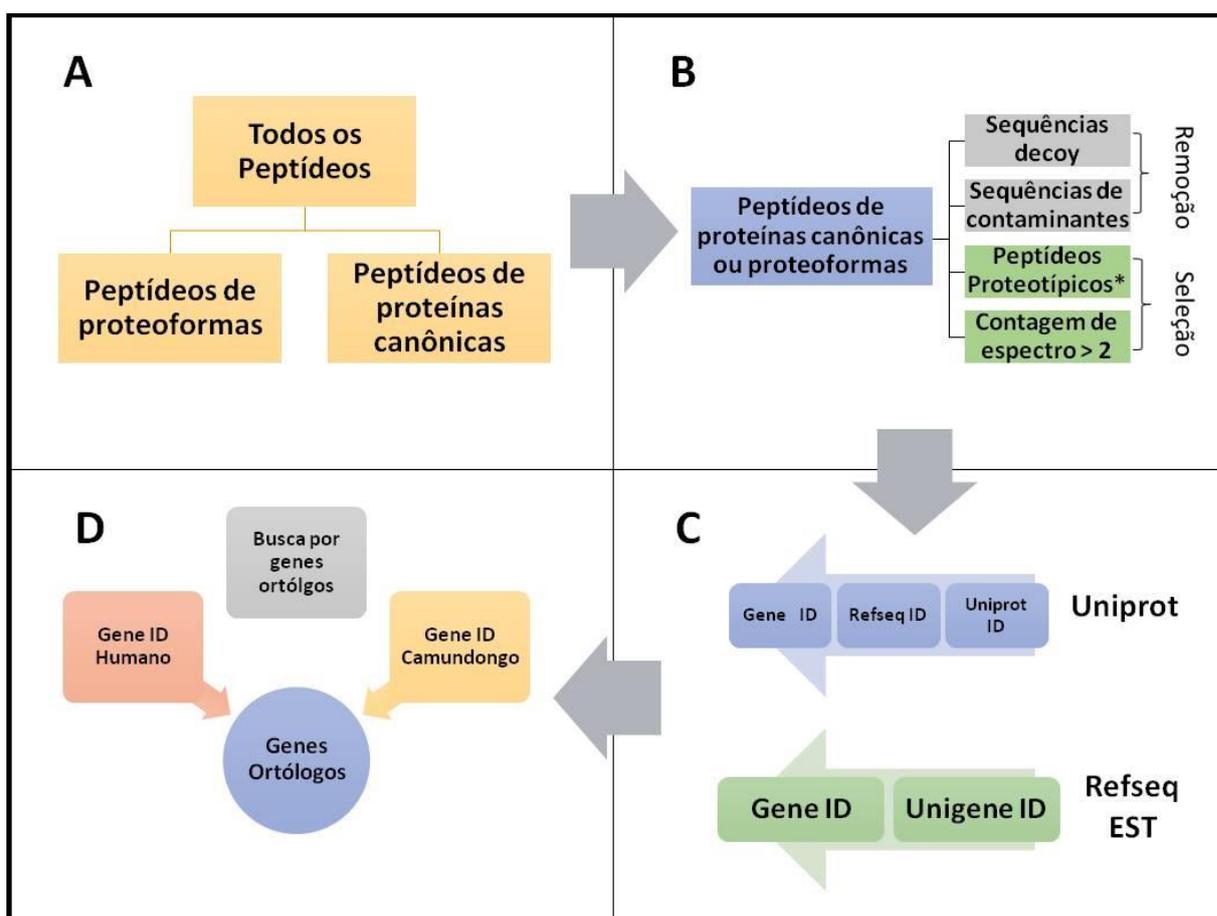


Figura 3.4 – Esquema do fluxo de análises para identificação de peptídeos e proteínas de genes ortólogos. *Para proteoformas também foram aceitos peptídeos compartilhados apenas entre outras proteoformas.

3.7 Comparação das proteínas identificadas em nossas análises com as proteínas inferidas nos artigos de origem dos conjuntos de dados de MS

Nós obtivemos as proteínas inferidas nos artigos de origem dos conjuntos de dados de MS de humano e camundongo analisados neste projeto, através de tabelas presentes nos materiais suplementares (Tabela 3.3). Destas tabelas nós extraímos apenas as proteínas inferidas por peptídeos proteotípicos e comparamos com nossos resultados, utilizando um programa escrito neste projeto.

Tabela 3.3 – Lista de tabelas suplementares contendo as proteínas inferidas nos artigos de origem dos dados de MS.

Organismo	Humano			Camundongo		
Região do Cérebro	Corpo Caloso	Nervo Óptico	Bulbo Olfatório	Corpo Caloso	Nervo Óptico	Bulbo Olfatório
Referência	(129)	(130)	(131)	(74)		
Tabela	tabela 1	tabela 1	tabela S1	tabela 17		

3.8 Visualização dos eventos de *splicing* e confecção de figuras

As proteoformas e seus respectivos peptídeos foram alinhados aos genomas de humano e camundongo, utilizando a versão online do programa BLAT (134) e visualizadas com a versão online do programa *genome browser* (142). Após isso, foram confeccionadas figuras para a representação dos diferentes eventos de *splicing*, através do programa *powerpoint* (versão 2016).

4 RESULTADOS

4.1 Repositório SpliceProt e repositório personalizado de humano e camundongo

O repositório SpliceProt humano previamente desenvolvido por nosso grupo de pesquisa (113) foi composto por 159.719 sequências proteicas. O repositório de sequências proteicas personalizado humano também construído por nosso grupo de pesquisa (7), continha 20.150 sequências canônicas e 204.294 peptídeos não redundantes, totalizando 224.453 sequências.

Para a construção do repositório SpliceProt de camundongo, nós realizamos a identificação de 217.977 variantes de *splicing* a partir de 4.223.625 transcritos contidos na base de dados Unigene de camundongo. Em seguida, estes transcritos passaram por um processo de tradução *in silico*, utilizando o programa TRANSEQ, o que resultou em 84.278 sequências polipeptídicas não redundantes, dando origem ao repositório SpliceProt de camundongo.

A construção do repositório de sequências proteicas personalizado de camundongo, constituído por sequências proteicas canônicas do UniProt/SwissProt e peptídeos não-redundantes oriundos das proteoformas do SpliceProt e do UniProt/SwissProt, gerou 16.888 sequências canônicas e 156.889 peptídeos não redundantes, totalizando 173.777 sequências.

4.2 Proteínas Canônicas

Todas as proteínas foram inferidas a partir de pelo menos um peptídeo proteotípico com contagem de espectro maior que um. Após isto, foram obtidos os códigos de gene referentes às proteínas inferidas. Através desta abordagem, nós identificamos 1243 genes humanos e 4890 genes de camundongo nos dados de proteômica *shotgun* de corpo caloso. Nos dados de nervo óptico, nós identificamos 2186 genes de humano e 4836 genes de camundongo e nos de bulbo olfatório, 3733 humanos e 5753 de camundongo (Tabela 4.1).

Tabela 4.1 – Quantidade de genes, proteínas canônicas inferidas e peptídeos identificados nos dados de proteômica *shotgun* de corpo caloso, nervo óptico e bulbo olfatório do cérebro de humano e camundongo.

Região Cerebral	Corpo Caloso		Nervo Óptico		Bulbo Olfatório	
Organismo	Humano	Camundongo	Humano	Camundongo	Humano	Camundongo
Genes	1233	4834	2158	4790	3690	5692
Peptídeos	6336	29631	10895	24881	21835	48510
Proteínas	1234	4836	2160	4793	3692	5697

Ao analisarmos a distribuição dos genes, peptídeos e proteínas nas três regiões do cérebro estudadas neste projeto, observamos que 956 genes, 2064 peptídeos e 951 proteínas eram comuns ao corpo caloso, nervo óptico e bulbo olfatório de humano (Figura 4.1). Em camundongo, 3633 genes, 12121 peptídeos e 3592 proteínas eram comuns às três regiões do cérebro (Figura 4.2).

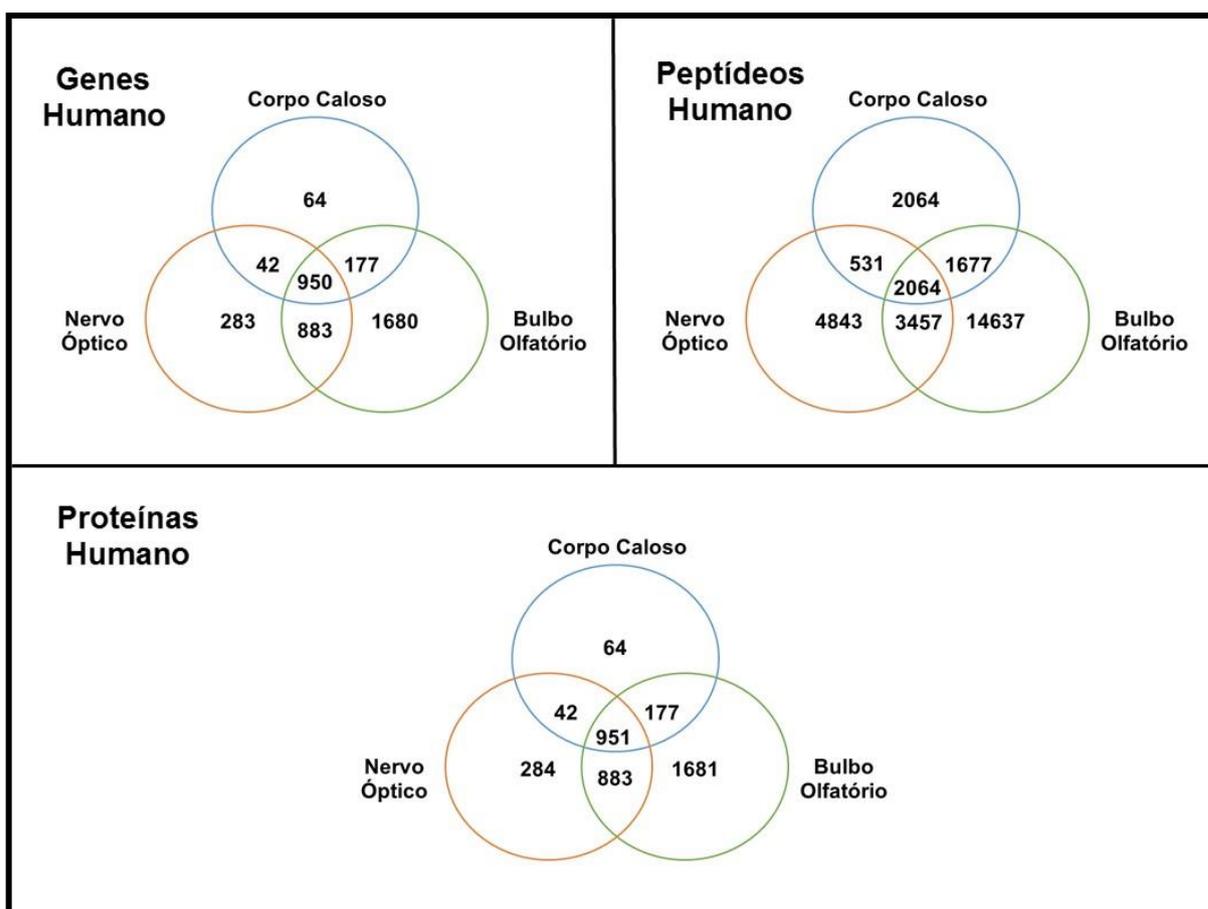


Figura 4.1 – Diagramas de Venn demonstrando a distribuição dos genes, peptídeos e proteínas entre os tecidos: corpo caloso, nervo óptico e bulbo olfatório de humano.

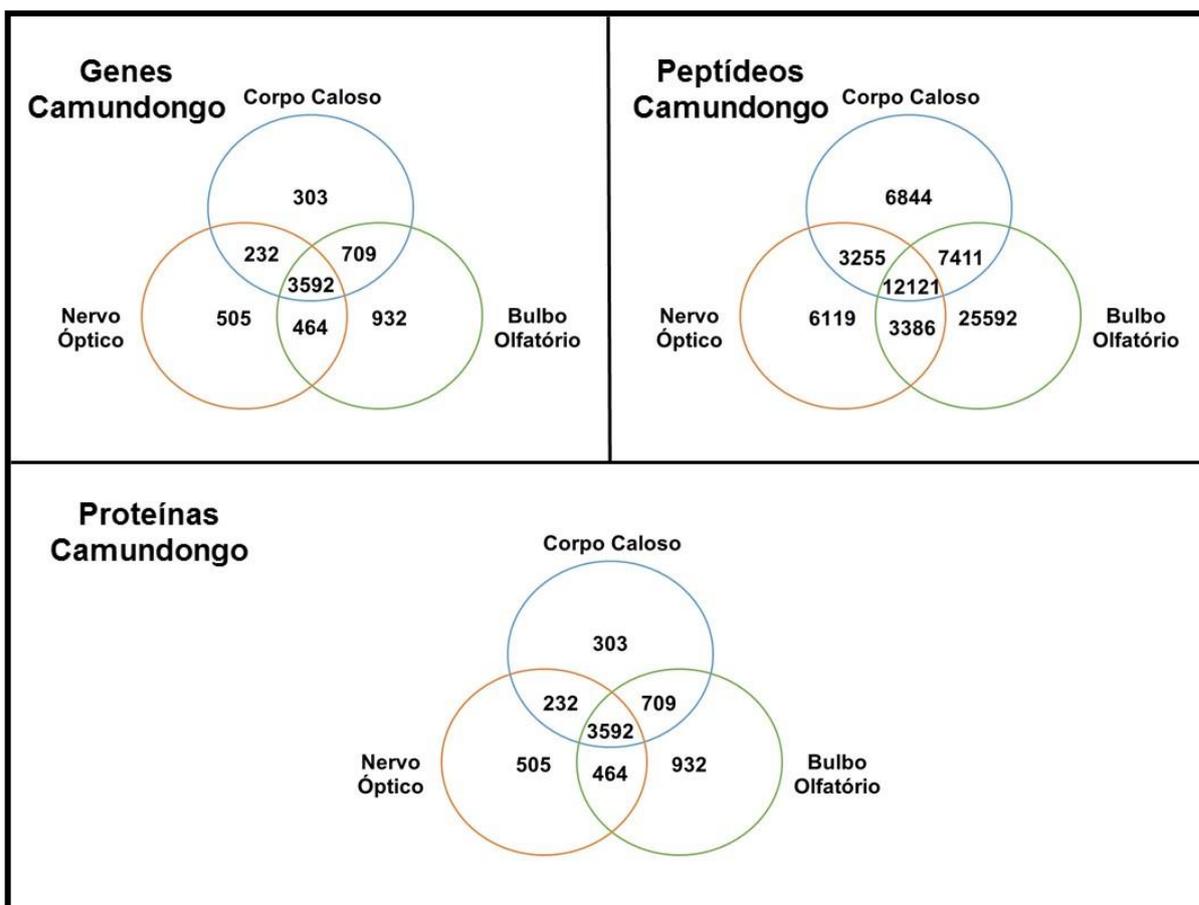


Figura 4.2 – Diagramas de Venn demonstrando a distribuição dos genes, peptídeos e proteínas entre os tecidos: corpo caloso, nervo óptico e bulbo olfatório de camundongo.

Dos resultados obtidos por nossas análises 1172 proteínas (95%) de corpo caloso, 1962 (90%) de nervo óptico e 3586 (97%) de bulbo olfatório foram comuns aos que os autores dos artigos de origem dos dados de MS de humano encontraram (Figura 4.3.1). Em camundongo, 2506 (51%) proteínas de corpo caloso, 2431 (51%) de nervo óptico e 2921 (51%) de bulbo olfatório foram identificadas tanto pelos autores quanto pelas nossas análises (Figura 4.3.2).

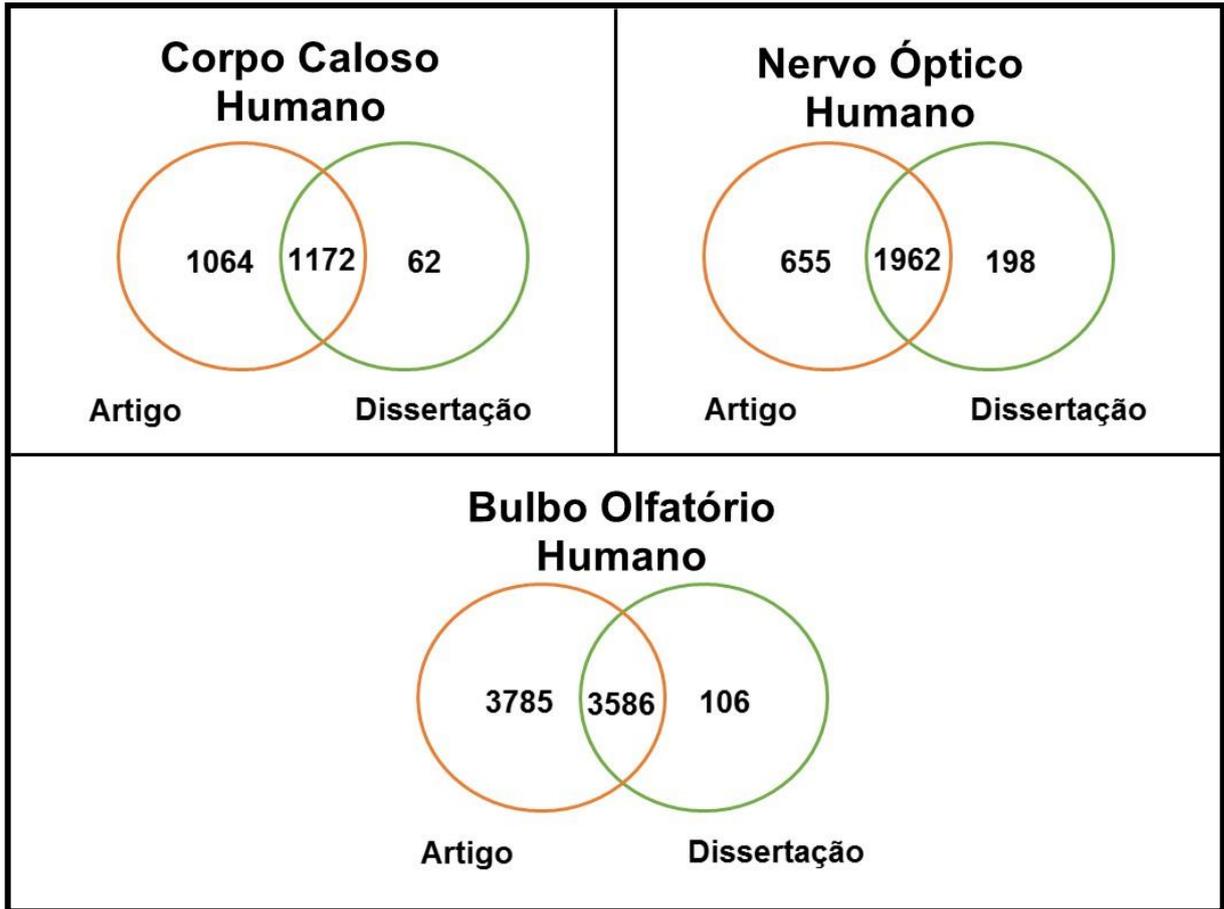


Figura 4.3.1 – Diagramas de Venn demonstrando a distribuição das proteínas identificadas pelos autores dos artigos de origem dos dados de MS e das proteínas identificadas em nossas análises em corpo caloso, nervo óptico e bulbo olfatório de humano.

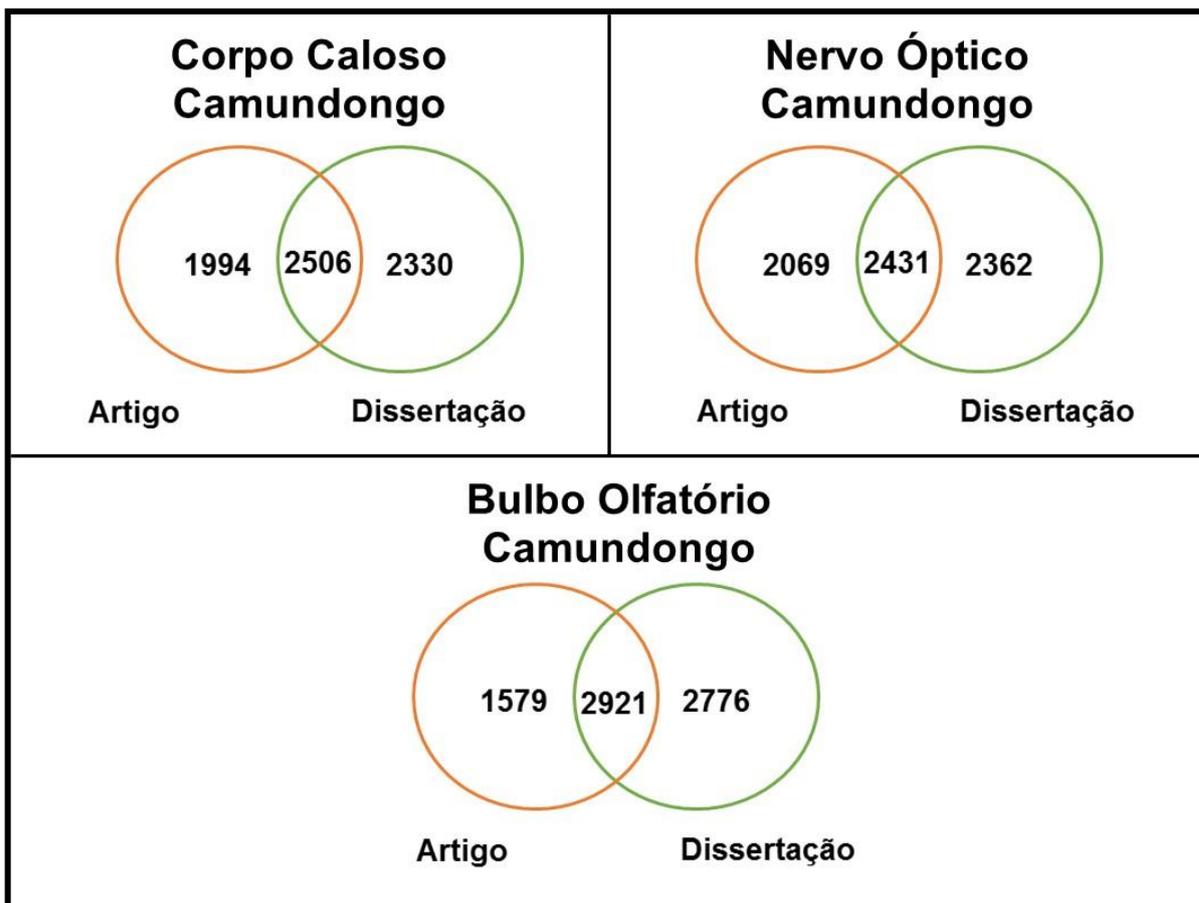


Figura 4.3.2 – Diagramas de Venn demonstrando a distribuição das proteínas identificadas pelos autores dos artigos de origem dos dados de MS e das proteínas identificadas em nossas análises em corpo caloso, nervo óptico e bulbo olfatório de camundongo.

4.2.1 Proteínas canônicas de genes ortólogos

Nós identificamos 1043 genes ortólogos entre humano e camundongo de proteínas canônicas de corpo caloso, 1705 genes de nervo óptico e 3055 genes de bulbo olfatório (Tabela 4.2).

Tabela 4.2 – Quantidade de genes ortólogos, proteínas canônicas inferidas e peptídeos de genes ortólogos identificados nos dados de proteômica *shotgun* de corpo caloso, nervo óptico e bulbo olfatório do cérebro de humano e camundongo.

Região Cerebral	Corpo Caloso		Nervo Óptico		Bulbo Olfatório	
Organismo	Humano	Camundongo	Humano	Camundongo	Humano	Camundongo
Genes	1043	1043	1705	1705	3055	3055
Peptídeos	5643	11698	9058	15050	19004	38813
Proteínas	1044	1043	1702	1702	3053	3052

Destes genes, nós observamos que 817 eram comuns aos três tecidos, 38 exclusivos de corpo caloso, 174 de nervo óptico e 1380 de bulbo olfatório (Figura 4.4).

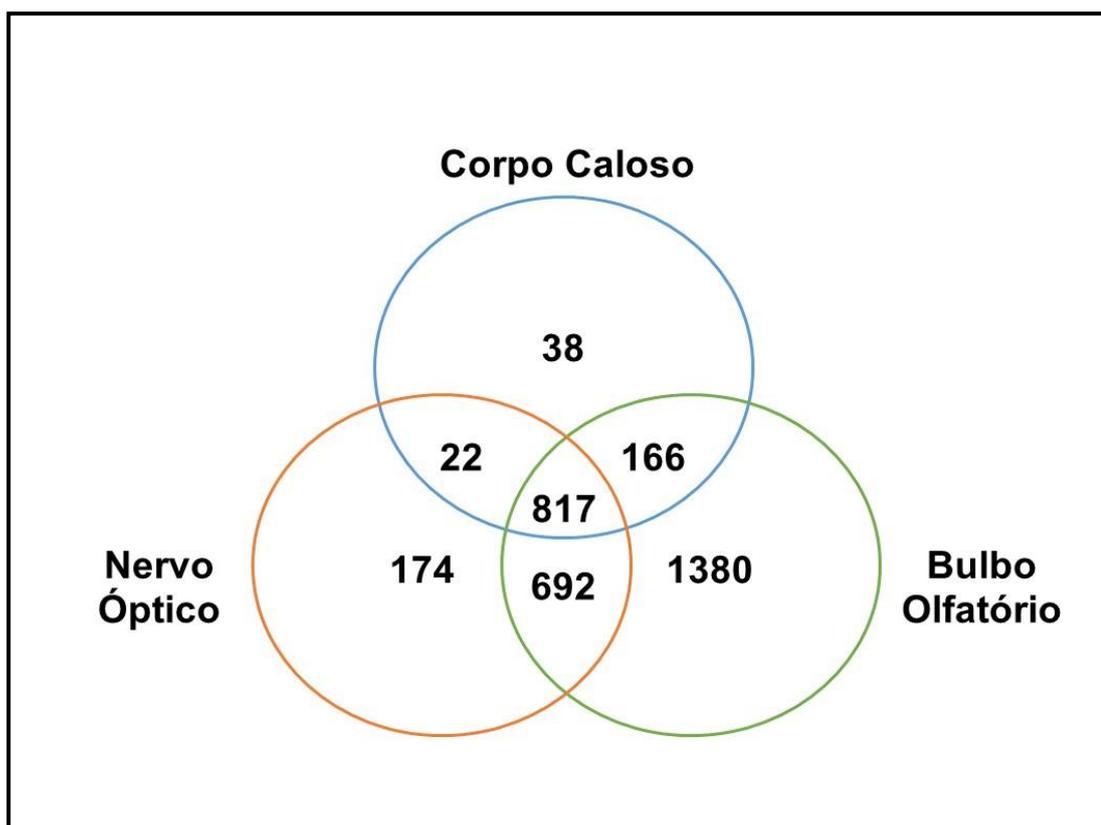


Figura 4.4 – Diagrama de Venn demonstrando os genes ortólogos de proteínas inferidas nos experimentos de proteômica *shotgun* dos três tecidos do cérebro de humano e camundongo analisados neste projeto: corpo caloso, nervo óptico e bulbo olfatório.

4.3 Proteoformas

Nos dados de MS de corpo caloso nós inferimos 10 proteoformas em humano e 3 proteoformas em camundongo. Já nos dados de MS de nervo óptico, nós inferimos 13 proteoformas em humano e 4 proteoformas em camundongo. Por fim, nos dados de MS de bulbo olfatório, nós inferimos 23 proteoformas em humano e 5 proteoformas em camundongo (Tabela 4.3). Para a inferência das proteoformas nós utilizamos apenas peptídeos proteotípicos os quais podem ser conferidos nas tabelas 4.4, 4.5 e 4.6.

Tabela 4.3 – Quantidade de genes, proteoformas e peptídeos inferidos e identificados nos dados de proteômica *shotgun* de corpo caloso, nervo óptico e bulbo olfatório do cérebro de humano e camundongo.

Região Cerebral	Corpo Caloso		Nervo Óptico		Bulbo Olfatório	
	Humano	Camundongo	Humano	Camundongo	Humano	Camundongo
Genes	10	3	13	4	23	5
Proteoformas	10	3	13	4	23	5
Peptídeos	12	3	14	4	28	5

Tabela 4.4 – Peptídeos proteotípicos de proteoformas identificados nos dados de MS de corpo caloso.

Espécie	Gene ID	Peptídeo	Contagem de Espectros
Humano	10529	GCFHCEVCK	3
	10529	TQEQISNVK	4
	1809	GSGSRPGIEGDTPR	4
	1809	TLDFDALSVMQR	3
	23564	GVPESLASGEGAGALPALDLGK	3
	2670	GAEM+15.994900M+15.994900ELNDR*	2
	302	GTNQLQEIINR	2
	51295	ELQNLQAPLVLR	14
	5315	GGVNLPGAAVDLPVSEK	2
	5690	DGYELSPTAAANFTR	2
	56948	VLVGGGTGFIGTALTQLLNAR	6
	7167	EAGVSDAVAQSTR	2
	Camundongo	12322	SSSSVQLMESSESTNTTIEDTK
20320		RPDEVPDAGPMK	2
210274		NSLYSDCIIEDK	3

*A anotação "M+15.994900" refere-se a um aminoácido metionina oxidado.

Tabela 4.5 – Peptídeos proteotípicos de proteoformas de genes ortólogos identificados nos dados de MS de nervo óptico.

Espécie	Gene ID	Peptídeo	Contagem de Espectros
Humano	10382	EAATFIGNSTAIQELFK	3
	1295	VEIPLASLR	2
	3043	GSAVTALWGK	6
	3845	QGVDDAFYTLVR	7
	4134	LSSTEEAVLNQAPQQK	2
	4134	SQSDLDFPGAADIPTR	13
	5315	GGVNLPGAAVDLPAVSEK	2
	5339	VPDVQDGELQLR	8
	5690	DGYELSPTAAANFTR	7
	57142	SSAVVDLLYWR	2
	60	HQGVM+15.994900VGIVTNWDDM+15.994900EK*	2
	81606	M+15.994900QVSTSRPAAAGLSCP DYLR	41
	8856	APAIPLHSAAAELASNHPR	3
	90231	GLGLALSPVSR	3
	Camundongo	12322	SSSSVQLMESSESTNTTIEDTK
12661		GCGKPISEEGATLGEK	3
13518		YSEVQVSIPDK	3
20320		RPDEVDPDAGPMK	5

*A anotação "M+15.9949900" refere-se a um aminoácido metionina oxidado.

Tabela 4.6 – Peptídeos proteotípicos de proteoforma identificadas nos dados de MS de bulbo olfatório.

Espécie	Gene ID	Peptídeo	Contagem de Espectros	
Humano	10382	EAATFIGNSTAIQELFK	3	
	1400	YGGMFAAVEGAYENK	2	
	1809	EPAPASPAPAGVEIR	6	
	1809	EVLQNLGPK	12	
	1809	TLDFDALSVGQR	2	
	2037	EEQEYEEEEVEEPRPAKPPVVK	2	
	213	LVRPEVDVM+15.994900CTAFHDNEETFLLR*	8	
	22924	QGQDVAPPNPVPQR	3	
	23705	AHSDYMLYVYDSR	2	
	2597	VIHDNFGIVEGLM+15.994900VTVHAITATQK	2	
	2597	VIHDNFGIVEGLMVTV	2	
	3043	VNVDVAVGGEALGSAVTALWGK	2	
	309	EDAQEIADTPSGDK	6	
	3329	ALMLQGVDLLAK	2	
	4134	EIPPFICEEPVAK	2	
	51062	ELGAVIDQVAAALWDQALYK	4	
	5339	VPDVQDGELQLR	4	
	5690	DGYELSPTAAANFTR	2	
	56948	VLVGGGTGFIGTALTQLLNAR	3	
	60	HQGVM+15.994900VGIVTNWDDMEK*	9	
	60	HQGVMVGIVTNWDDM+15.994900EK*	2	
	60	HQGVMVGIVTNWDDMEK	3	
	6262	VLDIANVLFHLEQVEHPQR	3	
	7419	SCSGVMEFSTSGHAYTDTGK	4	
	7534	TVFDEAIAELDTLSEESYK	2	
	818	SSSSVHLMEPQTTVVHNATDGIK	3	
	81	MLDAEDIVGTLRPDEK	9	
	998	NVFDEAILAALPETQPK	12	
	Camundongo	16706	FSFPAAYFIHHR	3
		20320	RPDEVPDAGPMK	2
27062		LAELVIEVLQQNEEHHAEGK	14	
83679		LNLQVLLSHVLGK	3	
94090		NVQVTLHTGLPVPDFYSSR	8	

*A anotação "M+15.994900" refere-se a um aminoácido metionina oxidado.

4.3.1 Proteoformas de genes ortólogos

Ao procurarmos por proteoformas de genes ortólogos entre humano e camundongo nos dados de MS analisados neste projeto, admitimos proteoformas inferidas tanto por peptídeos compartilhados entre outras proteoformas quanto por

peptídeos proteotípicos. A partir deste conjunto de dados, inferimos proteoformas de 5 genes ortólogos entre humano e camundongo em corpo caloso, 11 em nervo óptico e 19 em bulbo olfatório (Tabela 4.7). Os peptídeos identificados e suas respectivas contagens de espectros podem ser conferidos na Tabela 4.8.

Tabela 4.7 – Genes ortólogos entre humano e camundongo que expressam proteoformas inferidas nos dados de espectrometria de massas de corpo caloso, nervo óptico e bulbo olfatório.

Organismo	Humano		Camundongo		
	Região do Cérebro	Gene ID	Símbolo Gênico	Gene ID	Símbolo Gênico
Corpo Caloso		5315	PKM	18746	<i>Pkm</i>
		6709	SPTAN1	20740	<i>Sptan1</i>
		7168	TPM1	22003	<i>Tpm1</i>
		7170	TPM3	59069	<i>Tpm3</i>
		58473	PLEKHB1	27276	<i>Plekhb1</i>
Nervo Óptico		1759	DNM1	13429	<i>Dnm1</i>
		4134	MAP4	17758	<i>Map4</i>
		4155	MBP	17196	<i>Mbp</i>
		5315	PKM	18746	<i>Pkm</i>
		5339	PLEC	18810	<i>Plec</i>
		6709	SPTAN1	20740	<i>Sptan1</i>
		7168	TPM1	22003	<i>Tpm1</i>
		7170	TPM3	59069	<i>Tpm3</i>
		23114	NFASC	269116	<i>Nfasc</i>
	23136	EPB41L3	13823	<i>Epb41l3</i>	
	58473	PLEKHB1	27276	<i>Plekhb1</i>	
Bulbo Olfatório		1400	CRMP1	12933	<i>Crmp1</i>
		1759	DNM1	13429	<i>Dnm1</i>
		3190	HNRNPK	15387	<i>Hnrnpk</i>
		4134	MAP4	17758	<i>Map4</i>
		4137	MAPT	17762	<i>Mapt</i>
		5315	PKM	18746	<i>Pkm</i>
		5579	PRKCB	18751	<i>Prkcb</i>
		6812	STXBP1	20910	<i>Stxbp1</i>
		7168	TPM1	22003	<i>Tpm1</i>
		7169	TPM2	22004	<i>Tpm2</i>
		7170	TPM3	59069	<i>Tpm3</i>
		8573	CASK	12361	<i>Cask</i>
		9066	SYT7	54525	<i>Syt7</i>
		10313	RTN3	20168	<i>Rtn3</i>
		23114	NFASC	269116	<i>Nfasc</i>
	23136	EPB41L3	13823	<i>Epb41l3</i>	
	23499	MACF1	11426	<i>Macf1</i>	
	23705	CADM1	54725	<i>Cadm1</i>	
	65009	NDRG4	234593	<i>Ndr4</i>	

Legenda: Os genes ortólogos marcados em cinza possuem proteoformas inferidas nos dados das três regiões do cérebro de camundongo analisadas neste projeto: corpo caloso, nervo óptico e bulbo olfatório.

Tabela 4.8 – Peptídeos identificados nos experimentos de MS de humano e camundongo e suas respectivas contagens de espectros de acordo com as três regiões do cérebro analisadas neste projeto.

Amostras		Corpo Caloso		Nervo Optico		Bulbo Olfatório	
Gene	Peptídeo	Homem	Mouse	Homem	Mouse	Homem	Mouse
PLEC	AIQHEISSLK	-	-	-	2	-	-
	VPDVQDGELQLR	-	-	8	-	-	-
	SGGGSVGNQSVLDPAER	-	-	-	2	-	-
NFASC	VYSDTVQGQLR	-	-	-	6	-	-
	GPEPDTIIGYSGEDYPR	-	-	-	8	-	2
	EFTTPEGVPSAPR	-	-	-	3	-	-
	GPEPESVIGYSGEDLPSAPR	-	-	3	-	8	-
	IGYSGEDLPSAPR	-	-	-	-	2	-
MAP4	EIPPFICEEPVAK	-	-	-	-	2	-
	LNEIGLNVSMGQLK	-	-	-	10	-	-
	SSEEGPMLDQVPQQK	-	-	-	4	-	4
	ELGIAGLYLHSK	-	-	-	-	-	6
	EIGQDGFQWQR	-	-	-	3	-	2
	ALTAALAAEDELQSK	-	-	-	3	-	-
	NGLVSSCSVTGVMDSGQLNNK	-	-	-	2	-	-
	AAEFHPGLTEGK	-	-	-	4	-	5
	EIEVTATQSIPSLLEETPR	-	-	-	8	-	8
	QPAALTAALAAEDELQSK	-	-	-	2	-	-
	LSSTEEAVLNQAPQQK	-	-	2	-	-	-
	NSSFLEQNK	-	-	-	-	-	3
	VDSCPFICLGGEK	-	-	-	2	-	2
	ALSECYHLSVPPAVNLVDK	-	-	-	2	-	-
LEQIPEGSHK	-	-	-	4	-	3	
SQSDLDFPGAADIPTR	-	-	13	-	-	-	
TPM1	AEDSLLAAEEAAAK	-	-	2	-	2	-
	SLQEQADAAEER	2	2	5	-	7	4
	ETAADVASLNR	3	-	3	-	-	-
	SIDDLEDQLYQQLEQNR	-	-	-	-	4	-
	SIDDLEDQLYHQLEQNR	-	3	-	-	-	9
	AEDSLLAADETAAK	-	2	-	2	-	2

TPM2	SLIASEEEYSTK	-	-	-	-	-	3
	SLMASEEEYSTK	-	-	-	-	2	-
TPM3	IQVLQQQADDAEER	2	3	8	3	7	4
	EEHLCTQR	-	2	-	-	-	-
	EQAEAEVASLNR	2	3	2	2	4	3
	TIDDLEDK	-	3	-	-	-	-
NDRG4	YFLQGMGYM+15.994900PSASMTR	-	-	-	-	3	-
	YFLQGM+15.994900GYMPSASMTR	-	-	-	-	-	2
	YFLQGMGYMPSASMTR	-	-	-	-	3	4
SORBS1	SATASPQQPQAQQR	2	-	-	-	-	-
	STQDLSDVSTDEVGIPLR	-	5	-	-	-	-
MBP	SPLPSHAR	-	-	-	4	-	-
	GAYDAQGTLK	-	-	-	17	-	-
	PGFGYGGR	-	-	-	5	-	-
	FSWGAEGQKPGFGYGGR	-	-	-	7	-	-
	VPWLKPGR	-	-	7	-	-	-
SPTAN1	QEQIDNQYHSLLELGEK	4	-	8	-	-	-
	QGQIDNQYQSLLELGEK	-	9	-	13	-	-
PKM	LLFEELVR	-	14	-	9	-	18
	EAEAAMFHR	-	5	12	5	3	8
	EAEAAM+15.994900FHR	-	-	-	2	-	-
	ALIVLTESGR	-	-	5	-	-	-
	ASSHSTDLM EAMAMGSVEASYK	-	20	2	38	74	44
	ASSHSTDLM EAMAM+15.994900GSVEASYK	-	23	-	6	-	24
	CLAAALIVLTESGR	27	3	10	-	13	16
	ASSHSTDLM+15.994900EAMAMGSVEASYK	-	7	-	5	-	9
	ASSHSTDLM+15.994900EAM+15.994900AMGSVEASYK	-	-	-	-	-	5
	LFEELVR	2	-	20	-	-	-
	ASSHSTDLM EAM+15.994900AMGSVEASYK	-	7	-	8	-	8
MAMGSVEASYK	-	-	-	3	-	-	
ASSHSTDLM+15.994900EAMAM+15.994900GSVEASYK	-	2	-	-	-	-	
HNRNPK	IIPMLEEYQHYK	-	-	-	-	5	5
CRMP1	AWNTEDDLPVYLARPGSAAQTPR	-	-	-	-	-	2
	TIDFDAYSVGR	-	-	-	-	-	3

	YGGMFAAVEGAYENK	-	-	-	-	2	5
	YGGM+15.994900FAAVEGAYENK	-	-	-	-	-	5
MACF1	IGQSLSSLTSPAEQGVLSEK	-	-	-	-	2	-
	KPQASLPGMEQFSGACLEK	-	-	-	-	-	6
	QELPALQQNQSDLK	-	-	-	-	-	4
	QFHETAEPISDFLSVTEK	-	-	-	-	-	4
SYT7	LGEKPAPVPPPGEDALR	-	-	-	-	2	-
	EFILNISPYAPYGDPR	-	-	-	-	2	-
	QQSLQQPLSQNQQGR	-	-	-	-	-	3
	SNPGSWDHVVGQIR	-	-	-	-	-	4
EPB41L3	SLDGEVGTGQYATTK	-	-	-	13	-	4
	DSM+15.994900SAAEVGTGQYATTK	-	-	5	-	-	-
	DSMSAAEVGTGQYATTK	-	-	2	-	4	-
MAPT	AEEAGIGDTPSLEDEAAGHVTQ	-	-	-	-	2	-
	AEEAGIGDTPSLEDEAAGHVTQAR	-	-	-	-	5	-
	AEEAGIGDTPNQEDQAAGHVTQAR	-	-	-	-	-	25
PLEKHB1	TALM+15.994900EANSTPVR	-	2	-	3	-	-
	TALMEANSTPVR	-	2	-	-	-	-
	TALLEANSTPVR	3	-	17	-	-	-
RTN3	SCGSSCAVHDLIFWR	-	-	-	-	-	10
	SCSSSCAVHDLIFWR	-	-	-	-	4	-
CASK	HNADLVTYEEVVK	-	-	-	-	2	-
	TQSSSCEDLPSTTQPK	-	-	-	-	2	2
DMN1	TGLFTPDLAFEATVK	-	-	7	6	10	9
	CVDMVVELTATIR	-	-	-	-	4	-
PRKCB	NIDQSEFEGFSFVNSEFLKPEVK	-	-	-	-	2	6
STXBP1	WEVLIGSTHILTPTK	-	-	-	-	6	4
CADM1	AHSDYMLYVYDSR	-	-	-	-	2	5

Destes genes, apenas 3 tiveram proteoformas inferidas nas três regiões do cérebro de humano e camundongo analisadas neste projeto. Dois genes foram exclusivos de nervo óptico, 12 de bulbo olfatório e nenhum de corpo caloso (Figura 4.5).

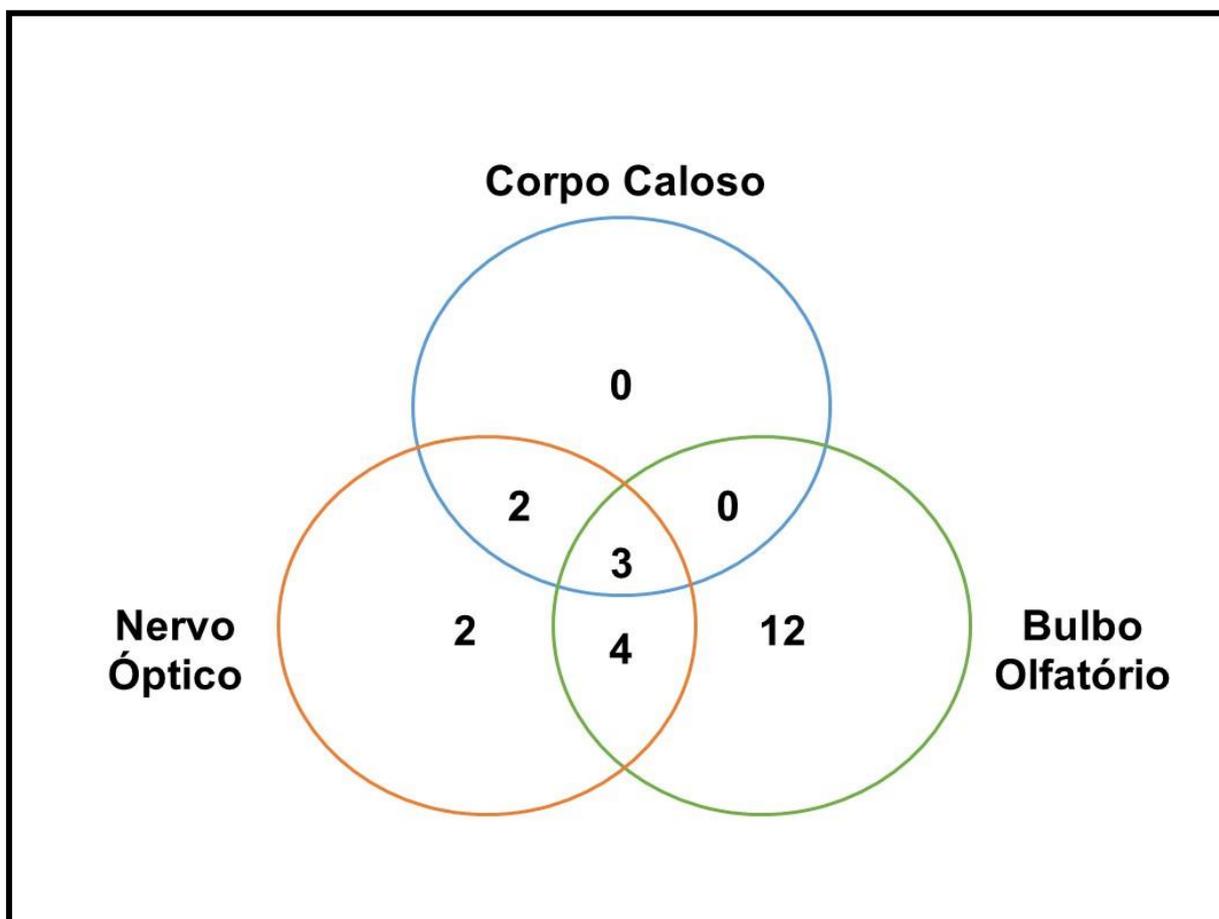


Figura 4.5 – Diagrama de Venn demonstrando os genes ortólogos de proteoformas inferidas nos experimentos de proteômica *shotgun* das três regiões do cérebro de humano e camundongo analisadas neste projeto: corpo caloso, nervo óptico e bulbo olfatório.

Nos próximos subtópicos, apresentaremos os diferentes eventos de *splicing* das proteoformas inferidas nos dados de MS e a localização dos seus respectivos peptídeos, de acordo com as três regiões do cérebro analisadas neste projeto.

4.3.1.1 Proteoformas exclusivas do nervo óptico

A partir das análises dos dados de MS de nervo óptico de humano e camundongo, inferimos proteoformas de dois genes ortólogos entre humano e camundongo (MBP e PLEC).

Para o gene MBP de humano, identificamos um peptídeo (VPWLKPGR) correspondente a duas proteoformas (P02686-3 e P02686-4/NP_002376). Este peptídeo mapeou inteiramente em um éxon e evidenciou um evento de uso alternativo de éxon (Figura 4.6.1). Para o gene ortólogo de camundongo (Mbp) nós identificamos quatro peptídeos (“FSWGAEGQKPGFGYGGR”; “PGFGYGGR”; “GAYDAQGTLSK” e “SPLPSHAR”) compartilhados entre as proteoformas P04370-4; P04370-5; P04370-6; P04370-9; P04370-10; P04370-11 e P04370-13. O mapeamento destes peptídeos ao genoma evidenciou o uso alternativo de 3 éxons distintos (Figura 4.6.2).

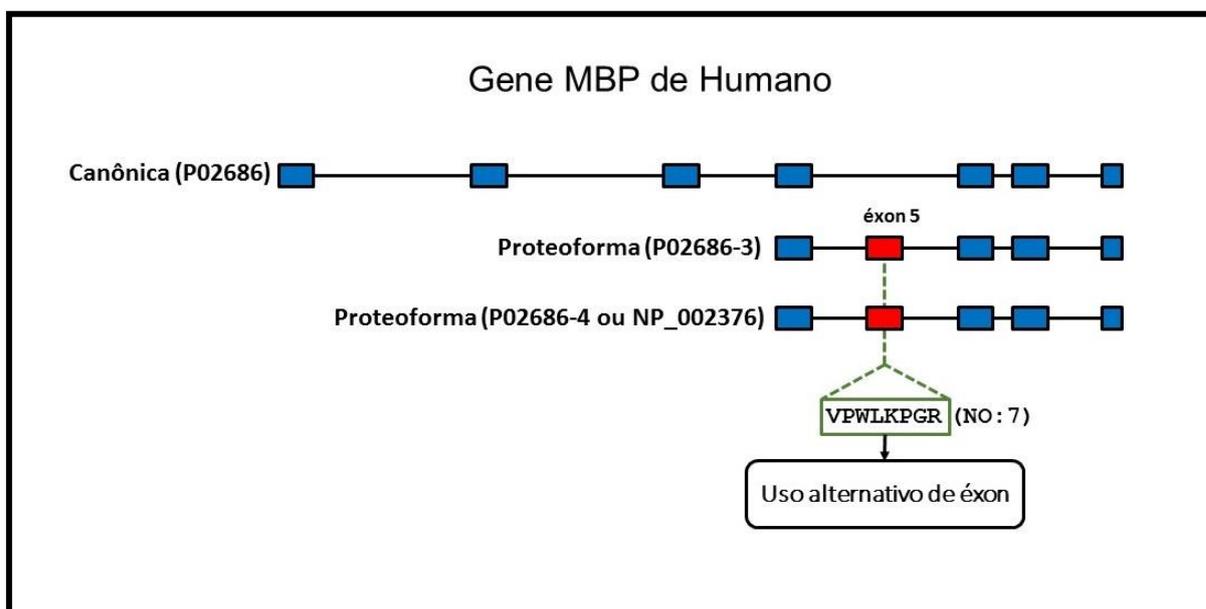


Figura 4.6.1 – Esquema representativo das diferenças no padrão de *splicing* da proteína canônica e proteoforma do gene MBP de humano. Retângulos vermelhos representam os éxons exclusivos de proteoformas onde os peptídeos identificados em nossas análises mapearam. Retângulos azuis representam os éxons das proteínas canônicas e proteoformas que não tiveram um peptídeo correspondente. O quadro verde representa o peptídeo identificado compartilhado entre proteoformas e as linhas pontilhadas verdes representam sua respectiva localização nos éxons das proteoformas. Entre parênteses a contagem de espectros de cada peptídeo no nervo óptico (NO).

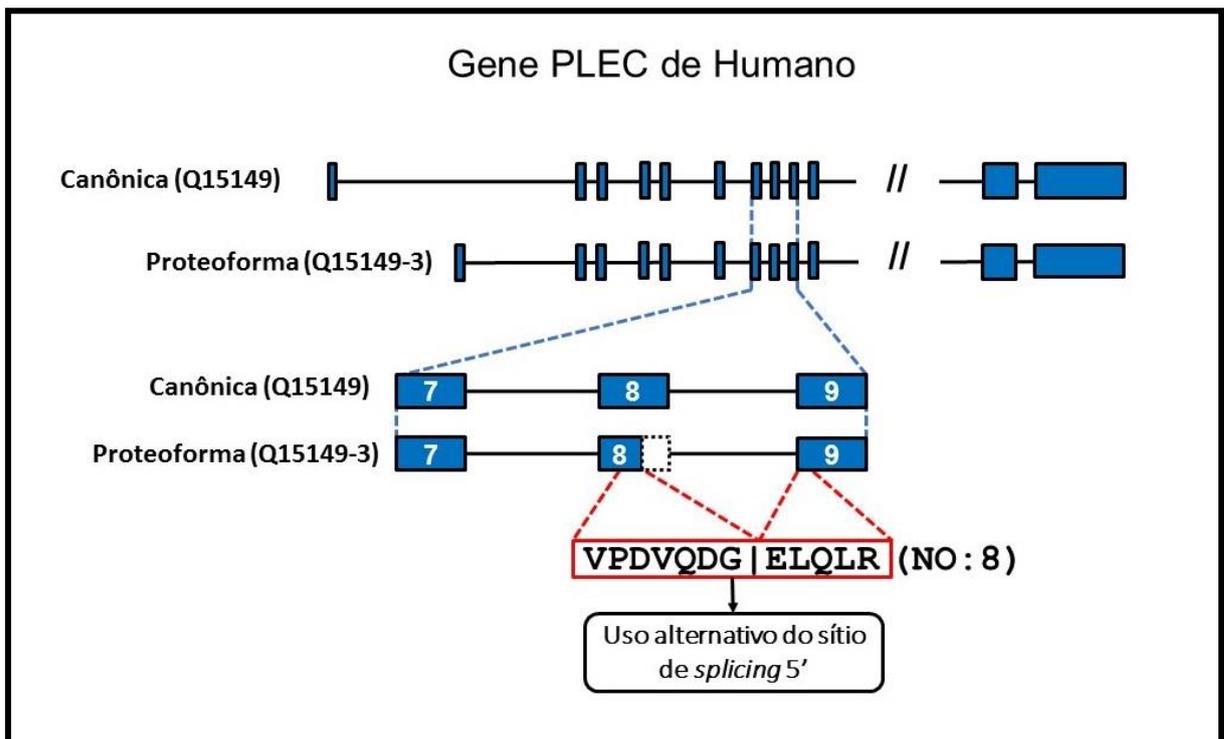


Figura 4.7.1 – Esquema representativo das diferenças no padrão de *splicing* da proteína canônica e proteoforma de *splicing* do gene PLEC de humano. Retângulos azuis representam os éxons das proteínas canônicas e proteoformas que não tiveram um peptídeo correspondente. Os retângulos brancos com contorno pontilhado representam a ausência de parte do éxon. O quadro vermelho representa o peptídeo identificado compartilhado entre proteoformas e as linhas pontilhadas vermelhas representam sua respectiva localização nos éxons das proteoformas. As linhas pontilhadas azuis representam a ampliação dos éxons. Entre parênteses a contagem de espectros de cada peptídeo no nervo óptico (NO).

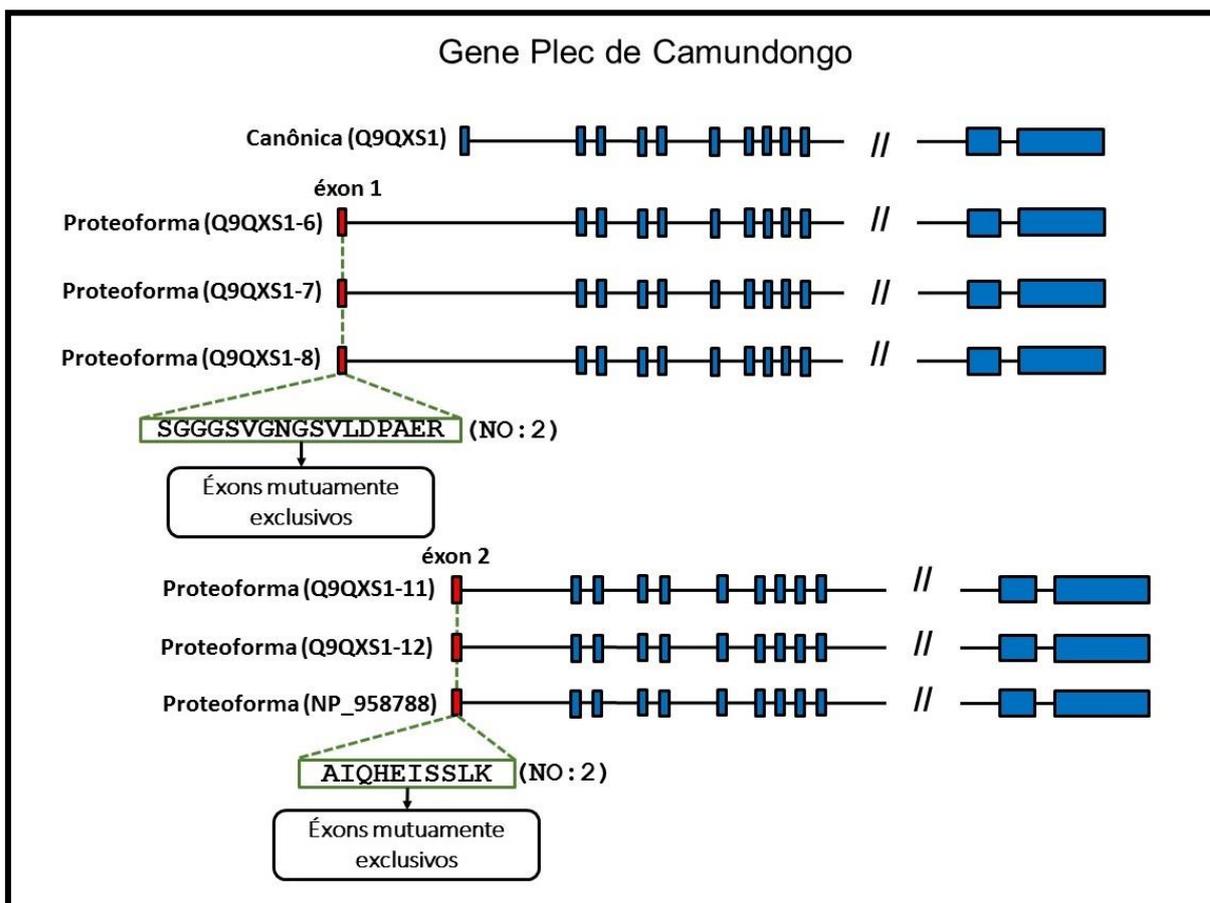


Figura 4.7.2 – Esquema representativo das diferenças no padrão de *splicing* da proteína canônica e proteoformas do gene Plec de camundongo. Retângulos vermelhos representam os éxons exclusivos de proteoformas onde os peptídeos identificados em nossas análises mapearam. Retângulos azuis representam os éxons das proteínas canônicas e proteoformas que não tiveram um peptídeo correspondente. Os quadros verdes representam os peptídeos identificados compartilhados entre proteoformas e as linhas pontilhadas verdes representam suas respectivas localizações nos éxons das proteoformas. Entre parênteses a contagem de espectros de cada peptídeo no nervo óptico (NO).

4.3.1.2 Proteoformas exclusivas de bulbo olfatório

Nos dados de MS de bulbo olfatório de humano e camundongo, inferimos proteoformas de 12 genes ortólogos (CRMP1, HNRNPK, MAPT, PRKCB, STXBP1, TPM2, CASK, SYT7, RTN3, MACF1, CADM1, NDRG4) através de peptídeos proteotípicos e compartilhados entre proteoformas.

Em humano, para o gene TPM2, identificamos um peptídeo (SLMASEEEYSTK) compartilhado entre as proteoformas P07951-2/NP_998839 e P07951-3. Em camundongo, para o gene ortólogo (Tpm2), identificamos um peptídeo proteotípico (SLIASEEEYSTK) referente à proteoforma P58774-2. Ao

mapearmos estes peptídeos aos genomas das respectivas espécies, observamos um evento de uso alternativo de éxon em humano (Figura 4.8.1) e camundongo (Figura 4.8.2).

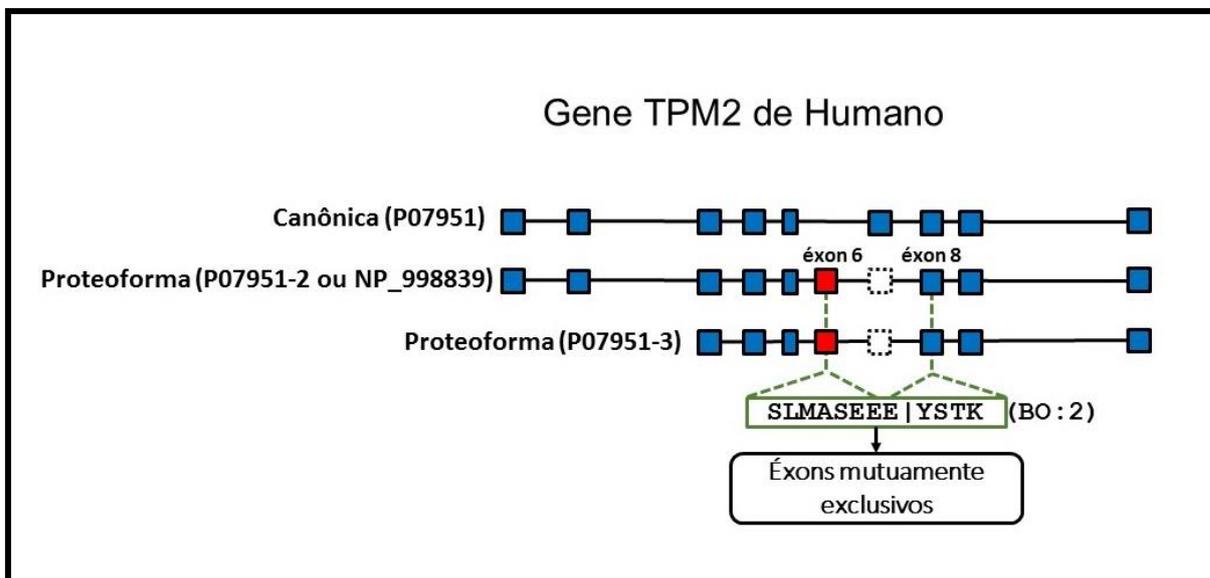


Figura 4.8.1 – Esquema representativo das diferenças no padrão de *splicing* da proteína canônica e proteoforma do gene TPM2 de humano. Retângulos azuis representam os éxons das proteínas canônicas e proteoformas que não tiveram um peptídeo correspondente. Os retângulos brancos com contorno pontilhado representam a ausência de um éxon. O quadro verde representa um peptídeo identificado compartilhado entre proteoformas e as linhas pontilhadas verdes representam sua respectiva localização nos éxons das proteoformas. Entre parênteses a contagem de espectros do peptídeo no bulbo olfatório (BO).

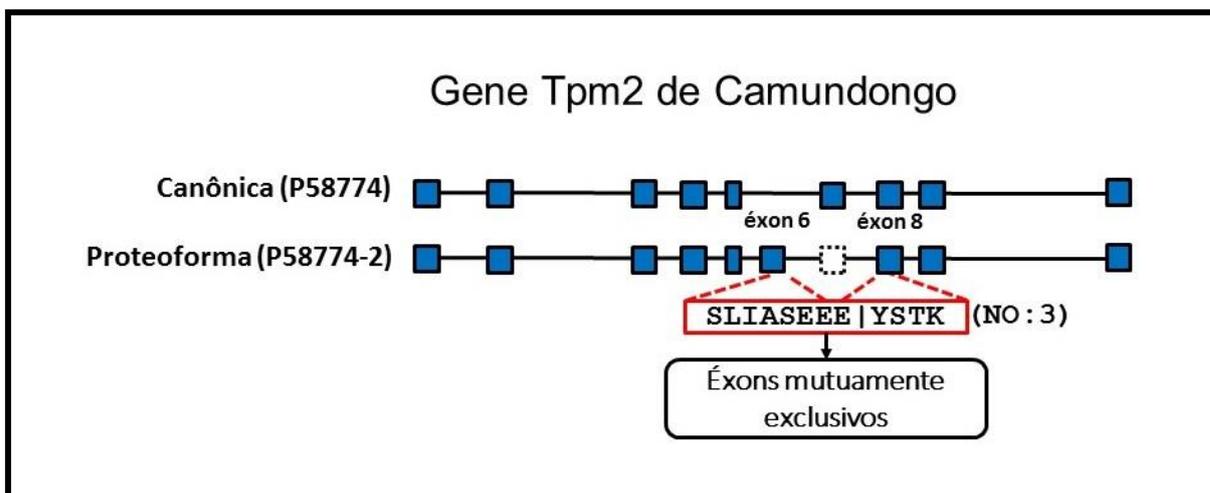


Figura 4.8.2 – Esquema representativo das diferenças no padrão de *splicing* da proteína canônica e proteoforma de *splicing* do gene Tpm2 de camundongo. Os éxons são representados pelos retângulos azuis. Os retângulos brancos com contorno pontilhado representam a ausência de um éxon. O quadro e linhas pontilhadas vermelhas representam o peptídeo proteotípico identificado e

sua respectiva localização nos éxons. Entre parênteses a contagem de espectros de cada peptídeo no bulbo olfatório (BO).

Para o gene NDRG4 de humano, identificamos um peptídeo (YFLQGMGYMPSASMTR) compartilhado entre as proteoformas Q9ULP0-2; Q9ULP0-3/NP_065198/NP_075061; Q9ULP0-5; Q9ULP0-6/NP_001123959; Q9ULP0-7; Q9ULP0-8. Em camundongo, para o gene ortólogo (NdrG4), identificamos o peptídeo “YFLQGMGYMPSASMTR” proteotípico da proteoforma Q8BTG7-2/NP_663577. Tanto o peptídeo de humano quanto o de camundongo mapearam em uma junção entre dois éxons e evidenciam um evento de uso alternativo de éxon (Figuras 4.9.1 e 4.9.2).

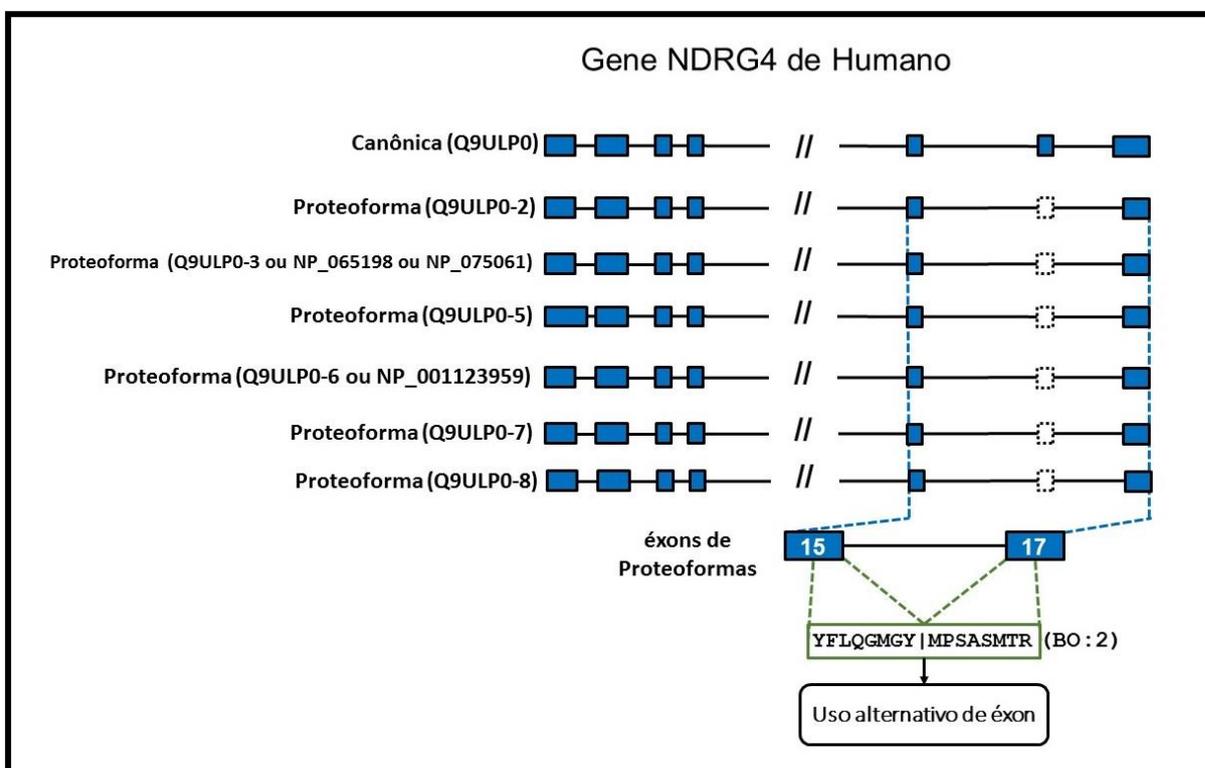


Figura 4.9.1 – Esquema representativo das diferenças no padrão de *splicing* da proteína canônica e proteoforma do gene NDRG4 de humano. Os éxons são representados pelos retângulos azuis. Os retângulos brancos com contorno pontilhado representam a ausência de um éxon. O quadro verde representa o peptídeo identificado compartilhado entre proteoformas e as linhas pontilhadas verdes representam sua respectiva localização nos éxons das proteoformas. As linhas pontilhadas azuis representam a ampliação dos éxons. Entre parênteses a contagem de espectros de cada peptídeo no bulbo olfatório (BO).

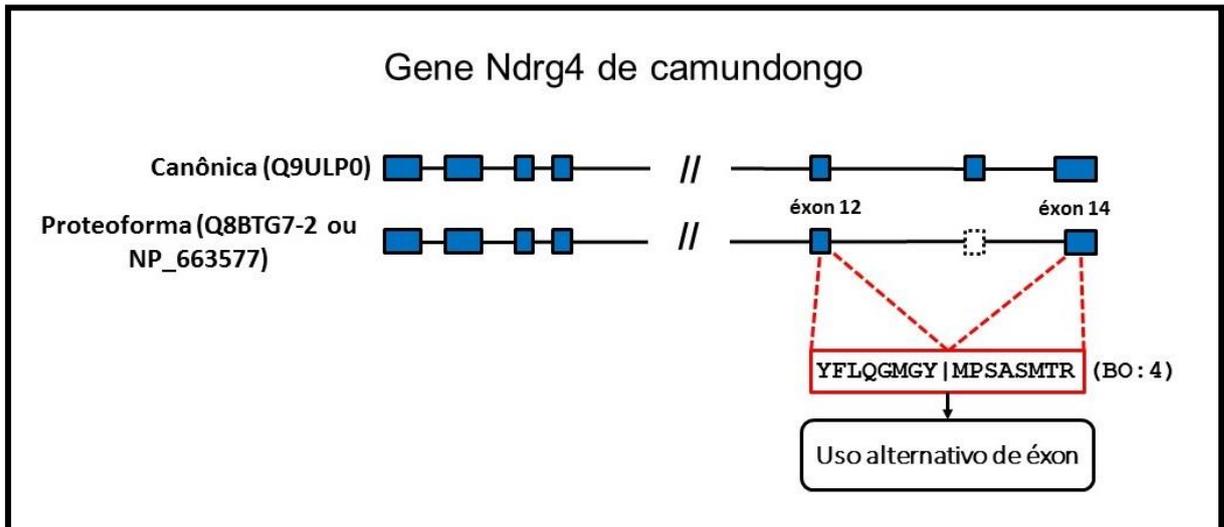


Figura 4.9.2 – Esquema representativo das diferenças no padrão de *splicing* da proteína canônica e proteoforma de *splicing* do gene Ndr4 de camundongo. Os éxons são representados pelos retângulos azuis. Os retângulos brancos com contorno pontilhado representam a ausência de um éxon. O quadro e linhas pontilhadas vermelhas representam o peptídeo proteotípicos identificado e sua respectiva localização nos éxons. Entre parênteses a contagem de espectros de cada peptídeo no bulbo olfatório (BO).

Para o gene HNRNPK de humano e seu ortólogo Hnrnpk de camundongo nós identificamos um peptídeo proteotípico (IIPTLEEYQHLYK) referente às proteoformas P61978-3 de humano e P61979-3 de camundongo. O peptídeo mapeou em uma junção entre dois éxons e evidenciou um evento de uso alternativo de éxon (Figura 4.10).

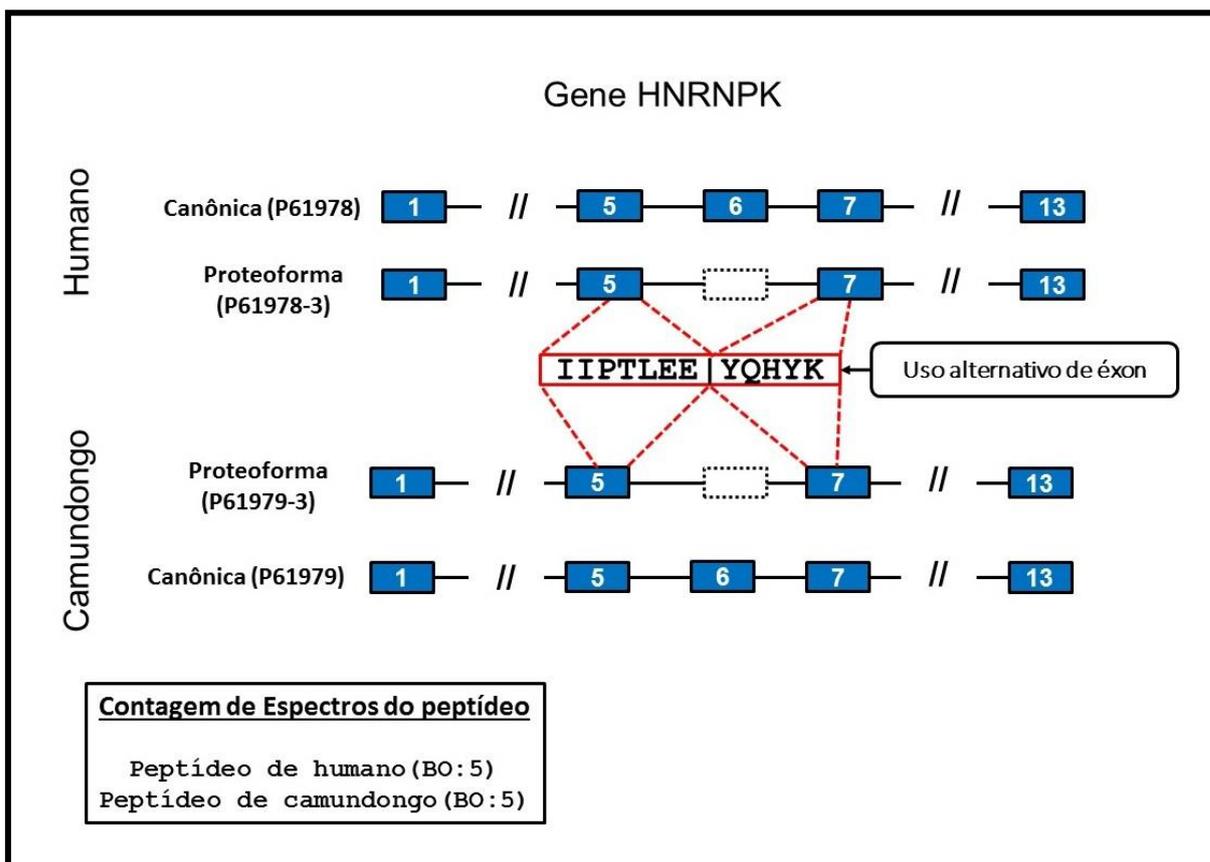


Figura 4.10 – Esquema representativo das diferenças no padrão de *splicing* da proteína canônica e proteoforma de *splicing* do gene HNRNPK de humano e camundongo. Os éxons são representados pelos retângulos azuis. Os retângulos brancos com contorno pontilhado representam a ausência de um éxon. O quadro e linhas pontilhadas vermelhas representam o peptídeo proteotípico identificado e sua respectiva localização nos éxons das proteoformas. Entre parênteses a contagem de espectros de cada peptídeo no bulbo olfatório (BO).

Para o gene CRMP1 humano, identificamos um peptídeo proteotípico (YGGMFAAVEGAYENK) que mapeou inteiramente em um éxon da proteoforma Q14194-2, representando um evento de uso alternativo de éxon. Em camundongo, para o gene ortólogo (Crmp1), identificamos três peptídeos proteotípicos (“AWNTEDDLPVYLARPGSAAQTPR”; “YGGMFAAVEGAYENK” e “TIDFDAYSVGR”) correspondentes a proteoforma de *splicing* NP_001129530. Estes peptídeos mapearam inteiramente em um éxon, evidenciando também um evento de uso alternativo de éxon (Figura 4.11).

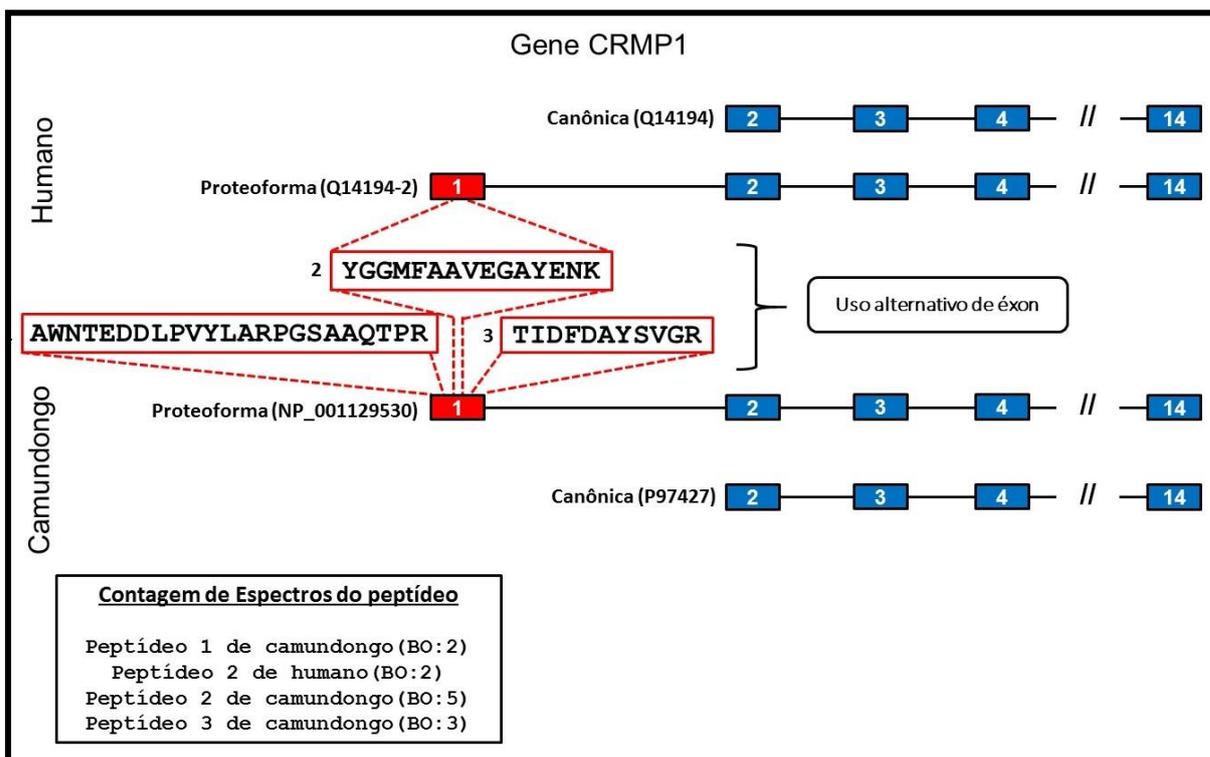


Figura 4.11 – Esquema representativo das diferenças no padrão de *splicing* da proteína canônica e proteoforma de *splicing* do gene CRMP1 de humano e seu ortólogo em camundongo (*Crmp1*). Retângulos vermelhos representam os éxons exclusivos de proteoformas onde os peptídeos identificados em nossas análises mapearam. Retângulos azuis representam os éxons das proteínas canônicas e proteoformas que não tiveram um peptídeo correspondente. O quadro e linhas pontilhadas vermelhas representam o peptídeo proteotípico identificado e sua respectiva localização no éxon da proteoforma. Entre parênteses a contagem de espectros de cada peptídeo no bulbo olfatório (BO).

Em relação ao gene *MACF1* humano nós identificamos um peptídeo (IGQSLSSLTSPAEGVLSSEK) compartilhado com as proteoformas Q9UPN3-2/NP_036222 e Q9UPN3-3. Este peptídeo mapeou inteiramente em um éxon destas proteoformas e representa um evento de uso alternativo de éxon (Figura 4.12.1). Em camundongo, para o gene ortólogo (*Macf1*), identificamos dois peptídeos (“QELPALQQNQSDLK” e “KPQASLPGMEQFSGACLEK”) compartilhados entre duas proteoformas (NP_001186065 e NP_001186066). O primeiro peptídeo mapeou inteiramente em uma porção de um éxon das proteoformas, evidenciando um evento de uso alternativo de sítio 5’ de *splicing*. O segundo peptídeo mapeou inteiramente em uma região do éxon das proteoformas que corresponde a um íntron na proteína canônica, indicando um evento de retenção de íntron. Nós identificamos também um peptídeo (QFHETAEPISDFLSVTEK) compartilhado com estas proteoformas e a proteína canônica, diferindo apenas em dois aminoácidos. Este peptídeo não

corresponde a um evento de *splicing*, porém, evidencia uma possível variação genética (Figura 4.12.2).

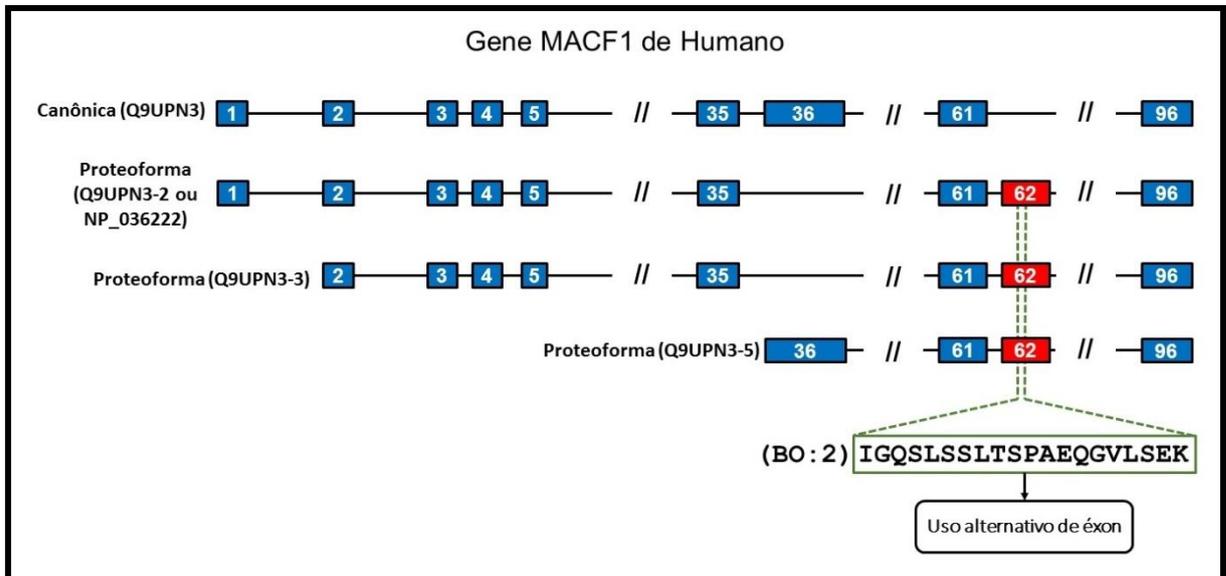


Figura 4.12.1 – Esquema representativo das diferenças no padrão de *splicing* da proteína canônica e proteoformas do gene MACF1 de humano. Retângulos vermelhos representam os éxons exclusivos de proteoformas onde os peptídeos identificados em nossas análises mapearam. Retângulos azuis representam os éxons das proteínas canônicas e proteoformas que não tiveram um peptídeo correspondente. O quadro verde representa um peptídeo identificado compartilhado entre proteoformas e as linhas pontilhadas verdes representam sua respectiva localização nos éxons das proteoformas. Entre parênteses a contagem de espectros de cada peptídeo no bulbo olfatório (BO).

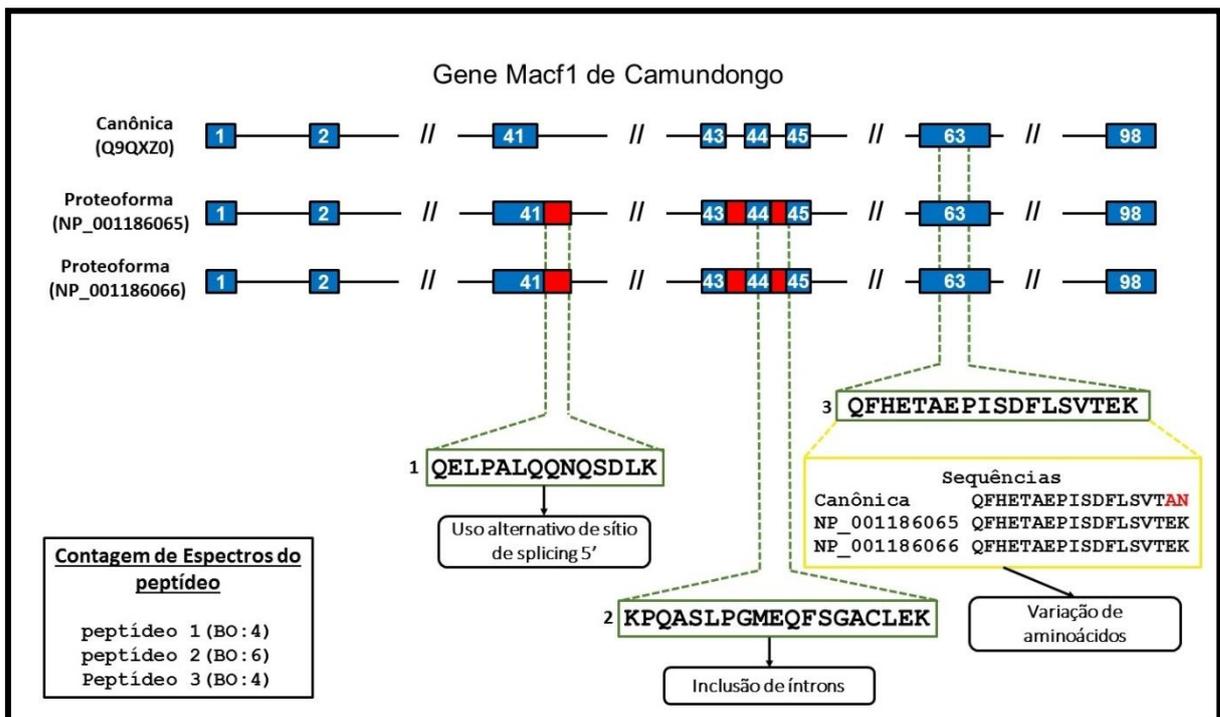


Figura 4.12.2 – Esquema representativo das diferenças no padrão de *splicing* da proteína canônica e proteoformas do gene Macf1 de camundongo. Retângulos vermelhos representam os éxons e introns retidos exclusivos de proteoformas onde os peptídeos identificados em nossas análises mapearam. Retângulos azuis representam os éxons das proteínas canônicas e proteoformas que não tiveram um peptídeo correspondente. Os quadros verdes representam os peptídeos identificados compartilhados entre proteoformas e as linhas pontilhadas verdes representam suas respectivas localizações nos éxons das proteoformas. O quadro e linhas amarelos representam as sequências do peptídeo nas proteoformas e a sequência correspondente na proteína canônica (os aminoácidos diferentes estão marcados em vermelho). Entre parênteses a contagem de espectros de cada peptídeo no bulbo olfatório (BO).

Para o gene SYT7 humano, identificamos dois peptídeos (“LGEKPAPVPPGEDALR” e “EFILNISPYAPYG DPR”) correspondentes a 5 proteoformas (O43581-2; O43581-3; O43581-4; O43581-5; O43581-6). Os peptídeos mapearam inteiramente em éxons evidenciando eventos de uso alternativo de éxons e éxons mutuamente exclusivos (Figura 4.13.1). Para o gene ortólogo de camundongo (Syf7), identificamos dois peptídeos proteotípicos (“QQSLQQPLSQNQQGR” e “SNPGSWDHSVVGQIR”) da proteoforma Q9R0N7-4. Estes peptídeos mapearam inteiramente em um éxon, indicando um evento de uso alternativo de éxon (Figura 4.13.2).

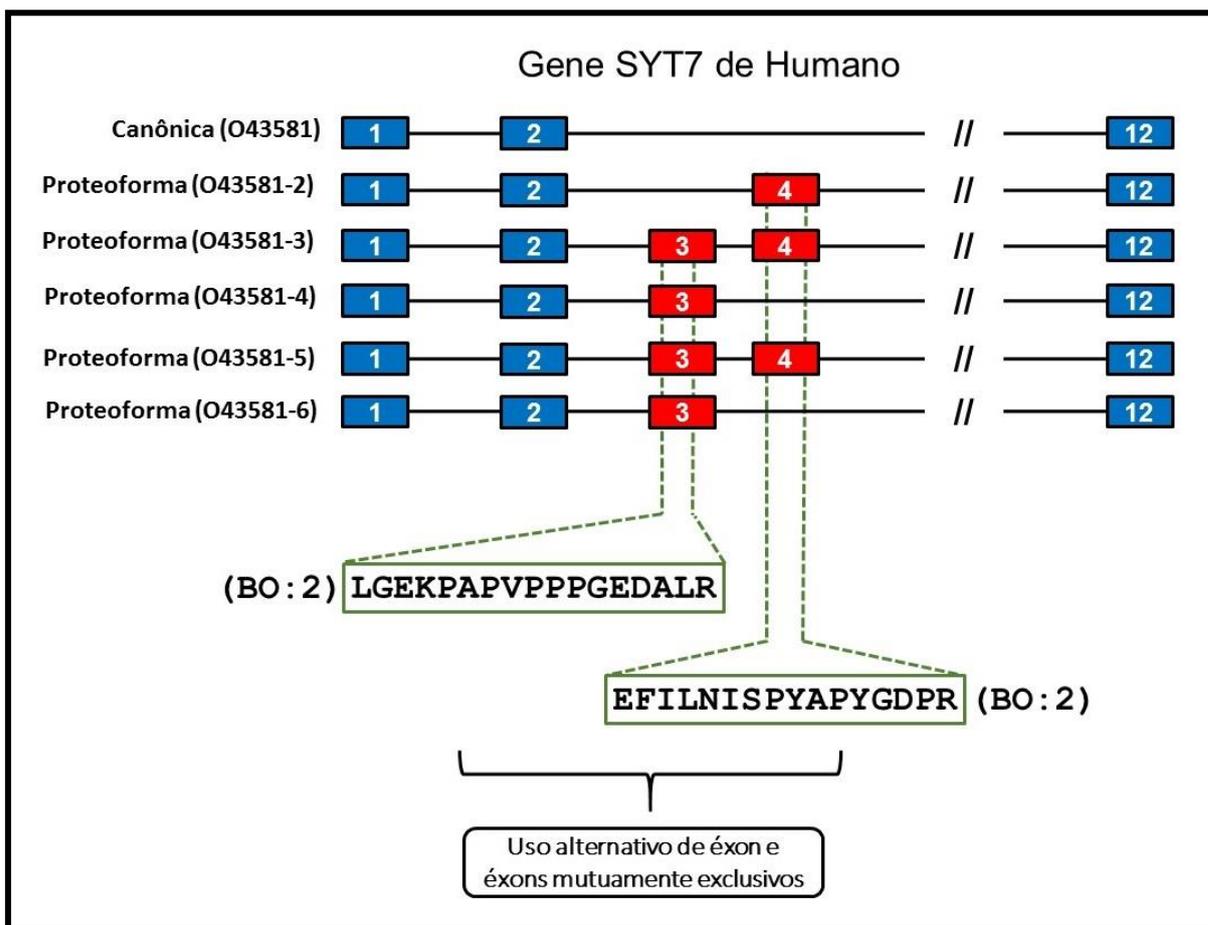


Figura 4.13.1 – Esquema representativo das diferenças no padrão de *splicing* da proteína canônica e proteoformas do gene SYT7 de humano. Retângulos vermelhos representam os éxons exclusivos de proteoformas onde os peptídeos identificados em nossas análises mapearam. Retângulos azuis representam os éxons das proteínas canônicas e proteoformas que não tiveram um peptídeo correspondente. Os quadros verdes representam os peptídeos identificados compartilhados entre proteoformas e as linhas pontilhadas verdes representam suas respectivas localizações nos éxons das proteoformas. Entre parênteses a contagem de espectros de cada peptídeo no bulbo olfatório (BO).

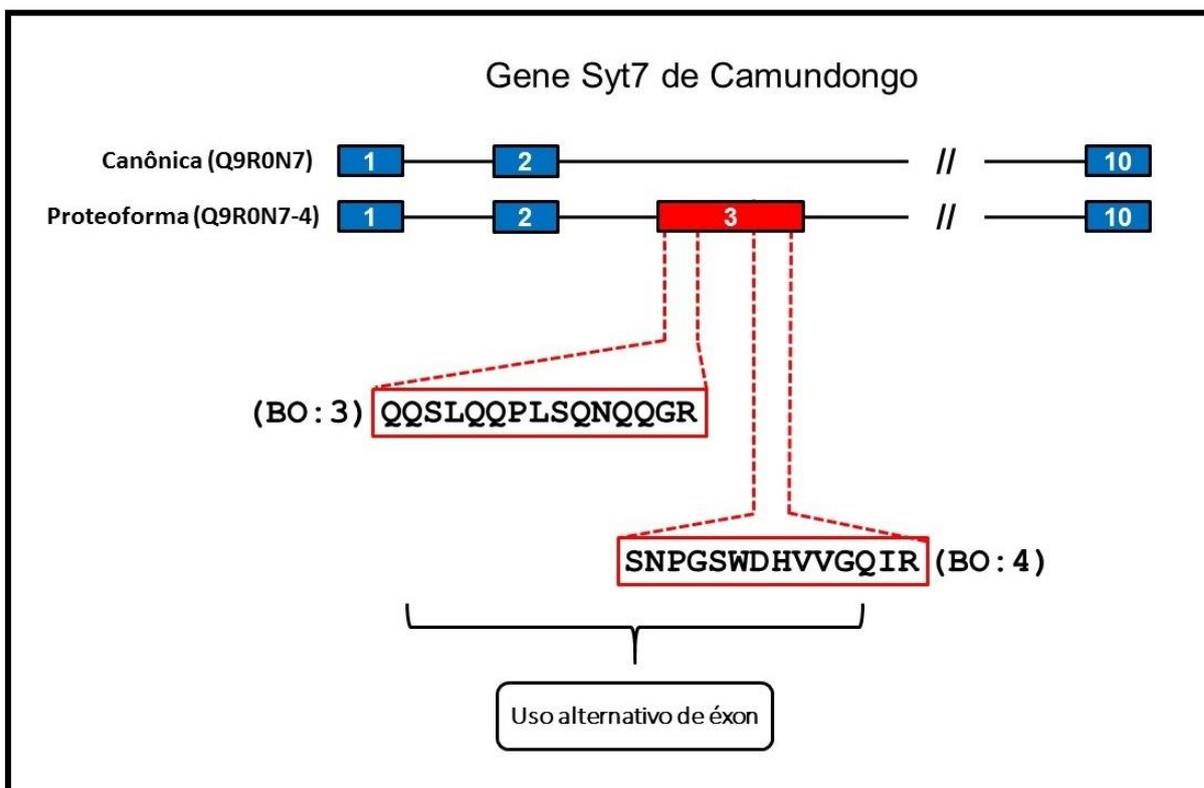


Figura 4.13.2 – Esquema representativo das diferenças no padrão de *splicing* da proteína canônica e proteoforma de *splicing* do gene Syt7 de camundongo. Retângulos vermelhos representam os éxons exclusivos de proteoformas onde os peptídeos identificados em nossas análises mapearam. Retângulos azuis representam os éxons das proteínas canônicas e proteoformas que não tiveram um peptídeo correspondente. Os quadros e linhas pontilhadas vermelhas representam os peptídeos proteotípicos identificados e suas respectivas localizações nos éxons das proteoformas. Entre parênteses a contagem de espectros de cada peptídeo no bulbo olfatório (BO).

Para o gene MAPT humano, identificamos dois peptídeos (“AEEAGIGDTPSLEDEAAGHVTQAR” e “AEEAGIGDTPSLEDEAAGHVTQ”) correspondentes as proteoformas P10636-2/NP_058525 e P10636-6/NP_058518. O primeiro peptídeo mapeou em uma junção entre dois éxons. O segundo peptídeo corresponde a uma porção do primeiro peptídeo e mapeou inteiramente em um éxon. Ambos evidenciaram um evento de uso alternativo de éxon (Figura 4.14.1). Para o gene ortólogo de camundongo (Mapt) nós identificamos um peptídeo (AEEAGIGDTPNQEDQAAGHVTQAR) correspondente as proteoformas P10637-3; P10637-4 e P10637-5. Este peptídeo mapeou em uma junção entre dois éxons e evidenciou um evento de uso alternativo de éxon (Figura 4.14.2).

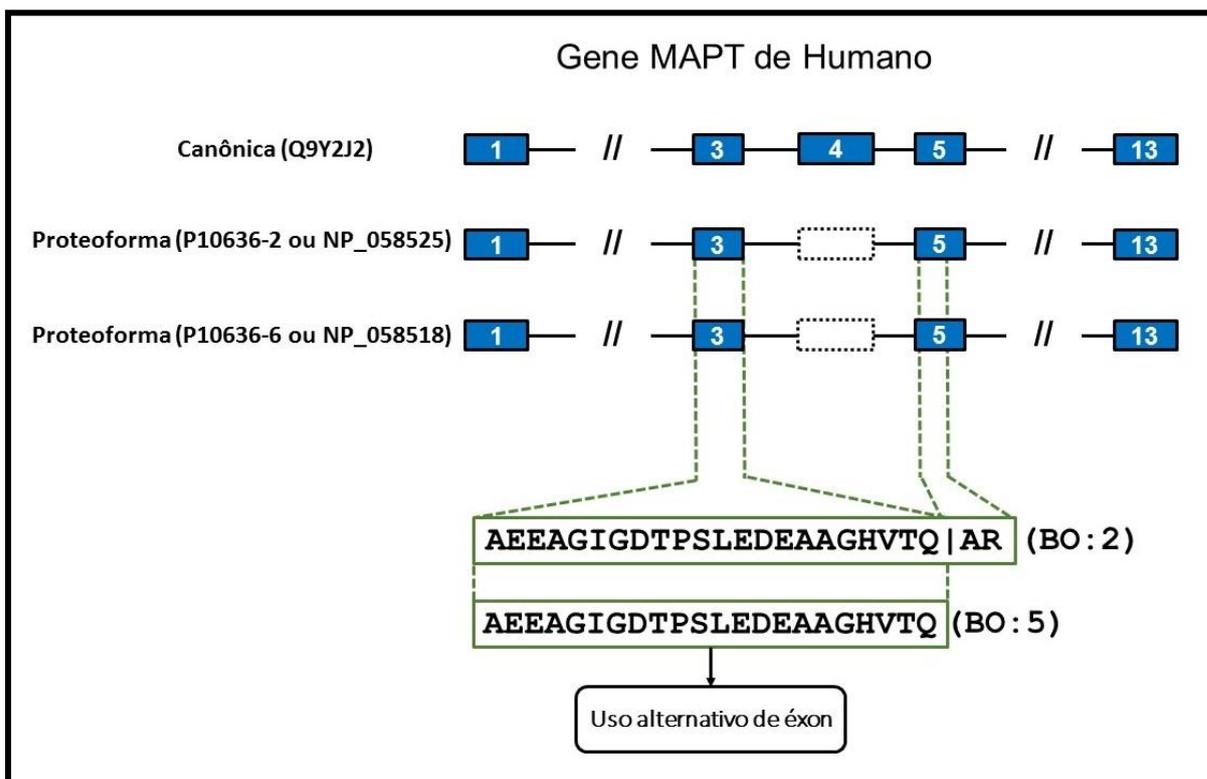


Figura 4.14.1 – Esquema representativo das diferenças no padrão de *splicing* da proteína canônica e proteoformas do gene MAPT de humano. Os éxons são representados pelos retângulos azuis. Os retângulos brancos com contorno pontilhado representam a ausência de um éxon. Os quadros verdes representam os peptídeos identificados compartilhados entre proteoformas e as linhas pontilhadas verdes representam suas respectivas localizações nos éxons das proteoformas. Entre parênteses a contagem de espectros de cada peptídeo no bulbo olfatório (BO).

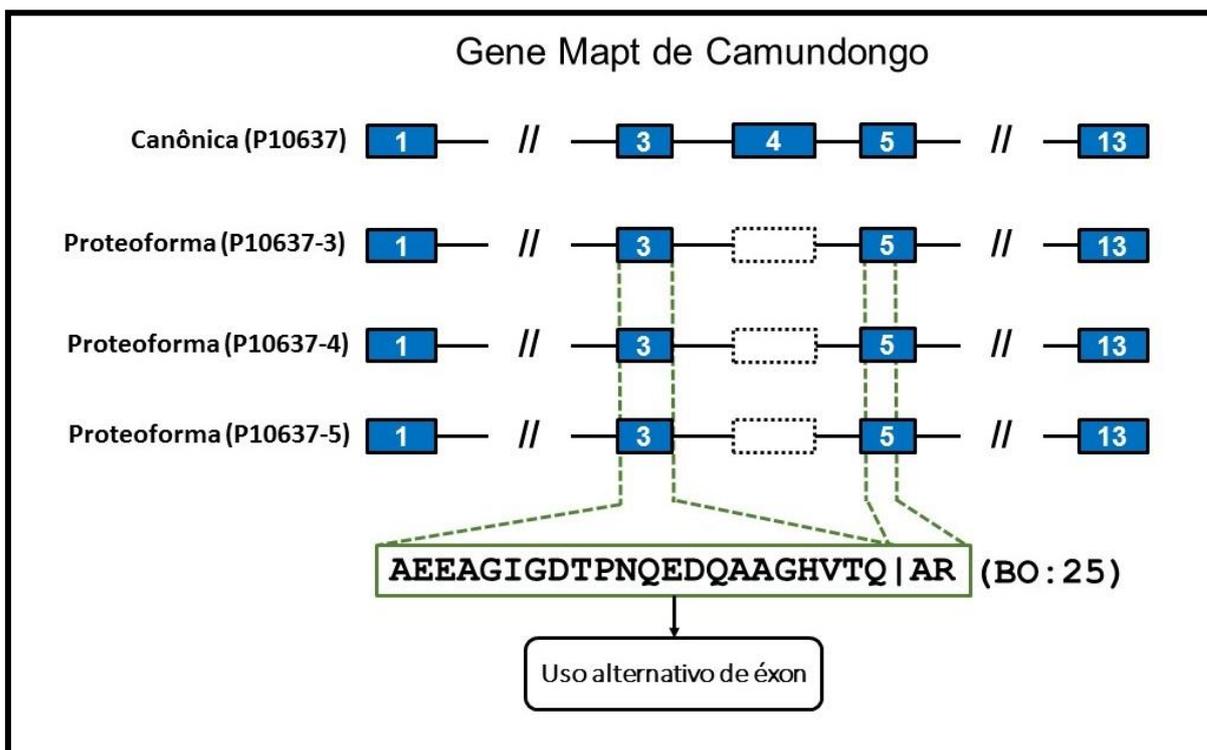


Figura 4.14.2 – Esquema representativo das diferenças no padrão de *splicing* da proteína canônica e proteoformas do gene Mapt de camundongo. Os éxons são representados pelos retângulos azuis. Os retângulos brancos com contorno pontilhado representam a ausência de um éxon. O quadro verde representa o peptídeo identificado compartilhado entre proteoformas e as linhas pontilhadas verdes representam sua respectiva localização nos éxons das proteoformas. Entre parênteses a contagem de espectros de cada peptídeo no bulbo olfatório (BO).

Para o gene RTN3 humano, identificamos um peptídeo (SCSSSCAVHDLIFWR) correspondente as proteoformas O95197-3/NP_006045; O95197-5/NP_958833 e O95197-6. Este peptídeo mapeou em uma junção entre dois éxons e demonstra um evento de uso alternativo de éxon (Figura 4.15.1). Para o gene ortólogo de camundongo (Rtn3) nós identificamos um peptídeo proteotípico (SCGSSCAVHDLIFWR) correspondente a proteoforma de *splicing* Q9ES97-3. O peptídeo mapeou em uma junção entre dois éxons evidenciando um evento de uso alternativo de dois éxons (Figura 4.15.2).

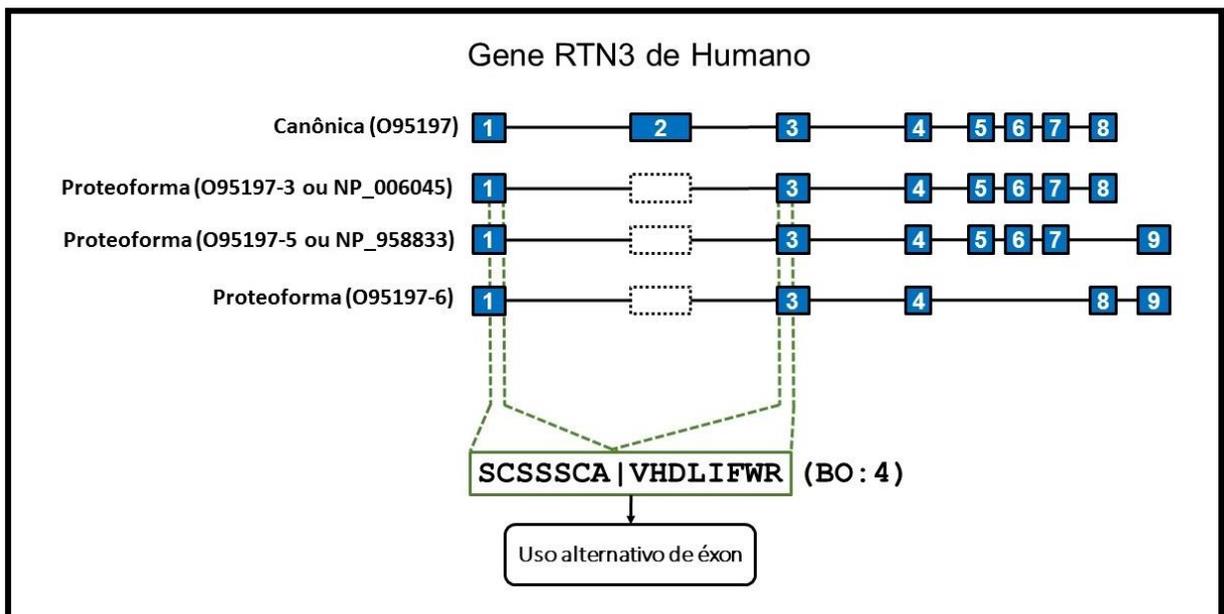


Figura 4.15.1 – Esquema representativo das diferenças no padrão de *splicing* da proteína canônica e proteoforma de *splicing* do gene RTN3 de humano. Os éxons são representados pelos retângulos azuis. Os retângulos brancos com contorno pontilhado representam a ausência de um éxon. O quadro verde representa o peptídeo identificado compartilhado entre proteoformas e as linhas pontilhadas verdes representa sua respectiva localização nos éxons das proteoformas. Entre parênteses a contagem de espectros de cada peptídeo no bulbo olfatório (BO).

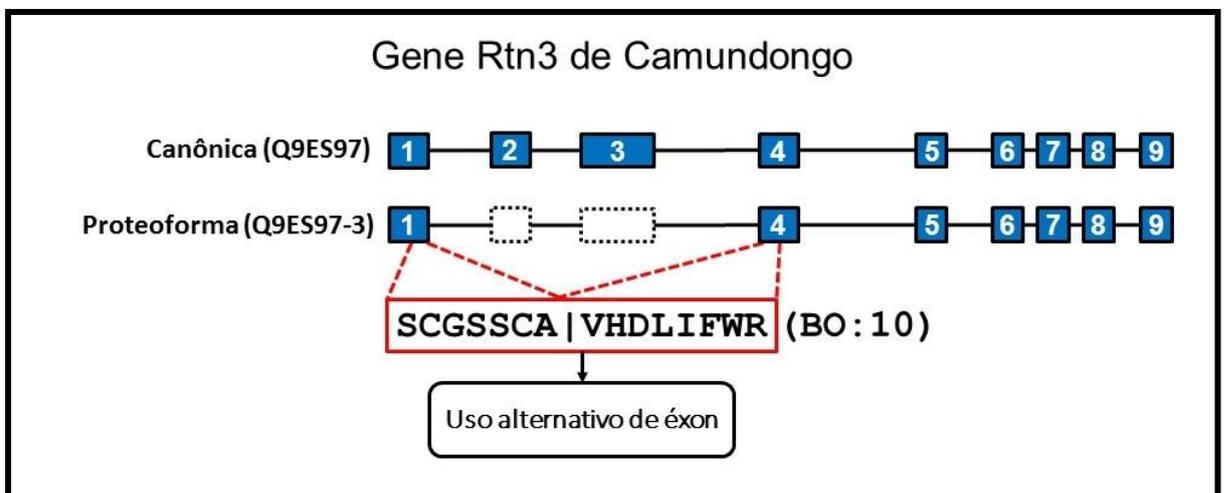


Figura 4.15.2 – Esquema representativo das diferenças no padrão de *splicing* da proteína canônica e proteoforma de *splicing* do gene Rtn3 de camundongo. Os éxons são representados pelos retângulos azuis. Os retângulos brancos com contorno pontilhado representam a ausência de dois éxons. O quadro e linhas pontilhadas vermelhas representam o peptídeo proteotípico identificado e sua respectiva localização nos éxons. Entre parênteses a contagem de espectros de cada peptídeo no bulbo olfatório (BO).

Para o gene CASK humano, identificamos dois peptídeos (“TQSSCEDLPSTTQPK” e “HNADLVTYEEVVK”) correspondentes as proteoformas O14936-2/NP_003679; O14936-3/NP_001119527; O14936-4/NP_001119526/NP_001119526; O14936-5 e O14936-6. Estes peptídeos mapearam em junções de éxons evidenciando eventos de uso alternativo de éxons e uso alternativo de sítio 5’ de *splicing* (Figura 4.16.1). Para o gene ortólogo de camundongo (Cask), identificamos um peptídeo proteotípico (TQSSCEDLPSTTQPK) correspondente a proteoforma de *splicing* O14936-2/NP_003679. Este peptídeo mapeou em uma junção entre dois éxons evidenciando um evento de uso alternativo de éxon (Figura 4.16.2).

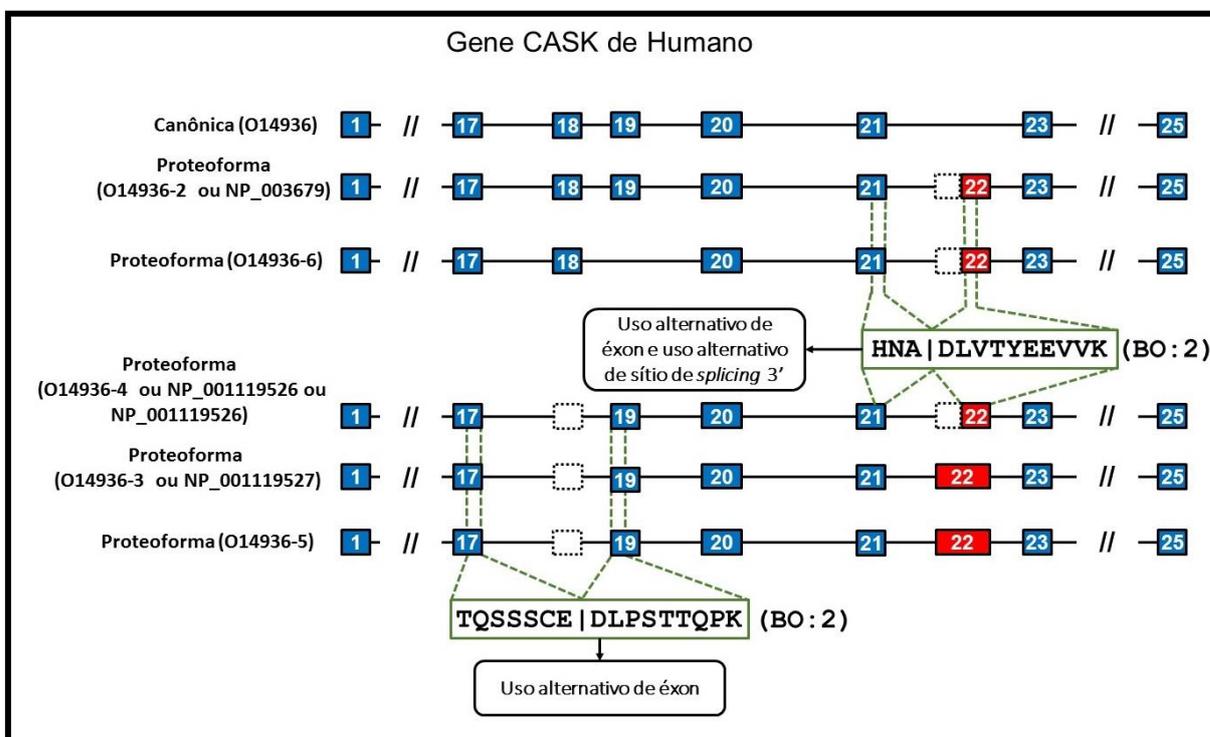


Figura 4.16.1 – Esquema representativo das diferenças no padrão de *splicing* da proteína canônica e proteoformas do gene CASK de humano. Retângulos vermelhos representam os éxons exclusivos de proteoformas onde os peptídeos identificados em nossas análises mapearam. Retângulos azuis representam os éxons das proteínas canônicas e proteoformas que não tiveram um peptídeo correspondente. Os retângulos brancos com contorno pontilhado representam a ausência de um éxon. Os quadros verdes representam os peptídeos identificados compartilhados entre proteoformas e as linhas pontilhadas verdes representam suas respectivas localizações nos éxons das proteoformas. Entre parênteses a contagem de espectros de cada peptídeo no bulbo olfatório (BO).

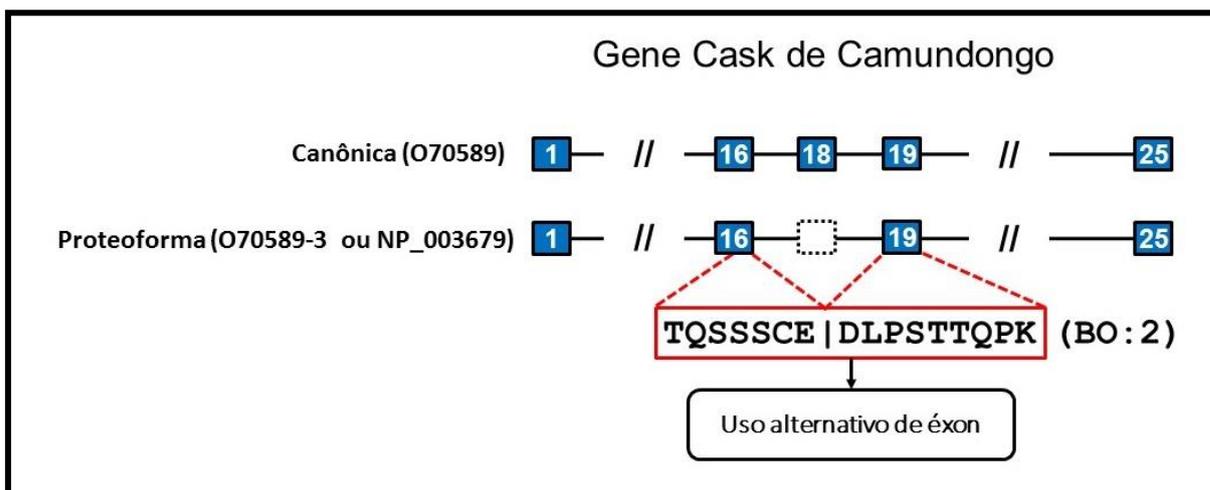


Figura 4.16.2 – Esquema representativo das diferenças no padrão de *splicing* da proteína canônica e proteoforma de *splicing* do gene Cask de camundongo. Os éxons são representados pelos retângulos azuis. O retângulo branco com contorno pontilhado representa a ausência de um éxon. O quadro e linhas pontilhadas vermelhas representam o peptídeo proteotípico e sua respectiva localização nos éxons. Entre parênteses a contagem de espectros de cada peptídeo no bulbo olfatório (BO).

Tanto para o gene PRKCB humano quanto para o seu ortólogo em camundongo (Prkcb), identificamos 1 peptídeo proteotípico (NIDQSEFEGFSFVNSEFLKPEVK) correspondente as proteoformas P05771-2/NP_002729 de humano e P68404-2/NP_NP_032881 de camundongo. Tanto em humano quanto em camundongo este peptídeo mapeou em uma porção de um éxon de uma proteoforma, correspondente a um íntron na proteína canônica, evidenciando um evento de uso alternativo de sítio de *splicing* 3' (Figura 4.17).

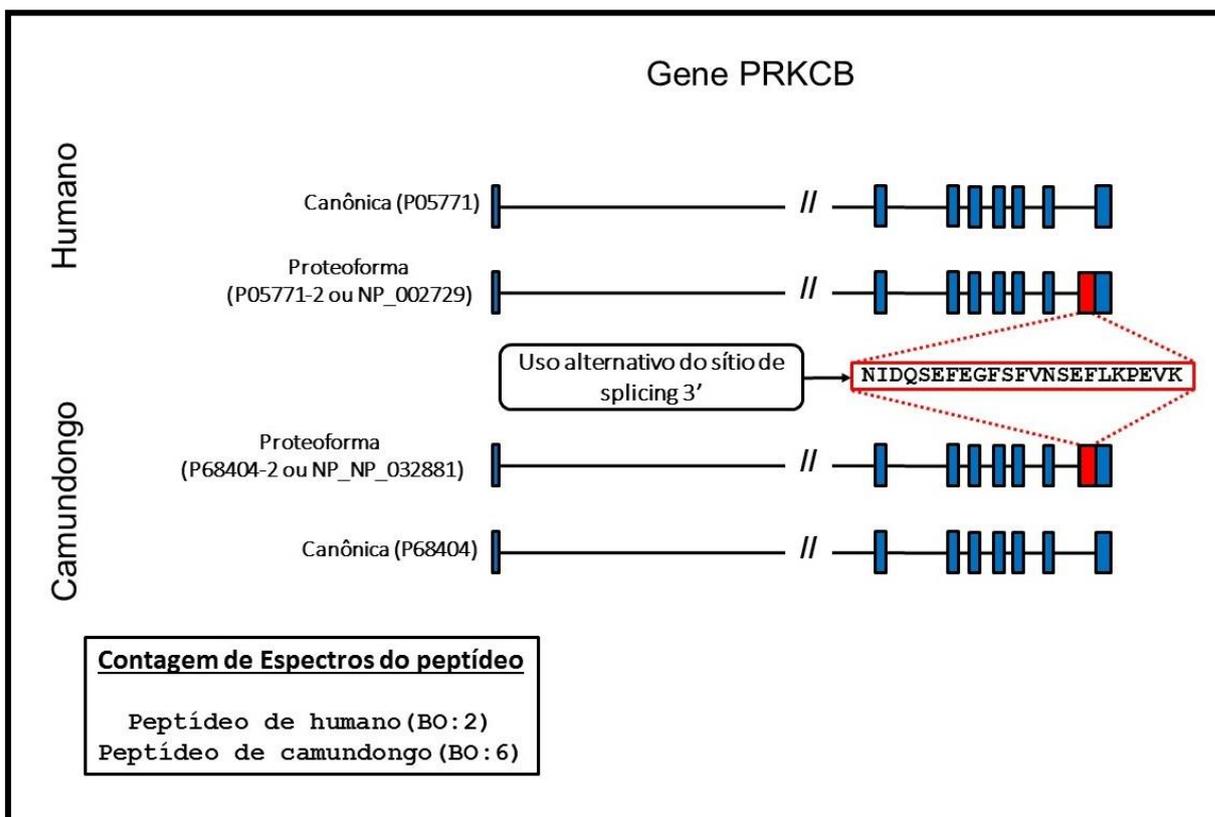


Figura 4.17 – Esquema representativo das diferenças no padrão de *splicing* da proteína canônica e proteoformas do gene PRKCB de humano e camundongo. Retângulos vermelhos representam os éxons exclusivos de proteoformas onde os peptídeos identificados em nossas análises mapearam. Retângulos azuis representam os éxons das proteínas canônicas e proteoformas que não tiveram um peptídeo correspondente. O quadro e linhas pontilhadas vermelhas representam o peptídeo proteotípico identificado e sua respectiva localização no éxon. Entre parênteses a contagem de espectros de cada peptídeo no bulbo olfatório (BO).

Para o gene STXBP1 humano e seu ortólogo em camundongo (*Stxbp1*), identificamos um peptídeo proteotípico (WEVLIGSTHILTPTK) correspondente as proteoformas P05771-2/NP_002729 de humano e O08599-2/NP_001107041 de camundongo. Tanto para humano como para camundongo este peptídeo mapeou em uma junção entre dois éxons, evidenciando um evento de uso alternativo de éxon (Figura 4.18).

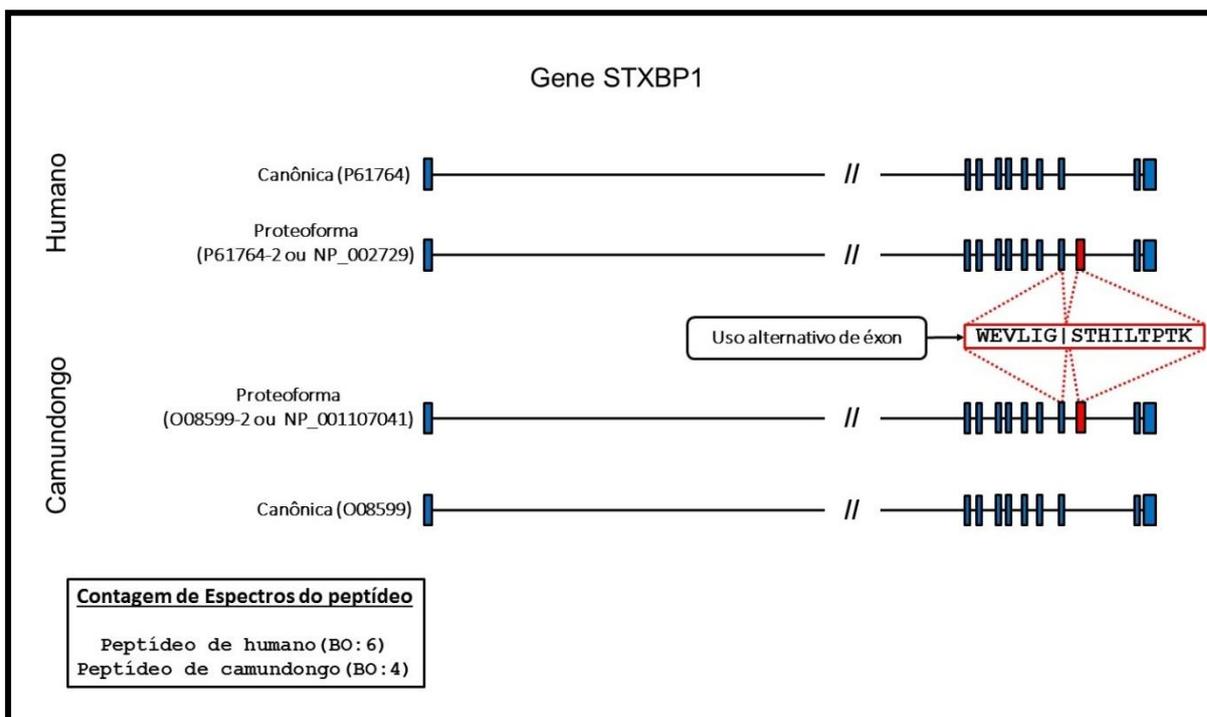


Figura 4.18 – Esquema representativo das diferenças no padrão de *splicing* da proteína canônica e proteoformas do gene STXBP1 de humano e camundongo. Retângulos vermelhos representam os éxons exclusivos de proteoformas onde os peptídeos identificados em nossas análises mapearam. Retângulos azuis representam os éxons das proteínas canônicas e proteoformas que não tiveram um peptídeo correspondente. O quadro e linhas pontilhadas vermelhas representam o peptídeo proteotípico identificado e sua respectiva localização nos éxons. Entre parênteses a contagem de espectros de cada peptídeo no bulbo olfatório (BO).

Para o gene CADM1 humano e seu respectivo ortólogo em camundongo (Cadm1), identificamos um peptídeo proteotípico (AHSDYMLYVYDSR) correspondente as proteoformas Q9BY67-5 de humano e Q8R5M8-4/NP_001020771 de camundongo. Tanto em humano quanto em camundongo este peptídeo mapeou em uma junção entre dois éxons, evidenciando um evento de uso alternativo de éxon (Figura 4.19).

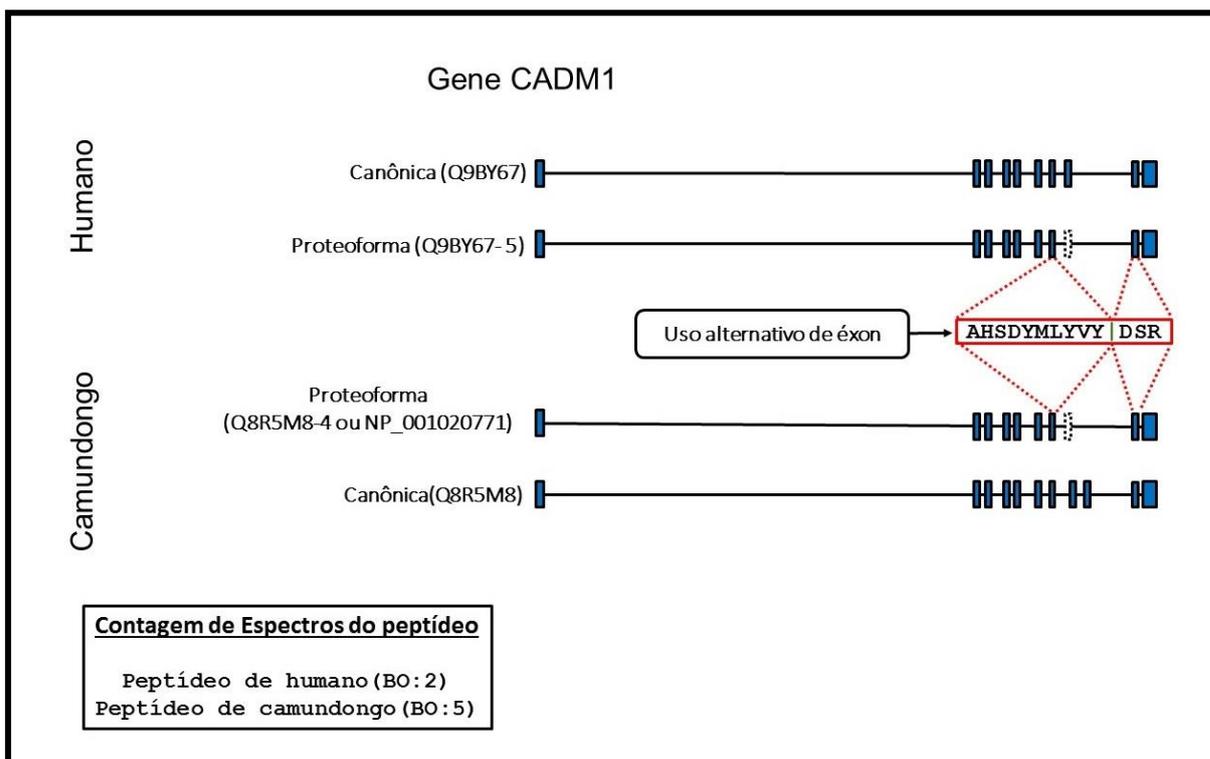


Figura 4.19 – Esquema representativo das diferenças no padrão de *splicing* da proteína canônica e proteoformas do gene CADM1 de humano e camundongo. Os éxons são representados pelos retângulos azuis. Os retângulos brancos com contorno pontilhado representam a ausência de um éxon. O quadro e linhas pontilhadas vermelhas representam o peptídeo proteotípico identificado em um ou mais experimentos de MS e sua respectiva localização nos éxons. Entre parênteses a contagem de espectros de cada peptídeo no bulbo olfatório (BO).

4.3.1.3 Proteoformas comuns ao nervo óptico e bulbo olfatório

Nos dados de MS de nervo óptico e bulbo olfatório, identificamos proteoformas de quatro genes ortólogos entre humano e camundongo (EPB41L3, NFASC, MAP4 e DNM1) comuns às estas regiões do cérebro.

Para o gene NFASC humano, identificamos dois peptídeos (“GPEPESVIGYSGEDLPSAPR” e “IGYSGEDLPSAPR”) compartilhados com as proteoformas O94856-12; O94856-9 ou NP_001005388 e O94856-4. Em camundongo, para o gene ortólogo (Nfasc), identificamos três peptídeos compartilhados com as proteoformas NP_001153790; NP_001153789; NP_001153788. O mapeamento dos peptídeos ao genoma evidenciou um evento de uso alternativo de éxons em humano (Figura 4.20.1) e camundongo (Figura 4.20.2).

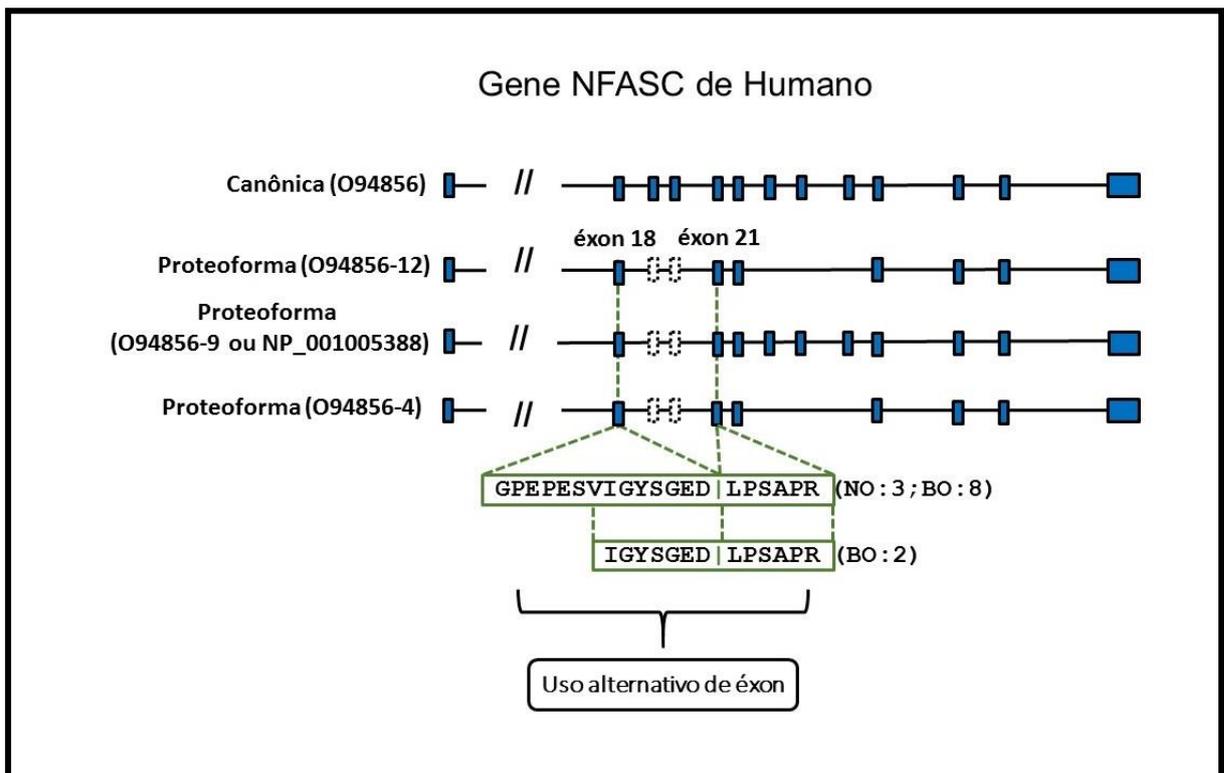


Figura 4.20.1 – Esquema representativo das diferenças no padrão de *splicing* da proteína canônica e proteoformas do gene NFASC de humano. Os éxons são representados pelos retângulos azuis. Os retângulos brancos com contorno pontilhado representam a ausência de dois éxons. Os quadros verdes representam os peptídeos identificados compartilhados entre proteoformas e as linhas pontilhadas verdes representam suas respectivas localizações nos éxons das proteoformas. Entre parênteses a contagem de espectros de cada peptídeo no nervo óptico (NO) e bulbo olfatório (BO).

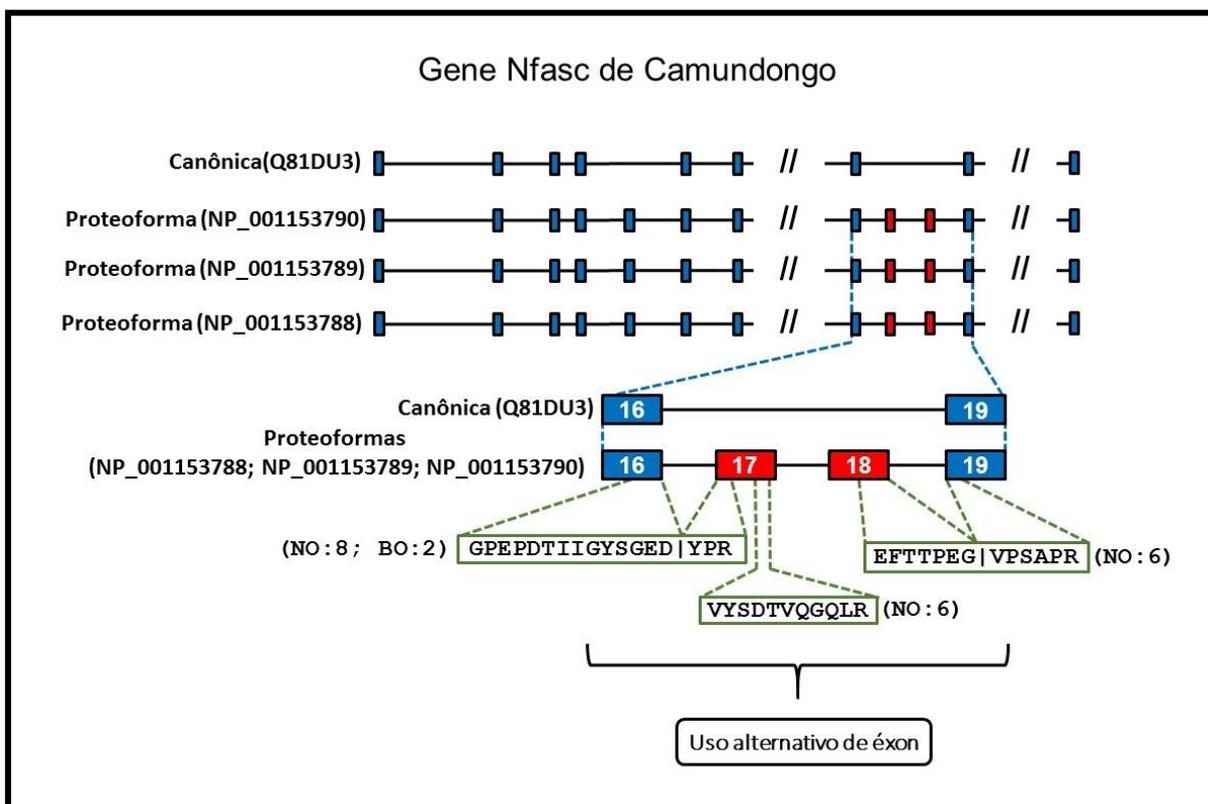


Figura 4.20.2 – Esquema representativo das diferenças no padrão de *splicing* da proteína canônica e proteoformas do gene Nfasc de camundongo. Retângulos vermelhos representam os éxons exclusivos de proteoformas onde os peptídeos identificados em nossas análises mapearam. Retângulos azuis representam os éxons das proteínas canônicas e proteoformas que não tiveram um peptídeo correspondente. Os quadros verdes representam os peptídeos identificados compartilhados entre proteoformas e as linhas pontilhadas verdes representam suas respectivas localizações nos éxons das proteoformas. As linhas pontilhadas azuis representam a ampliação dos éxons. Entre parênteses a contagem de espectros de cada peptídeo no nervo óptico (NO) e bulbo olfatório (BO).

Nós identificamos três peptídeos proteotípicos (“EIPPFICEEPVAK”; “LSSTEEAVLNQAPQQK” e “SQSDLDFPGAADIPTTR”) do gene MAP4 humano correspondentes a proteoforma P27816-3. Em camundongo, para o gene ortólogo (Map4), identificamos treze peptídeos proteotípicos (“LNEIGLNVSMGQLK”; “SSEEGPMLDQVPQQK”; “ELGIAGLYLHSK”; “EIGQDGFQWQR”; “ALTAALAAEDEQLSK”; “NGLVSSCSVTGVMSDNSGQLNNK”; “AAEFHPGLTEGK”; “EIEVTATQSIPSLLEETPR”; “QPAALTAALAAEDEQLSK”; “NSSFLEQNK”; “VDSCPFICLGGEK”; “ALSECYHLSVPPAVNLVDK”; “LEQIPEGSHK”) correspondentes a proteoforma P27546-4/NP_001192260. O mapeamento dos peptídeos ao genoma de humano (Figura 4.21.1) e camundongo (Figura 4.21.2) mostrou o evento de uso alternativo de um éxon.

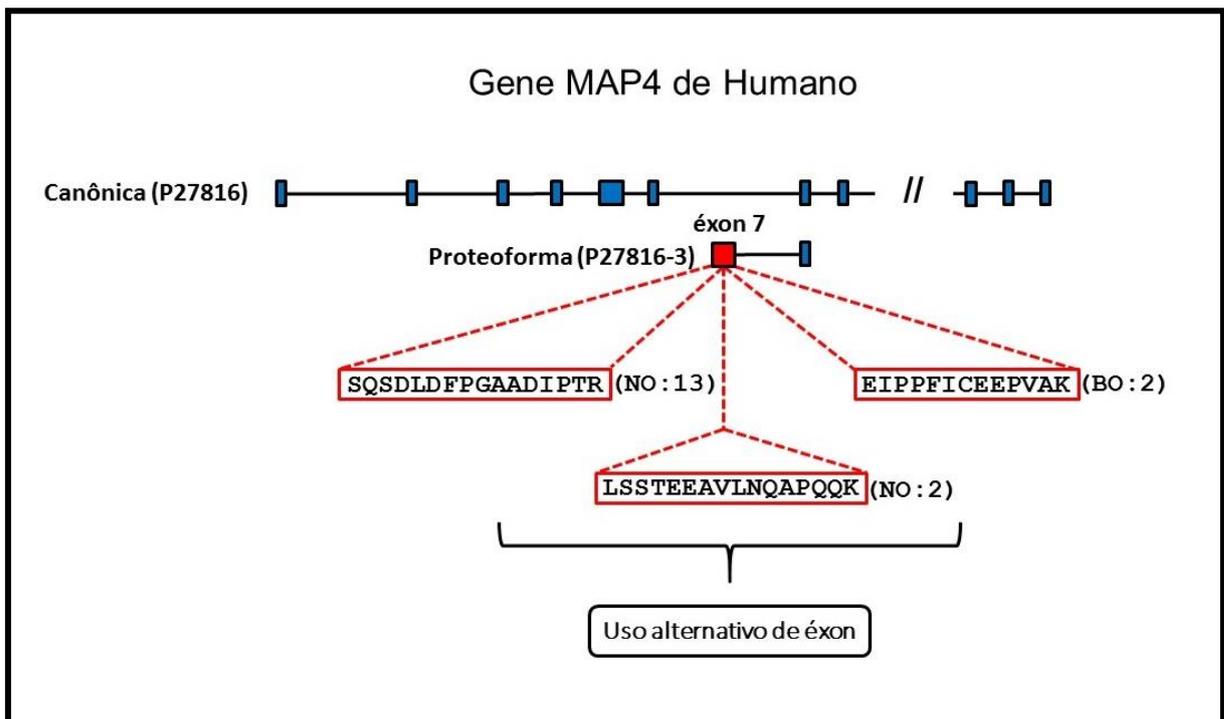


Figura 4.21.1 – Esquema representativo das diferenças no padrão de *splicing* da proteína canônica e proteoforma de *splicing* do gene MAP4 de humano. Retângulos vermelhos representam os éxons exclusivos de proteoformas onde os peptídeos identificados em nossas análises mapearam. Retângulos azuis representam os éxons das proteínas canônicas e proteoformas que não tiveram um peptídeo correspondente. O quadro e linhas pontilhadas vermelhas representam os peptídeos proteotípicos identificados e suas respectivas localizações nos éxons. Entre parênteses a contagem de espectros de cada peptídeo no nervo óptico (NO) e bulbo olfatório (BO).

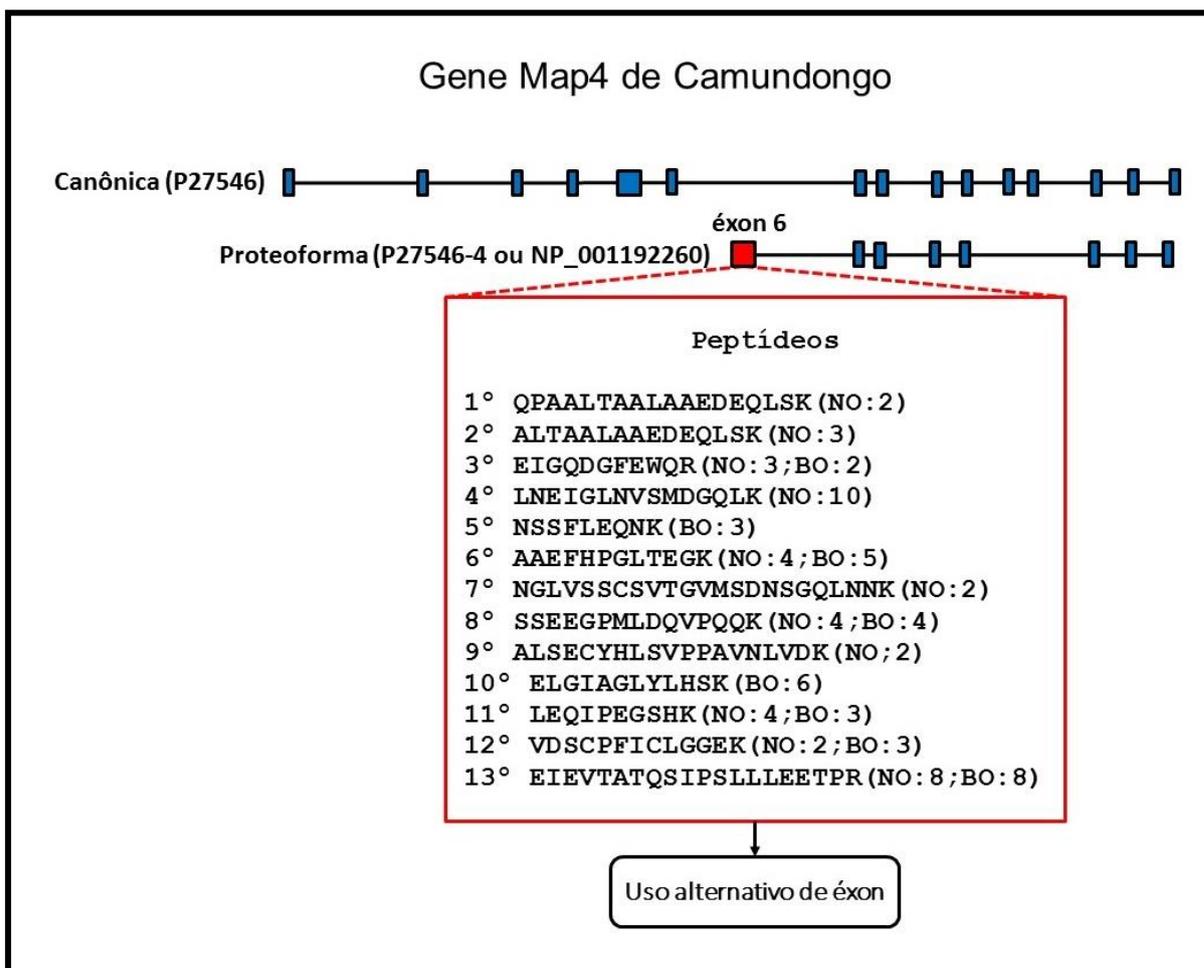


Figura 4.21.2 – Esquema representativo das diferenças no padrão de *splicing* da proteína canônica e proteoforma de *splicing* do gene Map4 de camundongo. Retângulos vermelhos representam os éxons exclusivos de proteoformas onde os peptídeos identificados em nossas análises mapearam. Retângulos azuis representam os éxons das proteínas canônicas e proteoformas que não tiveram um peptídeo correspondente. O quadro e linhas pontilhadas vermelhas representam os peptídeos proteotípicos identificados e suas respectivas localizações nos éxons. Entre parênteses a contagem de espectros de cada peptídeo no nervo óptico (NO) e bulbo olfatório (BO).

Para o gene EPB41L3 humano, identificamos um peptídeo (DSMSAAEVGTGQYATTK) correspondente a 3 proteoformas (Q9Y2J2-2; Q9Y2J2-3; Q9Y2J2-4). Este peptídeo mapeou em uma porção de um éxon das proteoformas que corresponde a um íntron na proteína canônica, indicando um evento de uso alternativo de sítio de *splicing* 3' (Figura 4.22.1). Para o gene ortólogo de camundongo (Epb41l3), identificamos um peptídeo (SLDGEVGTGQYATTK) correspondente a 2 proteoformas (Q9WV92-7 e Q9WV92-8). Este peptídeo mapeou em uma junção entre dois éxons das proteoformas, evidenciando um evento de uso alternativo de sítio de *splicing* 3' (Figura 4.22.2).

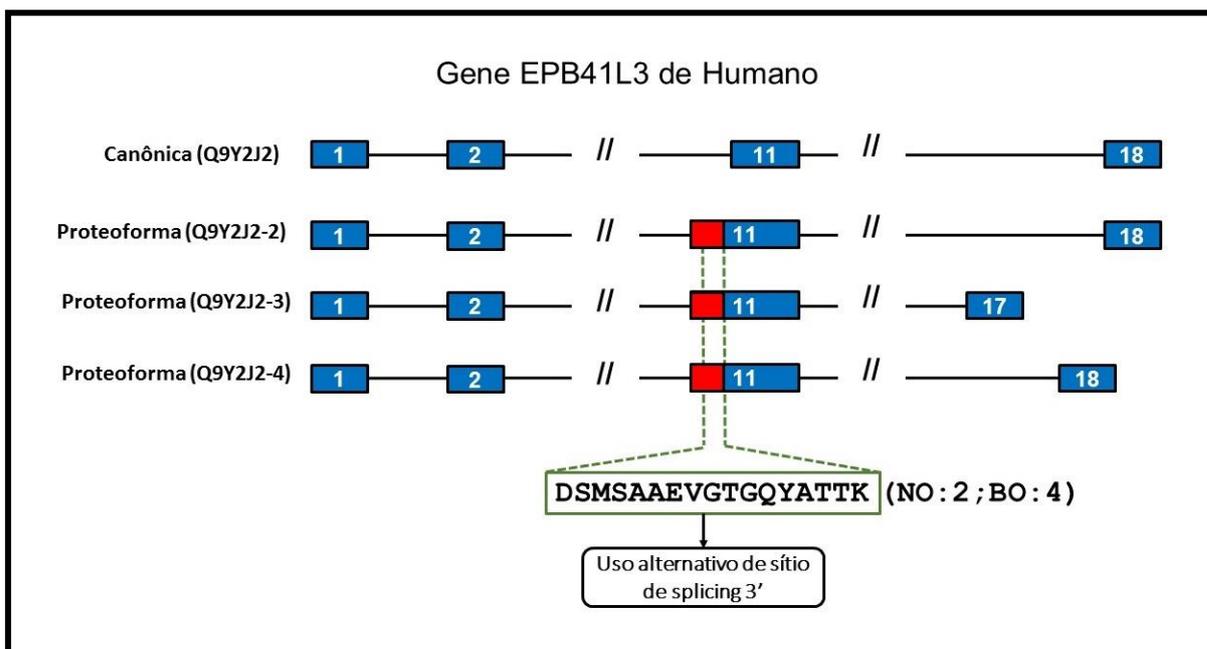


Figura 4.22.1 – Esquema representativo das diferenças no padrão de *splicing* da proteína canônica e proteoformas do gene EPB41L3 de humano. O Retângulo vermelho representa parte do éxon exclusiva das proteoformas onde o peptídeo identificado em nossas análises mapeou. Retângulos azuis representam os éxons das proteínas canônicas e proteoformas que não tiveram um peptídeo correspondente. O quadro verde representa o peptídeo identificado compartilhado entre proteoformas e as linhas pontilhadas verdes representam sua respectiva localização nos éxons das proteoformas. Entre parênteses a contagem de espectros de cada peptídeo no nervo óptico (NO) e bulbo olfatório (BO).

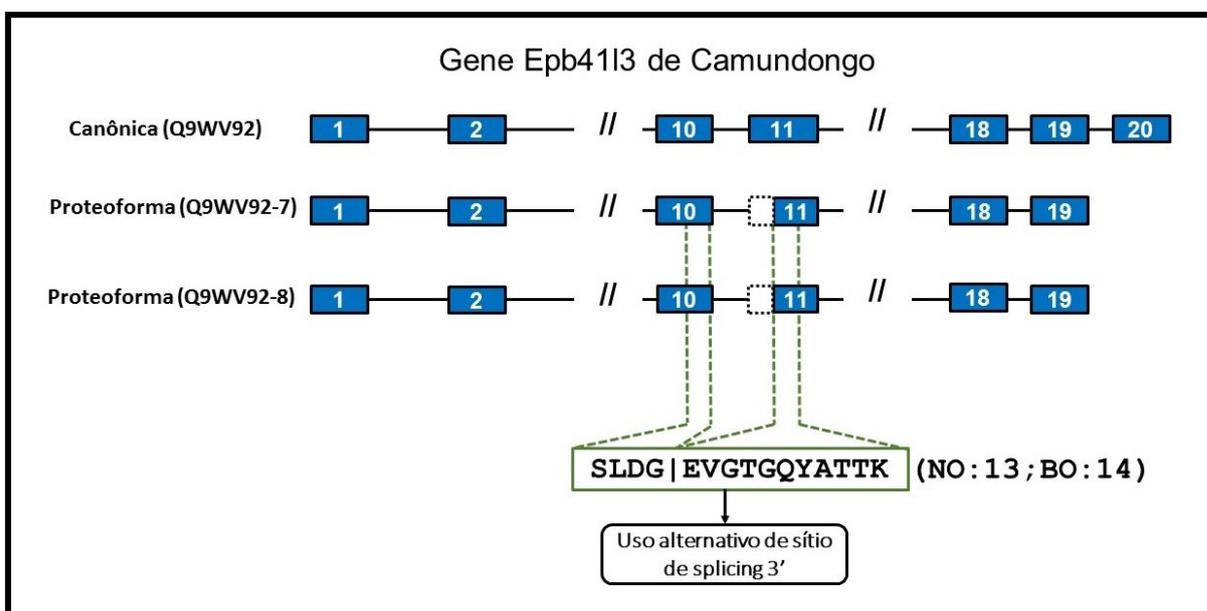


Figura 4.22.2 – Esquema representativo das diferenças no padrão de *splicing* da proteína canônica e proteoformas do gene Pb41l3 de camundongo. Os éxons são representados pelos retângulos azuis. Os retângulos brancos com contorno pontilhado representam a ausência de parte de um éxon. O quadro verde representa o peptídeo identificado compartilhado entre proteoformas e as linhas

pontilhadas verdes representam sua respectiva localização nos éxons das proteoformas. Entre parênteses a contagem de espectros de cada peptídeo no nervo óptico (NO) e bulbo olfatório (BO).

Para o gene DMN1 humano e seu ortólogo em camundongo (*Dmn1*), identificamos o peptídeo “TGLFTPDLAFEATVK” correspondente as proteoformas Q05193-2 e Q05193-5 de humano e P39053-3 de camundongo. Este peptídeo mapeou inteiramente em um éxon, tanto em humano quanto em camundongo, evidenciando um evento de uso alternativo de éxon (Figura 4.23).

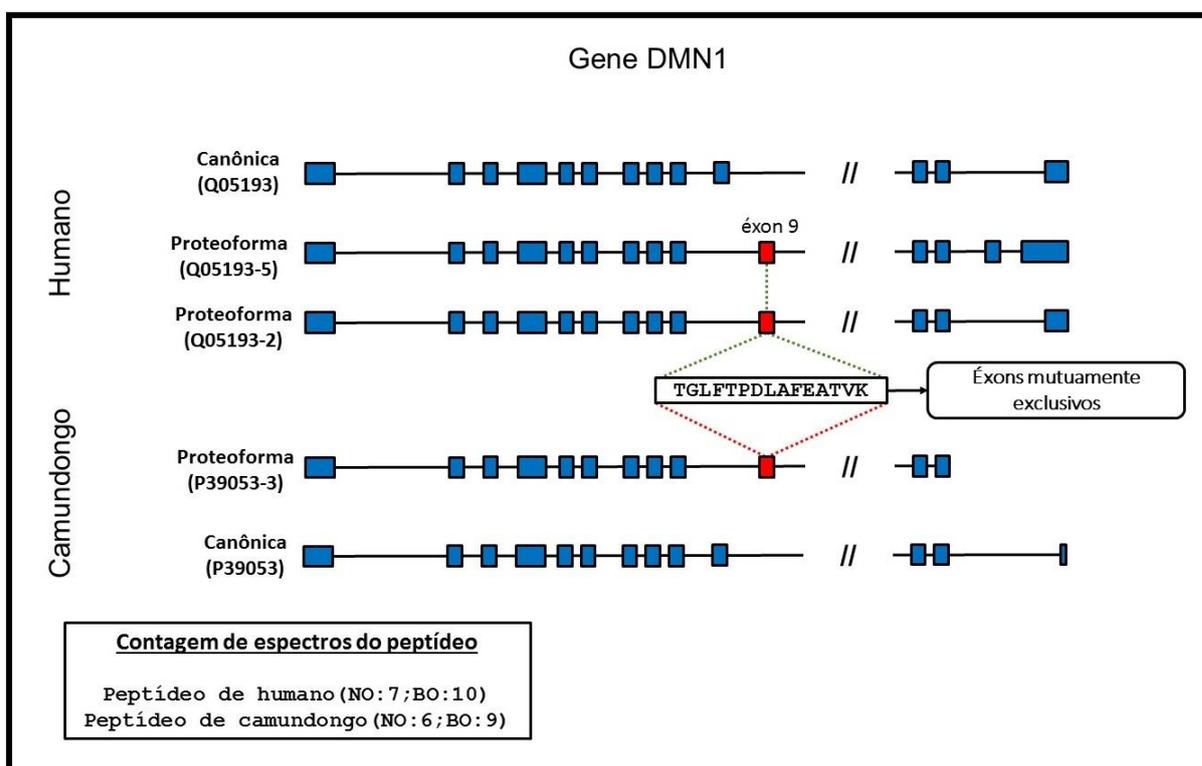


Figura 4.23 – Esquema representativo das diferenças no padrão de *splicing* da proteína canônica e proteoformas do gene DMN1 humano e camundongo. Retângulos vermelhos representam os éxons exclusivos de proteoformas onde os peptídeos identificados em nossas análises mapearam. Retângulos azuis representam os éxons das proteínas canônicas e proteoformas que não tiveram um peptídeo correspondente. As linhas pontilhadas verdes representam o peptídeo compartilhado entre proteoformas, identificado em um ou mais experimentos de MS e suas respectivas localizações no éxon das proteoformas de humano. As linhas pontilhadas vermelhas representam o peptídeo proteotípico identificado em um ou mais experimentos de MS e sua respectiva localização no éxon da proteoforma de camundongo. Entre parênteses a contagem de espectros de cada peptídeo no nervo óptico (NO) e bulbo olfatório (BO).

4.3.1.4 Proteoformas comuns ao nervo óptico e corpo caloso

Nós inferimos proteoformas de dois genes ortólogos entre humano e camundongo (SPTAN1 e PLEKHB1) nos dados de MS dos dois organismos, compartilhadas entre nervo óptico e corpo caloso.

Para o gene SPTAN1 humano, identificamos um peptídeo proteotípico (QEIQDNQYHSLLELGEK) correspondente a proteoforma Q13813-3/NP_001182461. Este peptídeo mapeou em uma junção entre dois éxons e representam um evento de uso alternativo de éxon (Figura 4.24.1). Para o gene ortólogo em camundongo (Sptan1), identificamos um peptídeo (QEIQDNQYQSLLELGEK) compartilhado com duas proteoformas (NP_001070022; NP_001171139). Este peptídeo também mapeou em uma junção entre dois éxons e evidência um evento de uso alternativo de éxon (Figura 4.24.2).

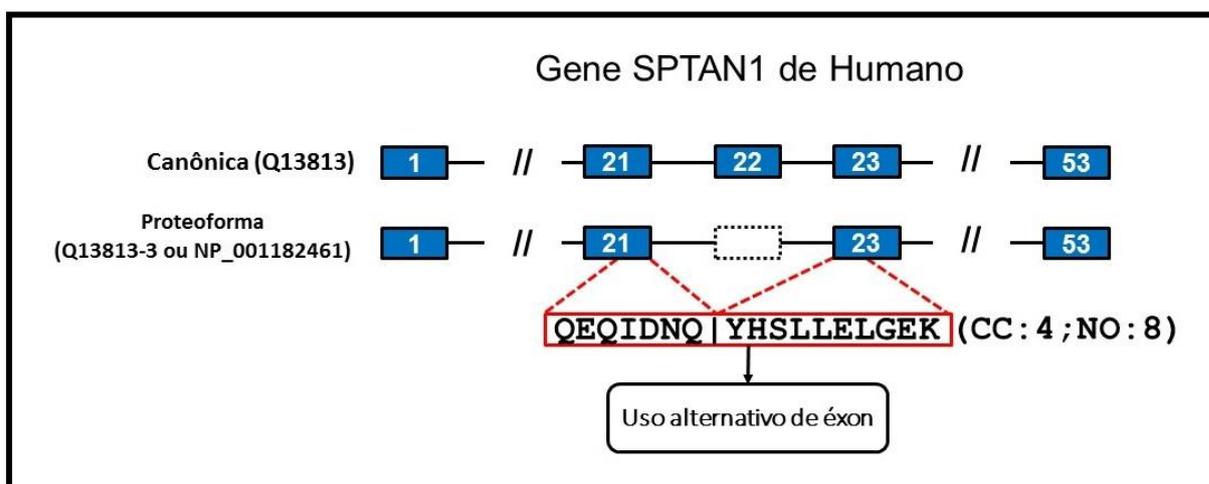


Figura 4.24.1 – Esquema representativo das diferenças no padrão de *splicing* da proteína canônica e proteoforma de *splicing* do gene SPTAN1 de humano. Os éxons são representados pelos retângulos azuis. O retângulo branco com contorno pontilhado representa a ausência de um éxon. O quadro e linhas pontilhadas vermelhas representam o peptídeo proteotípico identificado em um ou mais experimentos de MS e sua respectiva localização nos éxons. Entre parênteses a contagem de espectros de cada peptídeo no corpo caloso (CC) e nervo óptico (NO).

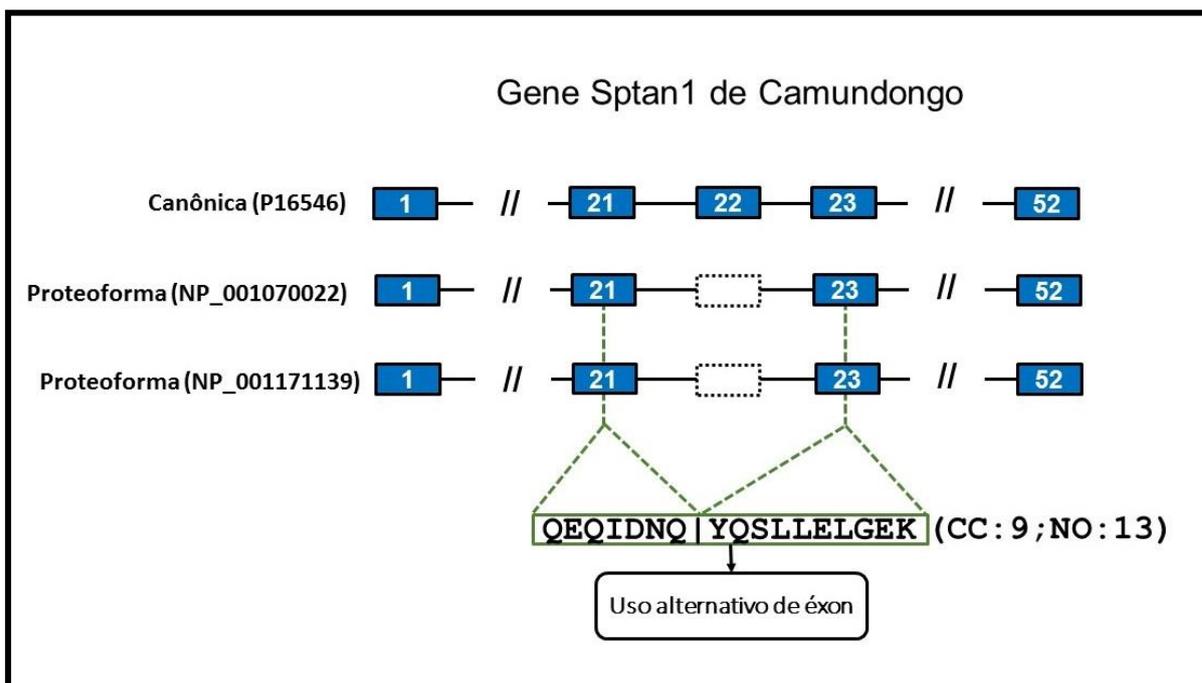


Figura 4.24.2 – Esquema representativo das diferenças no padrão de *splicing* da proteína canônica e proteoformas do gene Sptan1 de camundongo. Os éxons são representados pelos retângulos azuis. Os retângulos brancos com contorno pontilhado representam a ausência de um éxon. O quadro verde representa o peptídeo identificado compartilhado entre proteoformas e as linhas pontilhadas verdes representam sua respectiva localização nos éxons das proteoformas. Entre parênteses a contagem de espectros de cada peptídeo no corpo caloso (CC) e nervo óptico (NO).

Para o gene PLEKHB1 humano, identificamos um peptídeo (TALLEANSTPVR) correspondente as proteoformas Q9UF11-2/NP_001123505 e Q9UF11-4/NP_001123507/NP_001123508. Este peptídeo mapeou em uma junção entre dois éxons e evidenciou um evento de uso alternativo de éxon (Figura 4.25.1). Para o gene ortólogo em camundongo (Plekhh1), identificamos um peptídeo (TALMEANSTPVR) correspondente as proteoformas Q9QYE9-2/NP_001156654 e Q9QYE9-4/NP_001156656/NP_001156657/NP_001156659. Este peptídeo também mapeou em uma junção entre dois éxons e evidenciou um evento de uso alternativo de éxon (Figura 4.25.2).

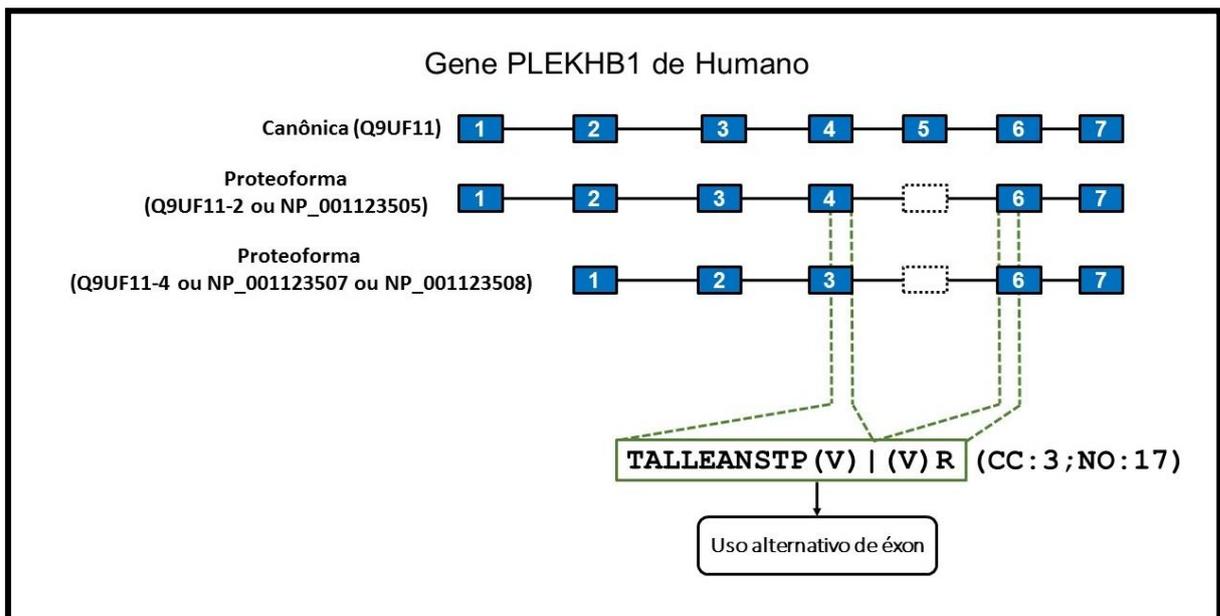


Figura 4.25.1 – Esquema representativo das diferenças no padrão de *splicing* da proteína canônica e proteoforma de *splicing* do gene PLEKHB1 de humano. Os éxons são representados pelos retângulos azuis. Os retângulos brancos com contorno pontilhado representam a ausência de um éxon. O quadro verde representa o peptídeo identificado compartilhado entre proteoformas e as linhas pontilhadas verdes representam sua respectiva localização nos éxons das proteoformas. Entre parênteses a contagem de espectros de cada peptídeo no corpo caloso (CC) e nervo óptico (NO).

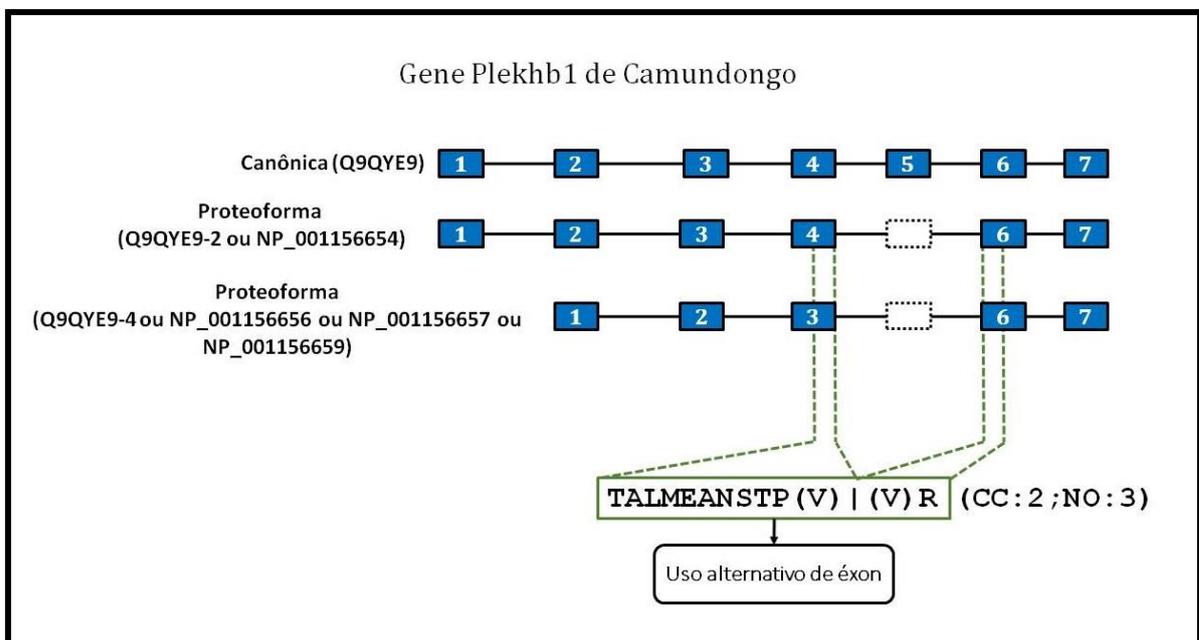


Figura 4.25.2 – Esquema representativo das diferenças no padrão de *splicing* da proteína canônica e proteoforma de *splicing* do gene Plekhhb1 de humano. Os éxons são representados pelos retângulos azuis. Os retângulos brancos com contorno pontilhado representam a ausência de um éxon. O quadro verde representa o peptídeo identificado compartilhado entre proteoformas e as linhas pontilhadas verdes representam sua respectiva localização nos éxons das proteoformas. Entre parênteses a contagem de espectros de cada peptídeo no corpo caloso (CC) e nervo óptico (NO).

4.3.1.5 Proteoformas comuns as três regiões do cérebro

Nós inferimos proteoformas de três genes ortólogos entre humano e camundongo (PKM, TPM1 e TPM3) em dados de MS de corpo caloso, nervo óptico e bulbo olfatório de ambos os organismos.

Para o gene PKM humano, identificamos cinco peptídeos proteotípicos (“EAEAAMFHR”; “LFEELVR”; “ASSHSTDLMEMAMGSVEASYK”; “CLAAALIVLTESGR” e “ALIVLTESGR”) da proteoforma de *splicing* P14618-2/NP_872270/NP_872271. Estes peptídeos mapearam inteiramente em um éxon e evidenciam um evento de uso alternativo de éxon. Para o gene ortólogo de camundongo, identificamos cinco peptídeos proteotípicos (“EAEAAMFHR”; “LLFEELVR”; “ASSHSTDLMEMAMGSVEASYK”; “MAMGSVEASYK” e “CLAAALIVLTESGR”) correspondentes a proteoforma P52480-2/NP_001240812. Estes peptídeos também mapearam inteiramente em um éxon e evidenciam um evento de uso alternativo de éxon (Figura 4.26).

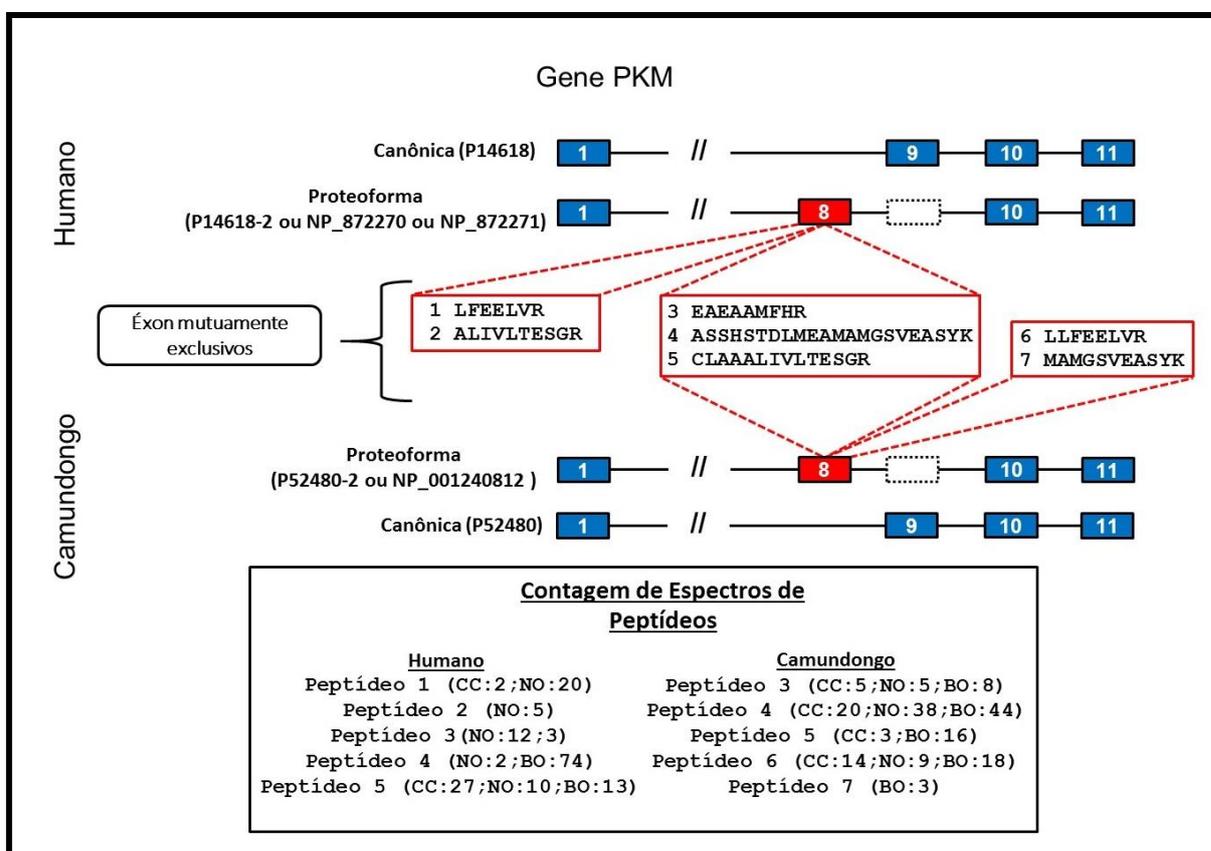


Figura 4.26 – Esquema representativo das diferenças no padrão de *splicing* da proteína canônica e proteoforma de *splicing* do gene PKM de humano e camundongo. Retângulos vermelhos representam os éxons exclusivos de proteoformas onde os peptídeos identificados em nossas análises mapearam.

Retângulos azuis representam os éxons das proteínas canônicas e proteoformas que não tiveram um peptídeo correspondente. Os retângulos brancos com contorno pontilhado representam a ausência de um éxon. Os quadros e linhas pontilhadas vermelhas representam os peptídeos proteotípicos identificados em um ou mais experimentos de MS e suas respectivas localizações no éxon da proteoforma. Entre parênteses a contagem de espectros de cada peptídeo no corpo caloso (CC) e nervo óptico (NO) e bulbo olfatório (BO).

Em humano, para o gene TPM1, identificamos dois peptídeos proteotípicos (“SLQEQADAAEER” e “SIDDLELEDQLYQQLEQNR”), correspondentes a proteoforma P09493-5/NP_001018008, um peptídeo (“ETAADVASLNR”) compartilhado entre as proteoformas P09493-2 e P09493-5/NP_001018008 e outro peptídeo (“AEDSLLAAEEAAAK”) compartilhado entre as proteoformas P09493-6; P09493-7/NP_001192260 e P09493-8/ NP_001192260. Em camundongo, para o gene ortólogo (Tpm1), identificamos um peptídeo proteotípico da proteoforma de *splicing* NP_001157721 e outro peptídeo compartilhado entre as proteoformas NP_001157727 e NP_001157728 (“SIDDLELEDQLYHQLEQNR”). O mapeamento dos peptídeos ao genoma de humano evidenciou eventos de uso alternativo de éxon e do sítio 3’ de *splicing* (Figura 4.27.1). Em camundongo os peptídeos mapeados ao genoma mostraram eventos de uso alternativo de éxon (Figura 4.27.2).

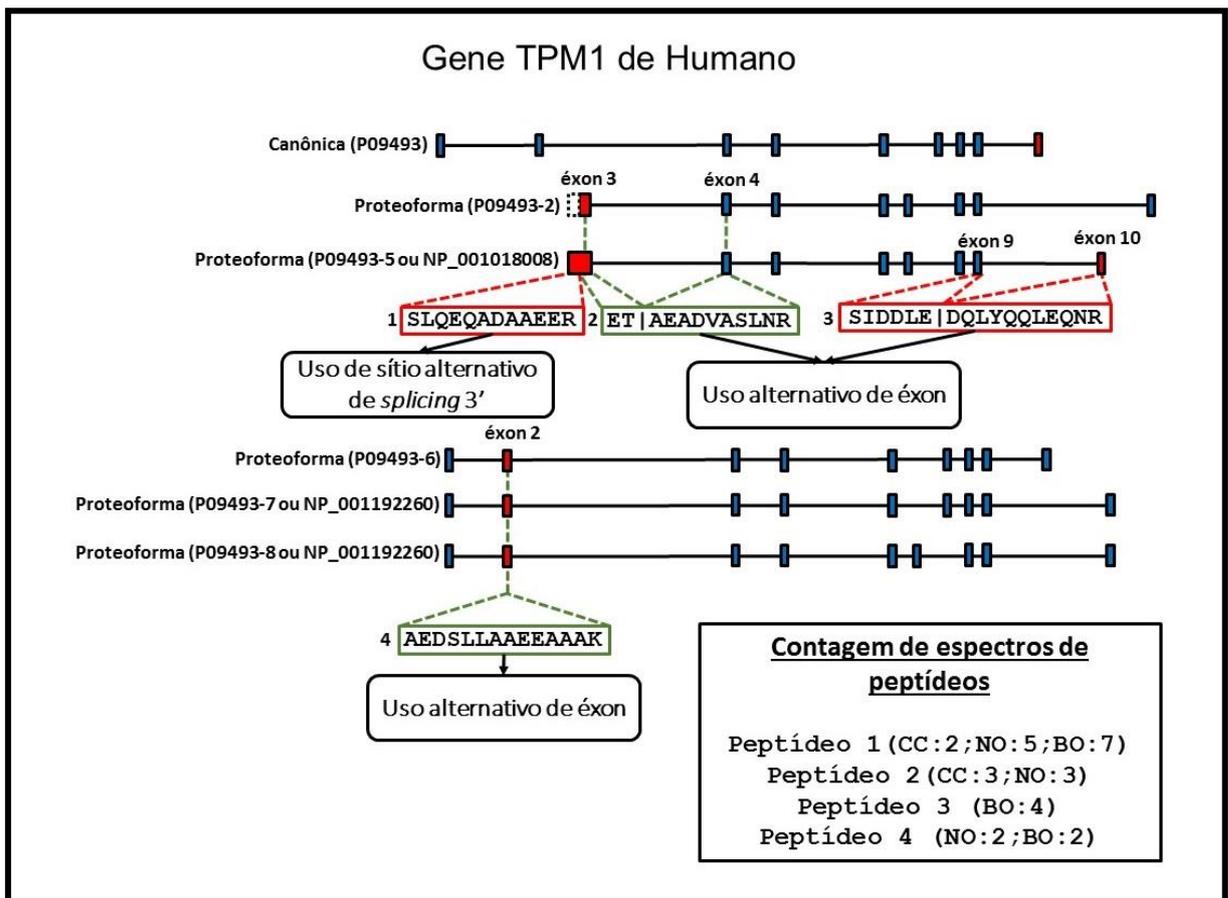


Figura 4.27.1 – Esquema representativo das diferenças no padrão de *splicing* da proteína canônica e proteoformas do gene TPM1 de humano. Retângulos vermelhos representam os éxons exclusivos de proteoformas onde os peptídeos identificados em nossas análises mapearam. Retângulos azuis representam os éxons das proteínas canônicas e proteoformas que não tiveram um peptídeo correspondente. Os retângulos brancos com contorno pontilhado representam a ausência de parte de um éxon. Os quadros e linhas pontilhadas vermelhas representam os peptídeos proteotípicos identificados em um ou mais experimentos de MS e suas respectivas localizações nos éxons. Os quadros verdes representam os peptídeos identificados compartilhados entre proteoformas e as linhas pontilhadas verdes representam suas respectivas localizações nos éxons das proteoformas. Entre parênteses a contagem de espectros de cada peptídeo no corpo caloso (CC) e nervo óptico (NO) e bulbo olfatório (BO).

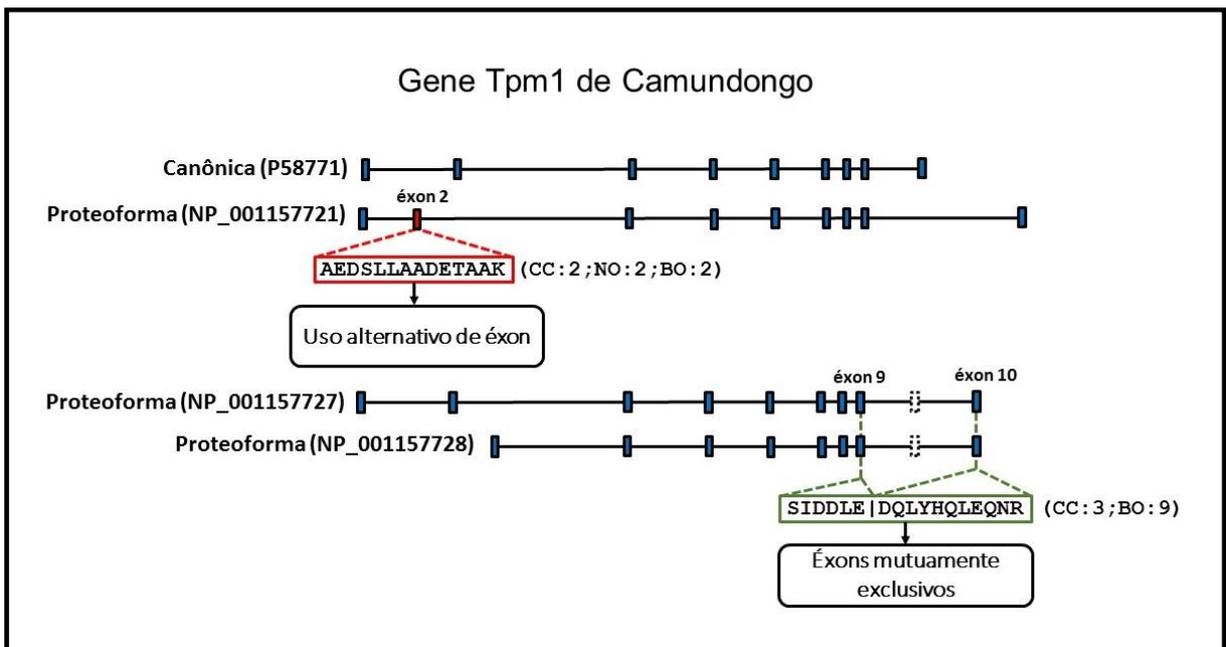


Figura 4.27.2 – Esquema representativo das diferenças no padrão de *splicing* da proteína canônica e proteoformas do gene Tpm1 de camundongo. Retângulos vermelhos representam os éxons exclusivos de proteoformas onde os peptídeos identificados em nossas análises mapearam. Retângulos azuis representam os éxons das proteínas canônicas e proteoformas que não tiveram um peptídeo correspondente. Os retângulos brancos com contorno pontilhado representam a ausência de um éxon. O quadro e linhas pontilhadas vermelhas representam o peptídeo proteotípico identificado e sua respectiva localização nos éxons. O quadro verde representa o peptídeo identificado compartilhado entre proteoformas e as linhas pontilhadas verdes representam sua respectiva localização nos éxons das proteoformas. Entre parênteses a contagem de espectros de cada peptídeo no corpo caloso (CC) e nervo óptico (NO) e bulbo olfatório (BO).

Nós identificamos dois peptídeos (“IQVLQQQADDAEER” e “EQAEAEVASLNR”) compartilhados entre as proteoformas P06753-2/NP_705935; P06753-3/NP_001036817; P06753-5/NP_001036816; P06753-4/NP_001036818 e P06753-6 do gene TPM3 de humano. Em camundongo, para o gene ortólogo (Tpm3), identificamos quatro peptídeos (“IQVLQQQADDAEER”; “EQAEAEVASLNR”; “TIDDLEDK” e “EEHLCTQR”) compartilhados entre as proteoformas NP_071709; NP_001258693; NP_001240669 e P21107-2/NP_001240667. O mapeamento dos peptídeos ao genoma das respectivas espécies mostrou eventos de uso alternativo de éxon em relação à forma proteica canônica em humano (Figura 4.28.1) e camundongo (Figura 4.28.2).

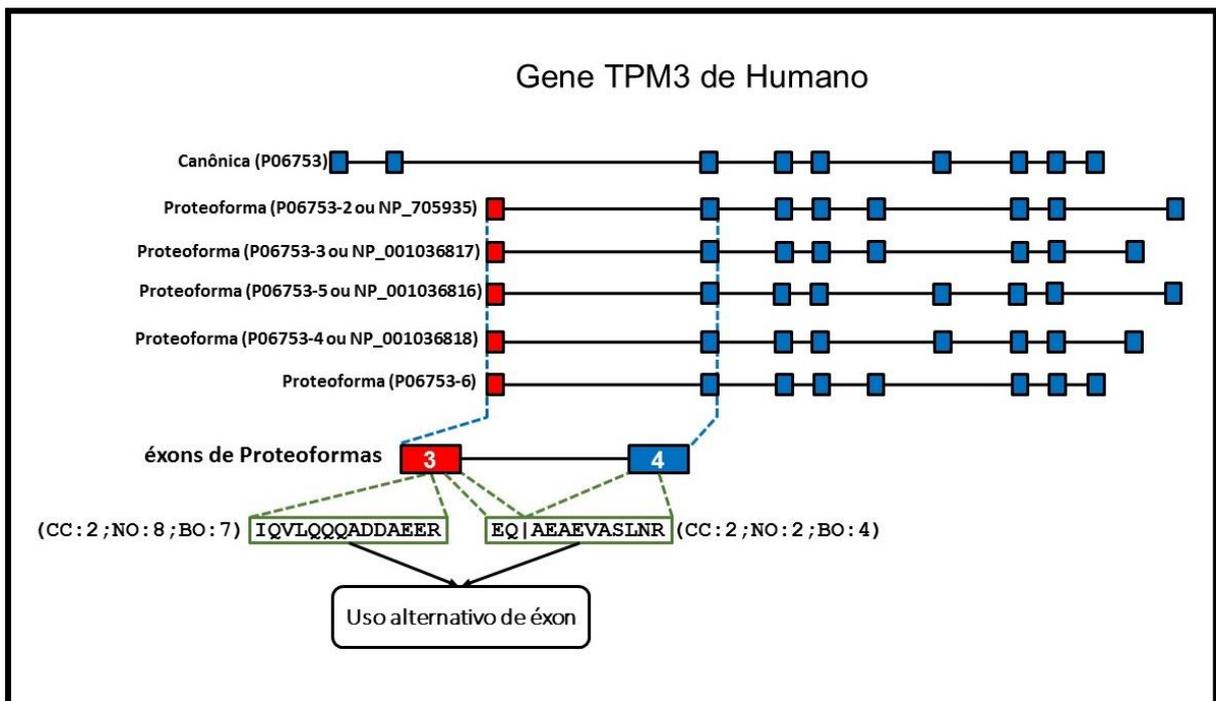


Figura 4.28.1 – Esquema representativo das diferenças no padrão de *splicing* da proteína canônica e proteoformas do gene TPM3 de humano. Retângulos vermelhos representam os éxons exclusivos de proteoformas onde os peptídeos identificados em nossas análises mapearam. Retângulos azuis representam os éxons das proteínas canônicas e proteoformas que não tiveram um peptídeo correspondente. Os quadros verdes representam os peptídeos identificados compartilhados entre proteoformas e as linhas pontilhadas verdes representam suas respectivas localizações nos éxons das proteoformas. As linhas pontilhadas azuis representam a ampliação dos éxons. Entre parênteses a contagem de espectros de cada peptídeo no corpo caloso (CC) e nervo óptico (NO) e bulbo olfatório (BO).

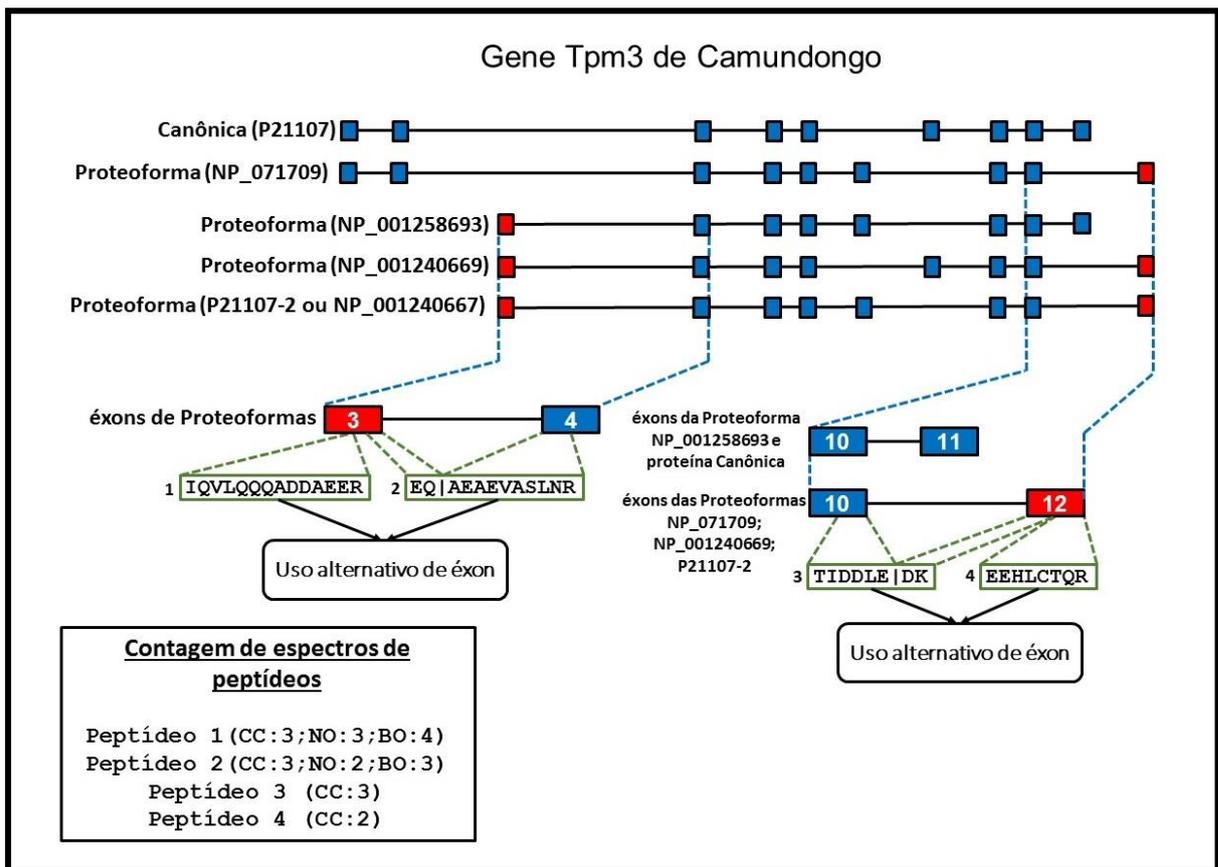


Figura 4.28.2 – Esquema representativo das diferenças no padrão de *splicing* da proteína canônica e proteoformas do gene Tpm3 de camundongo. Retângulos vermelhos representam os éxons exclusivos de proteoformas onde os peptídeos identificados em nossas análises mapearam. Retângulos azuis representam os éxons das proteínas canônicas e proteoformas que não tiveram um peptídeo correspondente. Os quadros verdes representam os peptídeos identificados compartilhados entre proteoformas e as linhas pontilhadas verdes representam suas respectivas localizações nos éxons das proteoformas. As linhas pontilhadas azuis representam a ampliação dos éxons. Entre parênteses a contagem de espectros de cada peptídeo no corpo caloso (CC) e nervo óptico (NO) e bulbo olfatório (BO).

4.3.1.6 Proteoformas de humano e camundongo inferidas por peptídeos proteotípicos idênticos

Como demonstrado nos subtópicos anteriores as proteoformas dos genes PRKCB, STXBP1, CADM1, HNRNPK, PKM e CRMP1 de humano e de seus ortólogos em camundongo foram inferidas por peptídeos proteotípicos idênticos. Ao alinharmos as sequências destas proteoformas de humano e camundongo observamos que as do gene PRKCB possuem 99% de identidade (Figura 29), as do gene STXBP1 100% de identidade (Figura 30), as do gene CADM1 98% de identidade (Figura 31), as do gene HNRNPK 99% de identidade (Figura 32), as do gene PKM possuem 98% de identidade (Figura 33) e as do gene CRMP1 98% de

identidade (Figura 34). As características principais que estas proteoformas possuem em comum estão resumidas na tabela 4.9.

Range 1: 1 to 673 Graphics				▼ Next Match	▲ Previous Match
Score	Expect	Method	Identities	Positives	Gaps
1395 bits(3612)	0.0	Compositional matrix adjust.	665/673(99%)	668/673(99%)	0/673(0%)
Query 1	MADPAAGPPPSEGEESTVRFARKGALRQKNVHEVKNHKFTARFFKQPTFCSHCTDFIWGF				60
Sbjct 1	MADPAAGPPPSEGEESTVRFARKGALRQKNVHEVKNHKFTARFFKQPTFCSHCTDFIWGF				60
Query 61	GKQGFQCQVCCFVWHKRCHEFVTFSCPGADKGPASDDPRSKHKFKIHTYSSPTFCDCGCS				120
Sbjct 61	GKQGFQCQVCCFVWHKRCHEFVTFSCPGADKGPASDDPRSKHKFKIHTYSSPTFCDCGCS				120
Query 121	LLYGLIHQGMKCDTCMMNVHKRCVMNVPSLCGTDHTERRGRIYIQAIDRDVLIIVLRDA				180
Sbjct 121	LLYGLIHQGMKCDTCMMNVHKRCVMNVPSLCGTDHTERRGRIYIQAIDR+VLIV+VRDA				180
Query 181	KNLVMPDMPNGLSDPYVKLKLIPDPKSESKQKTKTIKCSLNPEWNETFRFQLKESDKDRRL				240
Sbjct 181	KNLVMPDMPNGLSDPYVKLKLIPDPKSESKQKTKTIKCSLNPEWNETFRFQLKESDKDRRL				240
Query 241	SVEIWDWDLTSRNDFMGSLSFGISSELQKASVDGWFKLLSQEEGEYFNVPVPEGSEANEE				300
Sbjct 241	SVEIWDWDLTSRNDFMGSLSFGISSELQKA VDGWFKLLSQEEGEYFNVPVPEGSE NEE				300
Query 301	LRQKFERAKISQGTKVPEEKTNTVSKFDNNGNRDRMKLTDNFNLMVLGKGSFGKVMMLSE				360
Sbjct 301	LRQKFERAKI QGTK PEEKT NT+SKFDNNGNRDRMKLTDNFNLMVLGKGSFGKVMMLSE				360
Query 361	RKGTDELYAVKILKDDVVIQDDVECTMVEKRVLALPGKPPFLTQLHSCFQTMDRLYFVM				420
Sbjct 361	RKGTDELYAVKILKDDVVIQDDVECTMVEKRVLALPGKPPFLTQLHSCFQTMDRLYFVM				420
Query 421	EYVNGGDLMYHIQQVGRFKEPHAVFYAAEIAIGLFFLQSKGIIYRDLKLDNVMMLDSEGHI				480
Sbjct 421	EYVNGGDLMYHIQQVGRFKEPHAVFYAAEIAIGLFFLQSKGIIYRDLKLDNVMMLDSEGHI				480
Query 481	KIADFGMCKENIWDGVTTKTF CGTPDYIAPEIIAYQPYGKSVDWAFGVLLYEMLAGQAP				540
Sbjct 481	KIADFGMCKENIWDGVTTKTF CGTPDYIAPEIIAYQPYGKSVDWAFGVLLYEMLAGQAP				540
Query 541	FEGEDEDELFSIMEHNVAYPKSMSKEAVAICKGLMTKHPGKRLGCGPEGERDIKEHAFF				600
Sbjct 541	FEGEDEDELFSIMEHNVAYPKSMSKEAVAICKGLMTKHPGKRLGCGPEGERDIKEHAFF				600
Query 601	RYIDWEKLERKEIQPPYKPKACGRNAENFDRFFTRHPPVLTTPDQEVIRNIDQSEFEFEGFS				660
Sbjct 601	RYIDWEKLERKEIQPPYKPKACGRNAENFDRFFTRHPPVLTTPDQEVIRNIDQSEFEFEGFS				660
Query 661	FVNSEFLKPEVKS	673			
Sbjct 661	<u>FVNSEFLKPEVKS</u>	673			

Figura 4.29 – Alinhamento entre a proteoforma P05771-2 (*Query*) do gene PRKCB de humano e a proteoforma P68404-2 (*Sbjct*) do gene Prkcb de camundongo. O peptídeo proteotípico identificado nos dados de MS está marcado em vermelho.

Range 1: 1 to 603 Graphics				▼ Next Match	▲ Previous Match
Score	Expect	Method	Identities	Positives	Gaps
1251 bits(3238)	0.0	Compositional matrix adjust.	603/603(100%)	603/603(100%)	0/603(0%)
Query 1		MAPIGLKAVVGEKIMHDVIKKVKKKGEWKVLVVDQLSMRMLSSCCKMTDIMTEGITIVED		60	
Sbjct 1		MAPIGLKAVVGEKIMHDVIKKVKKKGEWKVLVVDQLSMRMLSSCCKMTDIMTEGITIVED		60	
Query 61		INKRREPLPSLEAVYLITPSEKSVHSLISDFKDPPTAKYRAAHVFFTDSCPDALFNELVK		120	
Sbjct 61		INKRREPLPSLEAVYLITPSEKSVHSLISDFKDPPTAKYRAAHVFFTDSCPDALFNELVK		120	
Query 121		SRAAKVIKTLTEINIAFLPYESQVYSLDSADSFQSFYSPHKAQMKNPILERLAEQIATLC		180	
Sbjct 121		SRAAKVIKTLTEINIAFLPYESQVYSLDSADSFQSFYSPHKAQMKNPILERLAEQIATLC		180	
Query 181		ATLKEYPAVRYRGEYKDNALLAQLIQDKLDAYKADDPTMGEGPDKARSQLLILDRGFDP		240	
Sbjct 181		ATLKEYPAVRYRGEYKDNALLAQLIQDKLDAYKADDPTMGEGPDKARSQLLILDRGFDP		240	
Query 241		SPVLHELTFQAMSYDLLPIENDVYKYETSGIGEARVKEVLLDEDDDLWIALRHKHIAEVS		300	
Sbjct 241		SPVLHELTFQAMSYDLLPIENDVYKYETSGIGEARVKEVLLDEDDDLWIALRHKHIAEVS		300	
Query 301		QEVTRSLKDFSSSKRMNTGEKTTMRDLSQMLKKMPQYQKELSKYSTHLHLAEDCMKHYQG		360	
Sbjct 301		QEVTRSLKDFSSSKRMNTGEKTTMRDLSQMLKKMPQYQKELSKYSTHLHLAEDCMKHYQG		360	
Query 361		TVDKLCRVEQDLAMGTDAEAGEKIKDPMRAIVPILLDANVSTYDKIRIILLYIFLKNGITE		420	
Sbjct 361		TVDKLCRVEQDLAMGTDAEAGEKIKDPMRAIVPILLDANVSTYDKIRIILLYIFLKNGITE		420	
Query 421		ENLNKLIQHAQIPPEDSEIITNMAHLGVPIVDSTLRRRSKPERKERISEQTYQLSRWTP		480	
Sbjct 421		ENLNKLIQHAQIPPEDSEIITNMAHLGVPIVDSTLRRRSKPERKERISEQTYQLSRWTP		480	
Query 481		IIKDIMEDIEDKLDTKHYPYISTRSSASFSTTAVSARYGHWKKNKAPGEYRSGPRLIIF		540	
Sbjct 481		IIKDIMEDIEDKLDTKHYPYISTRSSASFSTTAVSARYGHWKKNKAPGEYRSGPRLIIF		540	
Query 541		ILGGVSLNEMRCAYEVTQANGKWEVLIGSTHILTPTKFLMDLRHPDFRESSRVSFEDQAP		600	
Sbjct 541		ILGGVSLNEMRCAYEVTQANGKWEVLIGSTHILTPTKFLMDLRHPDFRESSRVSFEDQAP		600	
Query 601	TME	603			
Sbjct 601	TME	603			

Figura 4.30 – Alinhamento entre a proteoforma P61764-2 (*Query*) do gene STXPB1 de humano e a proteoforma O08599-2 do gene (*Sbjct*) Stxbp1 de camundongo. O peptídeo proteotípico identificado nos dados de MS está marcado em vermelho.

Range 1: 1 to 417 Graphics				▼ Next Match	▲ Previous Match
Score	Expect	Method	Identities	Positives	Gaps
848 bits(2190)	0.0	Compositional matrix adjust.	408/417(98%)	409/417(98%)	3/417(0%)
Query 1	MASVVLPSGSQCAAAAA---AAAPPGLRLRLLLLLFSAAALIPTGDGQNLFTKDVTVIEG				57
Sbjct 1	MAS VLPSSGSQCAAAAA AAAPPGLRLRLLLLL SAAALIPTGDGQNLFTKDVTVIEG				60
Query 58	EVATISCQVNKSDDSVIQLLNPNRQTIYFRDFRPLKDSRFQLLNFSSELKVSLTNVSIS				117
Sbjct 61	EVATISCQVNKSDDSVIQLLNPNRQTIYFRDFRPLKDSRFQLLNFSSELKVSLTNVSIS				120
Query 118	DEGRYFCQLYTDPPQESYTTITVLVPPRNLMDIQKDTAVEGEEIEVNCTAMASKPATTI				177
Sbjct 121	DEGRYFCQLYTDPPQESYTTITVLVPPRNLMDIQKDTAVEGEEIEVNCTAMASKPATTI				180
Query 178	RWFKGNTELKKGSEVEEWSDMYTVTSQLMLKVHKEDDGVPVICQVEHPAVTGNLQTQRYL				237
Sbjct 181	RWFKGN ELKKGSEVEEWSDMYTVTSQLMLKVHKEDDGVPVICQVEHPAVTGNLQTQRYL				240
Query 238	EVQYKPVVHIQMTYPLQGLTREGDALELTCEAIGKQPVMVTWVRVDEMPQHAVLSGPN				297
Sbjct 241	EVQYKPVVHIQMTYPLQGLTREGDA ELTCEAIGKQPVMVTWVRVDEMPQHAVLSGPN				300
Query 298	LFINNLNKTDNGTYRCEASNIVGKAHSDYMLVYDSRAGEEGSIRAVDHAVIGGVAVVV				357
Sbjct 301	LFINNLNKTDNGTYRCEASNIVGKAHSDYMLVYDSRAGEEG+I AVDHAVIGGVAVVV				360
Query 358	FAMLCLLIILGRYFARHKGTYFTHEAKGADDAADADTAIINAEGGQNNSEEKKEYFI				414
Sbjct 361	FAMLCLLIILGRYFARHKGTYFTHEAKGADDAADADTAIINAEGGQNNSEEKKEYFI				417

Figura 4.31 – Alinhamento entre a proteoforma Q9BY67-5 (*Query*) do gene CADM1 de humano e a proteoforma Q8R5M8-4 (*Sbjct*) do gene Cadm1 de camundongo. O peptídeo proteotípico identificado nos dados de MS está marcado em vermelho.

Range 1: 1 to 435 Graphics				▼ Next Match	▲ Previous Match	
Score	Expect	Method	Identities	Positives	Gaps	
879 bits(2272)	0.0	Compositional matrix adjust.	434/435(99%)	435/435(100%)	0/435(0%)	
Query 1		METEQPEETFPNTE	TNGEFGK	PAEDMEE	EQAFKRSRNTDEMVELRILLQSKNAGAVIGK	60
Sbjct 1		METEQPEETFPNTE	TNGEFGK	PAEDMEE	EQAFKRSRNTDEMVELRILLQSKNAGAVIGK	60
Query 61		GGKNIKALRTDYNASVSPDSSGPERILSISADIETIGEILKKIIP	LEEYQHYKGSDFD			120
Sbjct 61		GGKNIKALRTDYNASVSPDSSGPERILSISADIETIGEILKKIIP	LEEYQHYKGSDFD			120
Query 121		CELRLLIHQSLAGGIIGVKGAKIKELRENTQTTIKLFQ	ECCPHSTDRVVLIGGK	PDRVVE		180
Sbjct 121		CELRLLIHQSLAGGIIGVKGAKIKELRENTQTTIKLFQ	ECCPHSTDRVVLIGGK	PDRVVE		180
Query 181		CIKIILDLISESPIKGRAQPYDPNFYDETYDYG	GFTMMFDDRGRPVGF	PMRGRGGFDRM		240
Sbjct 181		CIKIILDLISESPIKGRAQPYDPNFYDETYDYG	GFTMMFDDRGRPVGF	PMRGRGGFDRM		240
Query 241		PPGRGGRMPSPRRDYDDMSPRRGPPPPPPGRGGRGGS	RARNLPLPPPPPPRGDLMAYD			300
Sbjct 241		PPGRGGRMPSPRRDYDDMSPRRGPPPPPPGRGGRGGS	RARNLPLPPPPPPRGDLMAYD			300
Query 301		RRGRPGDRYDGMVGFSADETWDSAIDTWSPSEWQ	MAYEPQGGSGYDYSYAGGRGSYDGLG			360
Sbjct 301		RRGRPGDRYDGMVGFSADETWDSAIDTWSPSEWQ	MAYEPQGGSGYDYSYAGGRGSYDGLG			360
Query 361		GPIITTQVTIPKDLAGSIIGKGGQRIKQIRHESGASIK	IDEPLEGSEDRIITITGTQDQI			420
Sbjct 361		GPIITTQVTIPKDLAGSIIGKGGQRIKQIRHESGASIK	IDEPLEGSEDRIITITGTQDQI			420
Query 421		QNAQYLLQNSVKQYA	435			
Sbjct 421		QNAQYLLQNSVKQYS	435			

Figura 4.32 – Alinhamento entre a proteoforma P61978-3 (*Query*) do gene HNRNPK de humano e a proteoforma P61979-3 (*Sbjct*) do gene Hnrnpk de camundongo. O peptídeo proteotípico identificado nos dados de MS está marcado em vermelho.

Range 1: 1 to 531		Graphics		▼ Next Match	▲ Previous Match
Score	Expect	Method	Identities	Positives	Gaps
1071 bits(2769)	0.0	Compositional matrix adjust.	519/531(98%)	522/531(98%)	0/531(0%)
Query 1	MSKPHSEAGTAFIQTQQLHAAMADTFLEHMCRLDIDSPPITARNTGIICTIGPASRSVET			60	
Sbjct 1	M KPHSEAGTAFIQTQQLHAAMADTFLEHMCRLDIDS PITARNTGIICTIGPASRSVE			60	
Query 61	LKEMIKSGMNVARLNFSGHGTHEYHAETIKNVRTATESFASDPILYRPVAVALDTKGPEIR			120	
Sbjct 61	LKEMIKSGMNVARLNFSGHGTHEYHAETIKNVREATESFASDPILYRPVAVALDTKGPEIR			120	
Query 121	TGLIKSGSTAEEVLKKGATLKITLDNAYMEKCDENILWLDYKNICKVVEVGSKIYVDDGL			180	
Sbjct 121	TGLIKSGSTAEEVLKKGATLKITLDNAYMEKCDENILWLDYKNICKVVEVGSKIYVDDGL			180	
Query 181	ISLQVKQKGADFLVTEVENGGSLGSKKGVNLPAAVDLPAVSEKDIQDLKFGVEQDQVDMV			240	
Sbjct 181	ISLQVK+KGADFLVTEVENGGSLGSKKGVNLPAAVDLPAVSEKDIQDLKFGVEQDQVDMV			240	
Query 241	FASFIRKASDVHEVRKVLGEKGNIKIISKIENHEGVRRFDEILEASDGMVARGDLGIE			300	
Sbjct 241	FASFIRKA+DVHEVRKVLGEKGNIKIISKIENHEGVRRFDEILEASDGMVARGDLGIE			300	
Query 301	IPAEEKVFLAQKMMIGRCNRAGKPVICATQMLESMIKKPRPTRAEGSDVANAVLDGADCIM			360	
Sbjct 301	IPAEEKVFLAQKMMIGRCNRAGKPVICATQMLESMIKKPRPTRAEGSDVANAVLDGADCIM			360	
Query 361	LSGETAKGDYPLEAVRMOHLIAREAEAAAMFHRKLFEEELVRASSHSTDLEAMAMGSVEAS			420	
Sbjct 361	LSGETAKGDYPLEAVRMOHLIAREAEAAAMFHR LFEELVRASSHSTDLEAMAMGSVEAS			420	
Query 421	YKCLAAALIVLTESGRSAHQVARYRPRAPIIAVTRNPQTARQAHLYRGIFPVLCKDPVQE			480	
Sbjct 421	YKCLAAALIVLTESGRSAHQVARYRPRAPIIAVTRNPQTARQAHLYRGIFPVLCKD V			480	
Query 481	AWAEDVDLRVNFAMNVGKARGFFKKGDVVIVLTGWRPGSGFTNTMRVVPV			531	
Sbjct 481	AWAEDVDLRVN AM+VGKARGFFKKGDVVIVLTGWRPGSGFTNTMRVVPV			531	

Figura 4.33 – Alinhamento entre a proteoforma P14618-2 (*Query*) do gene PKM de humano e a proteoforma P52480-2 (*Sbjct*) do gene Pkm de camundongo. O peptídeo proteotípico identificado nos dados de MS está marcado em vermelho.

Range 1: 1 to 686 Graphics				▼ Next Match	▲ Previous Match
Score	Expect	Method	Identities	Positives	Gaps
1373 bits(3555)	0.0	Compositional matrix adjust.	662/686(97%)	675/686(98%)	0/686(0%)
Query 1	MADRRRAWNTEDDLPVYLARPGSAAQT	PRQKYGGMFAAVEGAYENKTI	DFDAYSVGRRGS	60	
Sbjct 1	MA+RRRAWNTEDDLPVYLARPGSAAQT	PRQKYGGMFAAVEGAYENKTI	DFDAYSVGRRGS	60	
Query 61	ARTPRSAGRPDVGLPGPGGSED	ASDVSEPSGS	SAVSSPGERDERPPTLRIRRPAPRDLP	120	
Sbjct 61	ARTPRSAGRPDVGLPGPGGSED	ASDVSEPSGS	SAVSSPGERDRDRPALRIRCPAPRDLP	120	
Query 121	LGRDNGQSDRLLIKGGRI	INDDQSLYADVLEDGLIKQIGENL	IIVPGGVKTI	EANGRMVI	180
Sbjct 121	LGRDNGQSDRLLI+GGRI	INDDQSLYADVLEDGLIKQIGENL	IIVPGGVKTI	EANGRMVI	180
Query 181	PGGIDVNTYLQKPSQGMT	AADFFQGT	RAALVGGTTMIIDHVPEPGSSLLTSFEKWHEA	240	
Sbjct 181	PGGIDVNTYLQKPSQGMT	AADFFQGT	KAALAGTTMIIDHVPEPGSSLLTSFEKWHEA	240	
Query 241	ADTKSCCDYSLHVDITS	WYDGVREELEVLVQDKGVNS	FQVYVYKDYQMSDSQLYEAF	300	
Sbjct 241	ADTKSCCDYSLHVDITS	WYDGVREELEVLVQDKGVNS	FQVYVYKDYQMSDSQLYEAF	300	
Query 301	FLKGLGAVILVHAENGDLIAQE	QKRILEMGITGPEGHALS	SRPEELEAEAVFRAITIAAGRI	360	
Sbjct 301	FLKGLGAVILVHAENGDLIAQE	QKRILEMGITGPEGHALS	SRPEELEAEAVFRAITIAAGRI	360	
Query 361	NCPVYITKVMKSAA	DIALARKKGPLVFG	GEPIAASLGTGTHYWSKNNAKAAAFVTSPP	420	
Sbjct 361	NCPVYITKVMKSAA	DIALARKKGPLVFG	GEPIAASLGTGTHYWSKNNAKAAAFVTSPP	420	
Query 421	LSPDPTTDPDYLTSLLACGDL	QVTGSGHCPYSTAQKAVGKDNFTLI	PEGVNGIEERMTVW	480	
Sbjct 421	LSPDPTTDPDYLTSLLACGDL	QVTGSGHCPYSTAQKAVGKDNFTLI	PEGVNGIEERMTVW	480	
Query 481	DKAVATGKMDENQFVAVT	STNAAKIFNLYPRKGRI	AVGSDADVVIWDPDKLKITAKSHK	540	
Sbjct 481	DKAVATGKMDENQFVAVT	STNAAKIFNLYPRKGRI	AVGSDADVVIWDPDKMKITAKSHK	540	
Query 541	SAVEYNIFEGMECHG	SPLVVISQKIVFEDGNI	VNKGMRGFI	PRKAFPEHLYQRVKIRN	600
Sbjct 541	S VEYNIFEGMECHG	SPLVVISQKIVFEDGNI	+V+KGMGRGFI	PRK FPEHLYQRV+IR+STVEYNIFEGMECHG	600
Query 601	KVFGLQVSRGMYDGPVVE	PATPKYATPAPS	AKSSPSKHQPPPIRNLHQS	NFSLSGAQI	660
Sbjct 601	KVFGL VSRGMYDGPVVE	PATPK+A PAPS	AKSSPSKHQPPPIRNLHQS	NFSLSGAQI	660
Query 661	DDNNPRRTGHRIV	APPGGRSNITSLG	686		
Sbjct 661	DDNNPRRTGHRIV	APPGGRSNITSLG	686		

Figura 4.34 – Alinhamento entre a proteoforma Q14194-2 (*Query*) do gene CRMP1 de humano e a proteoforma NP_001129530 (*Sbjct*) do gene Crmp1 de camundongo. O peptídeo proteotípico identificado nos dados de MS está marcado em vermelho.

Tabela 4.9 – Informações sobre as proteoformas de genes ortólogos entre humano e camundongo inferidas nos dados de MS a partir de peptídeos proteotípicos idênticos.

Símbolo Gênico	Identidade	Padrão de Splicing	Região do Cérebro
PKM	98%	Éxons mutuamente exclusivos	Nervo óptico, corpo caloso e bulbo olfatório
HNRNPK	99%	Uso alternativo de éxon	Bulbo olfatório
CRMP1	98%	Uso alternativo de éxon	Bulbo olfatório
PRKCB	99%	Uso alternativo do sítio de <i>splicing</i> 3'	Bulbo olfatório
STXBP1	100%	Uso alternativo de éxon	Bulbo olfatório
CADM1	98%	Uso alternativo de éxon	Bulbo olfatório

5 DISCUSSÃO

O presente trabalho apresentou a utilização de uma metodologia computacional para a avaliação da diversidade de *splicing* alternativo no proteoma do cérebro de humano e camundongo, através de uma abordagem proteogenômica. O desenvolvimento de diferentes metodologias para construção de repositórios de sequências proteicas personalizados tem sido um dos focos do nosso grupo de pesquisa. Neste projeto, foi empregado um destes métodos para a construção de um repositório de sequências proteicas personalizado de camundongo, baseado no modelo utilizado em dados de humano desenvolvido por Tavares e colaboradores (2017) (7).

Como demonstrado nos resultados, a construção do repositório SpliceProt foi baseada na metodologia das matrizes ternárias (132), desenvolvida pelo nosso grupo de pesquisa. Esta metodologia apresentou ótima eficiência na identificação de eventos de *splicing* em dados de transcriptômica e genômica, como demonstrado por Tavares e colaboradores (113). Entretanto, uma limitação que esta técnica possui é a análise de eventos de *splicing* nos éxons das extremidades das unidades transcricionais, uma vez que estes podem corresponder a artefatos de sequenciamento, ou até mesmo, serem frutos de um evento de uso alternativo de promotores. Portanto, nesta dissertação, optamos por continuar não considerando eventos de *splicing* nestas regiões, com o intuito de evitarmos a identificação de falsos resultados positivos.

Nós utilizamos tanto o repositório de sequências proteicas personalizado de humano, previamente desenvolvido por Tavares e colaboradores (2017) (7), quanto o de camundongo, gerado neste projeto, para a identificação de peptídeos em dados de espectrometria de massas de amostras proteicas de diferentes regiões do cérebro de humano e camundongo (corpo caloso, nervo óptico e bulbo olfatório). A partir destas análises, nós inferimos a expressão de um total de 7.086 proteínas canônicas de humano e 13.321 de camundongo. Acreditamos que inferimos uma maior quantidade de proteínas de camundongo devido à alta acurácia dos espectros de massas MS/MS quando comparado aos de humano, vide as especificações dos espectrômetros de massas utilizados em cada experimento.

Atualmente, uma das dificuldades ao se analisar dados de MS é a identificação de peptídeos proteotípicos, isto é, aqueles exclusivos de uma proteína que permitem sua inferência de maneira inequívoca (96). Em nosso estudo, nós inferimos a expressão de um total de 3.289 proteínas canônicas e 23 proteoformas de genes ortólogos entre humano e camundongo, das quais seis proteoformas (PKM, HNRNPK, CRMP1, PRKCB, STXBP1, CADM1) foram inferidas por peptídeos proteotípicos com a mesma sequência em humano e camundongo. Acreditamos que as proteoformas de genes ortólogos inferidas em humano e camundongo sejam as mesmas, devido ao padrão de *splicing* alternativo idêntico que possuem e frequentemente a elevada similaridade das sequências das proteoformas (Tabela 5.1).

A seguir, apresentaremos algumas informações funcionais disponíveis na literatura a respeito dos produtos destes genes e algumas considerações sobre cada uma destas proteoformas.

O gene PKM de humano e seu ortólogo em camundongo (Pkm) codificam para a proteína piruvato quinase M. Esta proteína é uma enzima glicolítica que catalisa a transferência de um grupo fosforil de fosfoenolpiruvato (PEP) para ADP, gerando ATP [revisado por (143)]. Em nossas análises, nós inferimos a proteoforma M1 do gene piruvato quinase M em humano (Uniprot: P14618-2) e em camundongo (Uniprot: P52480-2) a partir dos mesmos peptídeos em dados de MS das três regiões do cérebro analisadas neste projeto. Esta proteoforma é marcada pela presença do éxon 9 e ausência do éxon 10 (144) e normalmente é expressa em tecidos maduros (145). Entretanto, no cérebro em estágios de desenvolvimento tumoral tem sido observado um declínio abrupto de sua expressão dando lugar à super-expressão da forma canônica M2 (146).

O gene HNRNPK de humano e seu ortólogo em camundongo (Hnrnpk) codificam para a proteína Ribonucleoproteína nuclear heterogênea K. Além do núcleo, esta proteína é encontrada no citoplasma e mitocôndria, e está implicada em diversos processos como o remodelamento de cromatina, tradução (147), transcrição (148) e *splicing* (149). Foi observado que a formação de uma variante de *splicing* aberrante deste gene está relacionada à Síndrome de Au-Kline, caracterizada por incapacidade intelectual, defeitos cardíacos e alterações do tecido

conjuntivo e do esqueleto (150). Em nossas análises nós identificamos um peptídeo proteotípico da proteoforma 3 deste gene nos dados de MS de bulbo olfatório de humano (Uniprot: P61978-3) e camundongo (Uniprot: P61979-3). Atualmente, a proteoforma P61978-3 de humano não possui confirmação experimental, sendo este um achado inédito na literatura. Como mencionado anteriormente, acreditamos que o fato do mesmo peptídeo proteotípico ter sido identificado em humano e camundongo torna mais confiável a inferência desta proteoforma.

A proteína 1 relacionada à diidropirimidinase (tradução livre do inglês, *dihydropyrimidinase-related protein 1*) é codificada pelo gene CRMP1. Esta proteína é expressa no cérebro (151) e participa do processo de crescimento axonal (152,153). Nós identificamos em nossas análises a proteoforma 3 do Uniprot em dados de MS de humano (Q14194-2) e a proteoforma 1 do Refseq em dados de MS de camundongo (NP_001129530) a partir do mesmo peptídeo proteotípico. O fato de estas proteoformas possuírem 97% de identidade nos levou a acreditar que correspondem à mesma proteoforma. Outro fato importante a ser observado é que a proteoforma de camundongo não se encontra na base de dados Uniprot/Swiss-Prot. Portanto, a utilização do nosso repositório de sequências proteicas personalizado de camundongo possibilitou a identificação do espectro de massas correspondente ao peptídeo proteotípico desta proteoforma.

Nós identificamos um peptídeo proteotípico da proteoforma 2 da cadeia alfa-1 da tropomiosina (Tpm1) em dados de MS das três regiões do cérebro de camundongo. De maneira similar a proteoforma do gene de camundongo exposto no parágrafo anterior, este peptídeo foi identificado a partir de uma sequência oriunda do Refseq (NP_001157721) presente no repositório de camundongo desenvolvido neste projeto e ausente no banco de dados Uniprot/Swiss-Prot. Este achado reitera a importância de se utilizar repositórios de sequências proteicas personalizados para auxiliar a busca por correspondências entre os espectros de massas e sequências de peptídeos.

A proteína 1 de ligação à sintaxina (tradução livre do inglês, *Syntaxin-binding protein 1*) advinda do gene STXBP1 de humano e de seu ortólogo em camundongo (Stxbp1) participa do processo de tráfico de vesículas e liberação de neurotransmissores (154). Mutações no gene que codifica esta proteína estão

relacionadas a algumas formas de epilepsia infantil (155). Em um trabalho utilizando peixe zebra, mutações induzidas que causam perda de função deste gene geraram mutantes que apresentaram convulsões epiléticas (156). Também tem sido observado que esta proteína é super-expressa em casos de esquizofrenia e síndrome bipolar (157). Nós inferimos a proteoforma 2 deste gene em humano (Uniprot:P61764-2 ou Refseq:NP_002729) e camundongo (Uniprot:O08599-2 ou Refseq:NP_001107041) a partir de um mesmo peptídeo proteotípico. O alinhamento entre as sequências completas da proteoforma inferida em humano e camundongo apresentou 100% de identidade, o que nos levou a acreditar que estas são a mesma proteoforma expressas e provavelmente desempenham funções similares em ambos os organismos.

A molécula de adesão celular 1 é uma proteína advinda do gene CADM1 de humano. Esta proteína atua na adesão sináptica, desempenha um papel na formação de espinhos dendríticos, na montagem das sinapses, na migração neuronal, crescimento axonal e encaminhamento nos axônios de neurônios diferenciadores (158). Em nossas análises nós inferimos a proteoforma 5 desta proteína em humano (Uniprot: Q9BY67-5) e a proteoforma 4 em camundongo (Uniprot: Q8R5M8-4) a partir de um mesmo peptídeo proteotípico identificado em dados de MS de bulbo olfatório. Apesar de serem proteoformas com nomes diferentes na base de dados Uniprot, o alinhamento entre suas sequências completas apresentou 98% de identidade, com cobertura de 100%, e ambas possuem o mesmo peptídeo proteotípico identificado em nossas análises. Isto nos levou a acreditar que estas proteoformas ao menos são similares e provavelmente desempenham funções parecidas no contexto celular de ambos os organismos.

O gene PRKCB de humano e seu ortólogo Prkcb em camundongo codificam para a proteína quinase C beta. Foi demonstrado que a ativação das vias de transdução desta proteína diminui os agregados de proteínas beta amiloide, presentes no cérebro e vasos sanguíneos de pacientes com Mal de Alzheimer (159). Também foi observado que esta proteína participa da fosforilação de fatores transcricionais que estão envolvidos na proliferação e diferenciação celular (160). Nós inferimos em nossas análises a proteoforma beta II de humano (Uniprot: P05771-2) e de camundongo (Uniprot: P68404-2) a partir de um mesmo peptídeo proteotípico identificado em dados de MS de bulbo olfatório de humano e

camundongo. O fato das sequências completas das proteoformas de humano e camundongo possuírem 99% de identidade nos levou a acreditar que inferimos a mesma proteoforma em humano e camundongo a partir do mesmo peptídeo proteotípico.

Além disso, as proteoformas P14618-2 do gene PKM e P09493-5 do gene TPM1, inferidas em dados de MS das três regiões do cérebro de humano, assim como a proteoforma P61978-3 do gene HNRNPK, inferida em dados de MS de bulbo olfatório, também foram encontradas em dados de oligodendrócitos de humano em outra análise realizada pelo nosso grupo de pesquisa (7). Isto corrobora com os nossos resultados apresentados nesta dissertação.

O cérebro humano é um órgão complexo do ponto de vista molecular o que torna o entendimento de seu proteoma algo desafiador (161), especialmente devido ao *splicing* alternativo (48). Portanto, a inferência de proteoformas oriundas de *splicing* alternativo em diferentes regiões do cérebro de humano e camundongo contribuiu para uma melhor compreensão do proteoma deste órgão. A identificação de peptídeos proteotípicos idênticos, correspondentes a proteoformas de genes ortólogos entre humano e camundongo, tornou mais confiável à inferência destas proteoformas nos dados de MS.

Ademais, devido a questões éticas, muitas vezes são utilizados modelos murinos para se investigar o proteoma de tecidos e os resultados obtidos em camundongo são extrapolados ao ser humano, o que torna crucial a investigação comparativa entre o ser humano e o camundongo (63). Desta forma, a inferência de proteoformas de genes ortólogos realizada neste estudo contribui com o conhecimento a respeito das similaridades e diferenças entre o ser humano e camundongo. Neste sentido, atualmente, estamos viabilizando a confirmação experimental dos mRNAs por PCR quantitativo (RT-qPCR) em amostras de corpo caloso, nervo óptico e bulbo olfatório de camundongo.

6 CONCLUSÃO

A integração de dados de transcriptômica e proteômica, conhecida como proteogenômica, tem ganhado maior destaque ao longo dos últimos anos. Com o advento das tecnologias de alta vazão como RNA-seq e espectrometria de massas de proteínas os dados passaram a ser produzidos em larga escala, tornando a identificação de variantes de *splicing* algo desafiador. Esta dissertação apresentou a inferência de proteoformas em dados de espectrometria de massas de proteínas de três diferentes regiões do cérebro de humano e camundongo (corpo caloso, nervo óptico e bulbo olfatório), sendo tais resultados inéditos na literatura até o momento. Para tal, foi utilizado um repositório de sequências proteicas personalizado de humano, previamente desenvolvido, e de camundongo, desenvolvido neste projeto, a partir da tradução hipotética de ESTs e mRNAs de referências (Refseq). Portanto, acreditamos que a identificação de peptídeos compartilhados entre humano e camundongo, pertencentes à proteoformas de genes ortólogos, realizada neste projeto, contribuiu para o melhor conhecimento da diversidade de *splicing* do cérebro de humano e camundongo. Em termos de bioinformática, esta dissertação dá um passo adiante na questão de caracterização de proteoformas ortólogas de humano e camundongo, mas que há enormes desafios a serem superados na questão da automação no sistema de classificação de proteoformas ortólogas de duas espécies.

7 REFERÊNCIAS BIBLIOGRÁFICAS

1. Crick F. Central Dogma of Molecular Biology. *Nature*. Agosto de 1970;227(5258):561–3.
2. Shapiro JA. Revisiting the Central Dogma in the 21st Century. *Ann N Y Acad Sci*. Outubro de 2009;1178(1):6–28.
3. Behjati S, Tarpey PS. What is next generation sequencing? *Arch Dis Child Educ Pract Ed*. dezembro de 2013;98(6):236–8.
4. Kukurba KR, Montgomery SB. RNA Sequencing and Analysis. *Cold Spring Harb Protoc*. 13 de abril de 2015;2015(11):951–69.
5. Canas B. Mass spectrometry technologies for proteomics. *Brief Funct Genomic Proteomic*. 3 de fevereiro de 2006;4(4):295–320.
6. Ning K, Nesvizhskii AI. The utility of mass spectrometry-based proteomic data for validation of novel alternative splice forms reconstructed from RNA-Seq data: a preliminary assessment. *BMC Bioinformatics*. 2010;11(S11):S14.
7. Tavares R, Wajnberg G, Scherer N de M, Pauletti BA, Cassoli JS, Ferreira CG, et al. Unveiling alternative splice diversity from human oligodendrocyte proteome data. *J Proteomics*. 16 de janeiro de 2017;151:293–301.
8. Blanquart S, Varré J-S, Guertin P, Perrin A, Bergeron A, Swenson KM. Assisted transcriptome reconstruction and splicing orthology. *BMC Genomics*. 11 de novembro de 2016;17(Suppl 10):786.
9. Zambelli F, Pavesi G, Gissi C, Horner DS, Pesole G. Assessment of orthologous splicing isoforms in human and mouse orthologous genes. *BMC Genomics*. 1º de outubro de 2010;11:534.
10. Taneri B, Novoradovsky A, Snyder B, Gaasterland T. Databases for comparative analysis of human-mouse orthologous alternative splicing. In: *RECOMB Workshop on Comparative Genomics*. Springer; 2004. p. 123–131.
11. Gratuze M, Julien J, Morin F, Calon F, Hébert SS, Marette A, et al. High-fat, high-sugar, and high-cholesterol consumption does not impact tau pathogenesis in a mouse model of Alzheimer's disease-like tau pathology. *Neurobiol Aging*. 29 de julho de 2016;47:71–3.
12. Choi D-K, Pennathur S, Perier C, Tieu K, Teismann P, Wu D-C, et al. Ablation of the Inflammatory Enzyme Myeloperoxidase Mitigates Features of Parkinson's Disease in Mice. *J Neurosci*. 13 de julho de 2005;25(28):6594–600.
13. Schurov IL, Handford EJ, Brandon NJ, Whiting PJ. Expression of disrupted in schizophrenia 1 (DISC1) protein in the adult and developing mouse brain indicates its role in neurodevelopment. *Mol Psychiatry*. Setembro de 2004;9(12):1100–10.
14. Kawarabayashi T, Younkin LH, Saido TC, Shoji M, Ashe KH, Younkin SG. Age-Dependent Changes in Brain, CSF, and Plasma Amyloid β Protein in the

- Tg2576 Transgenic Mouse Model of Alzheimer's Disease. *J Neurosci.* 15 de janeiro de 2001;21(2):372–81.
15. Gilbert W. Why genes in pieces? *Nature.* 9 de fevereiro de 1978;271(5645):501–501.
 16. Proudfoot NJ, Furger A, Dye MJ. Integrating mRNA Processing with Transcription. *Cell.* Fevereiro de 2002;108(4):501–12.
 17. Early P, Rogers J, Davis M, Calame K, Bond M, Wall R, et al. Two mRNAs can be produced from a single immunoglobulin mu gene by alternative RNA processing pathways. *Cell.* junho de 1980;20(2):313–9.
 18. Smith LM, Kelleher NL. Proteoform: a single term describing protein complexity. *Nat Methods.* março de 2013;10(3):186–7.
 19. Schlüter H, Apweiler R, Holzhütter H-G, Jungblut PR. Finding one's way in proteomics: a protein species nomenclature. *Chem Cent J.* 9 de setembro de 2009;3:11.
 20. Modrek B, Lee CJ. Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss. *Nat Genet.* junho de 2003;34(2):177–80.
 21. Barbosa-Morais NL, Irimia M, Pan Q, Xiong HY, Gueroussov S, Lee LJ, et al. The evolutionary landscape of alternative splicing in vertebrate species. *Science.* 21 de dezembro de 2012;338(6114):1587–93.
 22. Wang Y, Liu J, Huang B, Xu Y, Li J, Huang L, et al. Mechanism of alternative splicing and its regulation (Review). *Biomed Rep [Internet].* 17 de dezembro de 2014 [citado 28 de novembro de 2016]; Disponível em: <http://www.spandidos-publications.com/10.3892/br.2014.407>
 23. Kim E, Magen A, Ast G. Different levels of alternative splicing among eukaryotes. *Nucleic Acids Res.* janeiro de 2007;35(1):125–31.
 24. Pan Q, Bakowski MA, Morris Q, Zhang W, Frey BJ, Hughes TR, et al. Alternative splicing of conserved exons is frequently species-specific in human and mouse. *Trends Genet.* 1º de fevereiro de 2005;21(2):73–7.
 25. Goodman SJ, Branda CS, Robinson MK, Burdine RD, Stern MJ. Alternative splicing affecting a novel domain in the *C. elegans* EGL-15 FGF receptor confers functional specificity. *Dev Camb Engl.* agosto de 2003;130(16):3757–66.
 26. Muriel JM, Dong C, Hutter H, Vogel BE. Fibulin-1C and Fibulin-1D splice variants have distinct functions and assemble in a hemicentin-dependent manner. *Dev Camb Engl.* outubro de 2005;132(19):4223–34.
 27. Demir E, Dickson BJ. fruitless splicing specifies male courtship behavior in *Drosophila*. *Cell.* 3 de junho de 2005;121(5):785–94.
 28. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet.* dezembro de 2008;40(12):1413–5.

29. Wang ET, Sandberg R, Luo S, Khrebtkova I, Zhang L, Mayr C, et al. Alternative Isoform Regulation in Human Tissue Transcriptomes. *Nature*. 27 de novembro de 2008;456(7221):470–6.
30. Lareau LF, Inada M, Green RE, Wengrod JC, Brenner SE. Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature*. 19 de abril de 2007;446(7138):926–9.
31. Faustino NA, Cooper TA. Pre-mRNA splicing and human disease. *Genes Dev*. 15 de fevereiro de 2003;17(4):419–37.
32. Espinoza M, de Silva R, Dickson DW, Davies P. Differential incorporation of tau isoforms in Alzheimer's disease. *J Alzheimers Dis JAD*. maio de 2008;14(1):1–16.
33. Kwok JBJ, Hallupp M, Loy CT, Chan DKY, Woo J, Mellick GD, et al. GSK3B polymorphisms alter transcription and splicing in Parkinson's disease. *Ann Neurol*. Dezembro de 2005;58(6):829–39.
34. Nakata K, Lipska BK, Hyde TM, Ye T, Newburn EN, Morita Y, et al. DISC1 splice variants are upregulated in schizophrenia and associated with risk polymorphisms. *Proc Natl Acad Sci*. 15 de setembro de 2009;106(37):15873–8.
35. Celotto AM, Graveley BR. Alternative splicing of the *Drosophila* Dscam pre-mRNA is both temporally and spatially regulated. *Genetics*. outubro de 2001;159(2):599–608.
36. Fagnani M, Barash Y, Ip JY, Misquitta C, Pan Q, Saltzman AL, et al. Functional coordination of alternative splicing in the mammalian central nervous system. *Genome Biol*. 12 de junho de 2007;8:R108.
37. Song JY, Ichtchenko K, Südhof TC, Brose N. Neuroligin 1 is a postsynaptic cell-adhesion molecule of excitatory synapses. *Proc Natl Acad Sci U S A*. 2 de fevereiro de 1999;96(3):1100–5.
38. Ushkaryov YA, Petrenko AG, Geppert M, Südhof TC. Neurexins: synaptic cell surface proteins related to the alpha-latrotoxin receptor and laminin. *Science*. 3 de julho de 1992;257(5066):50–6.
39. Ichtchenko K, Hata Y, Nguyen T, Ullrich B, Missler M, Moomaw C, et al. Neuroligin 1: a splice site-specific ligand for beta-neurexins. *Cell*. 5 de maio de 1995;81(3):435–43.
40. Ullrich B, Ushkaryov YA, Südhof TC. Cartography of neurexins: more than 1000 isoforms generated by alternative splicing and expressed in distinct subsets of neurons. *Neuron*. março de 1995;14(3):497–507.
41. Boucard AA, Chubykin AA, Comoletti D, Taylor P, Südhof TC. A splice code for trans-synaptic cell adhesion mediated by binding of neuroligin 1 to alpha- and beta-neurexins. *Neuron*. 20 de outubro de 2005;48(2):229–36.
42. Calarco JA, Superina S, O'Hanlon D, Gabut M, Raj B, Pan Q, et al. Regulation of vertebrate nervous system alternative splicing and development by an SR-related protein. *Cell*. 4 de setembro de 2009;138(5):898–910.

43. Quesnel-Vallières M, Irimia M, Cordes SP, Blencowe BJ. Essential roles for the splicing regulator nSR100/SRRM4 during nervous system development. *Genes Dev.* 1º de abril de 2015;29(7):746–59.
44. Raj B, Blencowe BJ. Alternative Splicing in the Mammalian Nervous System: Recent Insights into Mechanisms and Functional Roles. *Neuron.* 1º de julho de 2015;87(1):14–27.
45. Keppetipola N, Sharma S, Li Q, Black DL. Neuronal regulation of pre-mRNA splicing by polypyrimidine tract binding proteins, PTBP1 and PTBP2. *Crit Rev Biochem Mol Biol.* agosto de 2012;47(4):360–78.
46. Linares AJ, Lin C-H, Damianov A, Adams KL, Novitsch BG, Black DL. The splicing regulator PTBP1 controls the activity of the transcription factor Pbx1 during neuronal differentiation. *eLife.* 24 de dezembro de 2015;4:e09268.
47. Xue Y, Ouyang K, Huang J, Zhou Y, Ouyang H, Li H, et al. Direct conversion of fibroblasts to neurons by reprogramming PTB-regulated microRNA circuits. *Cell.* 17 de janeiro de 2013;152(1–2):82–96.
48. Yeo G, Holste D, Kreiman G, Burge CB. Variation in alternative splicing across human tissues. *Genome Biol.* 13 de setembro de 2004;5:R74.
49. Makeyev EV, Zhang J, Carrasco MA, Maniatis T. The MicroRNA miR-124 Promotes Neuronal Differentiation by Triggering Brain-Specific Alternative Pre-mRNA Splicing. *Mol Cell.* 3 de agosto de 2007;27(3):435–48.
50. Lipscombe D. Neuronal proteins custom designed by alternative splicing. *Curr Opin Neurobiol.* 1º de junho de 2005;15(3):358–63.
51. Luneau CJ, Williams JB, Marshall J, Levitan ES, Oliva C, Smith JS, et al. Alternative splicing contributes to K⁺ channel diversity in the mammalian central nervous system. *Proc Natl Acad Sci.* 1º de maio de 1991;88(9):3932–6.
52. Xie J, Black DL. A CaMK IV responsive RNA element mediates depolarization-induced alternative splicing of ion channels. *Nature.* 19 de abril de 2001;410(6831):936–9.
53. Merkin J, Russell C, Chen P, Burge CB. Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science.* 21 de dezembro de 2012;338(6114):1593–9.
54. Yeo GW, Nostrand EV, Holste D, Poggio T, Burge CB. Identification and analysis of alternative splicing events conserved in human and mouse. *Proc Natl Acad Sci.* 22 de fevereiro de 2005;102(8):2850–5.
55. Lin L, Shen S, Jiang P, Sato S, Davidson BL, Xing Y. Evolution of alternative splicing in primate brain transcriptomes. *Hum Mol Genet.* 1º de agosto de 2010;19(15):2958–73.
56. Semple BD, Blomgren K, Gimlin K, Ferriero DM, Noble-Haeusslein LJ. Brain development in rodents and humans: Identifying benchmarks of maturation and vulnerability to injury across species. *Prog Neurobiol.* 2013;0:1–16.

57. Richards LJ, Plachez C, Ren T. Mechanisms regulating the development of the corpus callosum and its agenesis in mouse and human. *Clin Genet.* outubro de 2004;66(4):276–89.
58. Oddo S, Caccamo A, Shepherd JD, Murphy MP, Golde TE, Kaye R, et al. Triple-transgenic model of Alzheimer's disease with plaques and tangles: intracellular Abeta and synaptic dysfunction. *Neuron.* 31 de julho de 2003;39(3):409–21.
59. Lin CH, Tallaksen-Greene S, Chien WM, Cearley JA, Jackson WS, Crouse AB, et al. Neurological abnormalities in a knock-in mouse model of Huntington's disease. *Hum Mol Genet.* 15 de janeiro de 2001;10(2):137–44.
60. Batzoglou S, Pachter L, Mesirov JP, Berger B, Lander ES. Human and Mouse Gene Structure: Comparative Analysis and Application to Exon Prediction. *Genome Res.* 1º de julho de 2000;10(7):950–8.
61. Strand AD, Aragaki AK, Baquet ZC, Hodges A, Cunningham P, Holmans P, et al. Conservation of Regional Gene Expression in Mouse and Human Brain. *PLOS Genet.* 20 de abril de 2007;3(4):e59.
62. Herculano-Houzel S, Watson C, Paxinos G. Distribution of neurons in functional areas of the mouse cerebral cortex reveals quantitatively different cortical zones. *Front Neuroanat* [Internet]. 21 de outubro de 2013 [citado 17 de setembro de 2016];7. Disponível em: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3800983/>
63. Treuting PM, Dintzis SM. *Comparative Anatomy and Histology: A Mouse and Human Atlas (Expert Consult)*. Academic Press; 2011. 474 p.
64. Koch C, Reid RC. Neuroscience: Observatories of the mind. *Nature.* 22 de março de 2012;483(7390):397–8.
65. Oberheim NA, Takano T, Han X, He W, Lin JHC, Wang F, et al. Uniquely Hominid Features of Adult Human Astrocytes. *J Neurosci.* 11 de março de 2009;29(10):3276–87.
66. Yue F, Cheng Y, Breschi A, Vierstra J, Wu W, Ryba T, et al. A comparative encyclopedia of DNA elements in the mouse genome. *Nature.* 20 de novembro de 2014;515(7527):355–64.
67. Meyer N, Richter N, Fan Z, Siemonsmeier G, Pivneva T, Jordan P, et al. Oligodendrocytes in the Mouse Corpus Callosum Maintain Axonal Function by Delivery of Glucose. *Cell Rep.* 27 de fevereiro de 2018;22(9):2383–94.
68. Choi HJ, Sun D, Jakobs TC. Isolation of intact astrocytes from the optic nerve head of adult mice. *Exp Eye Res.* 1º de agosto de 2015;137:103–10.
69. Lodovichi C, Belluscio L, Katz LC. Functional topography of connections linking mirror-symmetric maps in the mouse olfactory bulb. *Neuron.* 24 de abril de 2003;38(2):265–76.

70. Harper A. Mouse models of neurological disorders—A comparison of heritable and acquired traits. *Biochim Biophys Acta BBA - Mol Basis Dis.* outubro de 2010;1802(10):785–95.
71. Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet.* 13 de março de 2012;13(4):227–32.
72. Wang H, Qian W-J, Chin MH, Petyuk VA, Barry RC, Liu T, et al. Characterization of the Mouse Brain Proteome Using Global Proteomic Analysis Complemented with Cysteinyl-Peptide Enrichment. *J Proteome Res.* fevereiro de 2006;5(2):361–9.
73. Taraslia VK, Kouskoukis A, Anagnostopoulos AK, Stravopodis DJ, Margaritis LH, Tsangaris GT. Proteomic Analysis of Normal Murine Brain Parts. *Cancer Genomics - Proteomics.* 1º de maio de 2013;10(3):125–54.
74. Sharma K, Schmitt S, Bergner CG, Tyanova S, Kannaiyan N, Manrique-Hoyos N, et al. Cell type- and brain region-resolved mouse brain proteome. *Nat Neurosci.* 2 de novembro de 2015;18(12):1819–31.
75. Wang J, Gu Y, Wang L, Hang X, Gao Y, Wang H, et al. HUPPO BPP pilot study: A proteomics analysis of the mouse brain of different developmental stages. *PROTEOMICS.* 1º de novembro de 2007;7(21):4008–15.
76. Wasinger VC, Cordwell SJ, Cerpa-Poljak A, Yan JX, Gooley AA, Wilkins MR, et al. Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium*. *Electrophoresis.* julho de 1995;16(7):1090–4.
77. Rabilloud T. *Proteome Research: Two-Dimensional Gel Electrophoresis and Identification Methods.* Springer Science & Business Media; 2013. 256 p.
78. Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature.* 13 de março de 2003;422(6928):198–207.
79. Minjarez B, Calderón-González KG, Rustarazo MLV, Herrera-Aguirre ME, Labra-Barrios ML, Rincon-Limas DE, et al. Identification of proteins that are differentially expressed in brains with Alzheimer’s disease using iTRAQ labeling and tandem mass spectrometry. *J Proteomics.* 29 de abril de 2016;139:103–21.
80. Núñez EV, Guest PC, Martins-de-Souza D, Domont GB, Nogueira FCS. Application of iTRAQ Shotgun Proteomics for Measurement of Brain Proteins in Studies of Psychiatric Disorders. In: *Proteomic Methods in Neuropsychiatric Research* [Internet]. Springer, Cham; 2017 [citado 12 de março de 2018]. p. 219–27. (Advances in Experimental Medicine and Biology). Disponível em: https://link.springer.com/chapter/10.1007/978-3-319-52479-5_18
81. Sarkis GA, Mangaonkar MD, Moghieb A, Lelling B, Guertin M, Yadikar H, et al. The Application of Proteomics to Traumatic Brain and Spinal Cord Injuries. *Curr Neurol Neurosci Rep.* 1º de março de 2017;17(3):23.
82. Mertins P, Mani DR, Ruggles KV, Gillette MA, Clauser KR, Wang P, et al. Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature.* junho de 2016;534(7605):55–62.

83. Huang Z, Ma L, Huang C, Li Q, Nice EC. Proteomic profiling of human plasma for cancer biomarker discovery. *PROTEOMICS*. 1º de março de 2017;17(6):n/a-n/a.
84. Rolland DCM, Basrur V, Jeon Y-K, McNeil-Schwalm C, Fermin D, Conlon KP, et al. Functional proteogenomics reveals biomarkers and therapeutic targets in lymphomas. *Proc Natl Acad Sci*. 20 de junho de 2017;114(25):6581–6.
85. Neverova I, Van Eyk JE. Role of chromatographic techniques in proteomic analysis. *J Chromatogr B*. 5 de fevereiro de 2005;815(1–2):51–63.
86. Parker CE, Warren MR, Mocanu V. Mass Spectrometry for Proteomics. In: Alzate O, organizador. *Neuroproteomics* [Internet]. Boca Raton (FL): CRC Press/Taylor & Francis; 2010 [citado 24 de janeiro de 2018]. (Frontiers in Neuroscience). Disponível em: <http://www.ncbi.nlm.nih.gov/books/NBK56011/>
87. Steen H, Mann M. The abc's (and xyz's) of peptide sequencing. *Nat Rev Mol Cell Biol*. Setembro de 2004;5(9):699–711.
88. Verheggen K, Raeder H, Berven FS, Martens L, Barsnes H, Vaudel M. Anatomy and evolution of database search engines—a central component of mass spectrometry based proteomic workflows. *Mass Spectrom Rev* [Internet]. 13 de setembro de 2017 [citado 16 de janeiro de 2018]; Disponível em: <http://doi.wiley.com/10.1002/mas.21543>
89. UniProt: a hub for protein information. *Nucleic Acids Res*. 28 de janeiro de 2015;43(Database issue):D204–12.
90. Pruitt KD, Brown GR, Hiatt SM, Thibaud-Nissen F, Astashyn A, Ermolaeva O, et al. RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res*. 1º de janeiro de 2014;42(Database issue):D756–63.
91. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, et al. GENCODE: The reference human genome annotation for The ENCODE Project. *Genome Res*. setembro de 2012;22(9):1760–74.
92. Nesvizhskii AI. A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. *J Proteomics*. 10 de outubro de 2010;73(11):2092–123.
93. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B Methodol*. 1995;57(1):289–300.
94. Empirical Bayes Analysis of a Microarray Experiment. *J Am Stat Assoc*. 2001;10.
95. Käll L, Storey JD, MacCoss MJ, Noble WS. Posterior Error Probabilities and False Discovery Rates: Two Sides of the Same Coin. *J Proteome Res*. janeiro de 2008;7(1):40–4.
96. Bhattacharya DA. Proteotypic Peptides. In: Dubitzky W, Wolkenhauer O, Cho K-H, Yokota H, organizadores. *Encyclopedia of Systems Biology* [Internet]. Springer New York; 2013 [citado 7 de dezembro de 2017]. p. 1800–1800.

Disponível em: http://link.springer.com/referenceworkentry/10.1007/978-1-4419-9863-7_1475

97. Pathan M, Samuel M, Keerthikumar S, Mathivanan S. Unassigned MS/MS Spectra: Who Am I? In: Proteome Bioinformatics [Internet]. Humana Press, New York, NY; 2017 [citado 5 de dezembro de 2017]. p. 67–74. (Methods in Molecular Biology). Disponível em: https://link.springer.com/protocol/10.1007/978-1-4939-6740-7_6
98. Nesvizhskii AI. Proteogenomics: concepts, applications and computational strategies. *Nat Methods*. 1º de novembro de 2014;11(11):1114–25.
99. Jaffe JD, Berg HC, Church GM. Proteogenomic mapping as a complementary method to perform genome annotation. *Proteomics*. janeiro de 2004;4(1):59–77.
100. Sheynkman GM, Shortreed MR, Cesnik AJ, Smith LM. Proteogenomics: Integrating Next-Generation Sequencing and Mass Spectrometry to Characterize Human Proteomic Variation. *Annu Rev Anal Chem Palo Alto Calif*. 12 de junho de 2016;9(1):521–45.
101. Wu CH. Integrative Bioinformatics for Genomics and Proteomics. *J Biomol Tech JBT*. outubro de 2011;22(Suppl):S4.
102. O’Leary NA, Wright MW, Brister JR, Ciuffo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*. 4 de janeiro de 2016;44(Database issue):D733–45.
103. Boguski MS, Lowe TM, Tolstoshev CM. dbEST--database for “expressed sequence tags”. *Nat Genet*. agosto de 1993;4(4):332–3.
104. Ng PC, Kirkness EF. Whole genome sequencing. *Methods Mol Biol Clifton NJ*. 2010;628:215–26.
105. Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, et al. Targeted Capture and Massively Parallel Sequencing of Twelve Human Exomes. *Nature*. 10 de setembro de 2009;461(7261):272–6.
106. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. janeiro de 2009;10(1):57–63.
107. Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS. Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. *Science*. 10 de abril de 2009;324(5924):218–23.
108. Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet TIG*. junho de 2000;16(6):276–7.
109. Stajich JE, Block D, Boulez K, Brenner SE, Chervitz SA, Dagdigian C, et al. The Bioperl Toolkit: Perl Modules for the Life Sciences. *Genome Res*. outubro de 2002;12(10):1611–8.

110. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*. 1º de junho de 2009;25(11):1422–3.
111. Li J, Su Z, Ma Z-Q, Slebos RJC, Halvey P, Tabb DL, et al. A Bioinformatics Workflow for Variant Peptide Detection in Shotgun Proteomics. *Mol Cell Proteomics MCP* [Internet]. maio de 2011 [citado 16 de março de 2018];10(5). Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3098595/>
112. Menon R, Omenn GS. Identification of Alternatively Spliced Transcripts using a Proteomic Informatics Approach. *Methods Mol Biol Clifton NJ*. 2011;696:319–26.
113. Tavares R, de Miranda Scherer N, Pauletti BA, Araújo E, Folador EL, Espindola G, et al. SpliceProt: A protein sequence repository of predicted human splice variants. *PROTEOMICS*. fevereiro de 2014;14(2–3):181–5.
114. Frenkel-Morgenstern M, Gorohovski A, Vucenovic D, Maestre L, Valencia A. ChiTaRS 2.1—an improved database of the chimeric transcripts and RNA-seq data with novel sense–antisense chimeric RNA transcripts. *Nucleic Acids Res*. 28 de janeiro de 2015;43(Database issue):D68–75.
115. Evans VC, Barker G, Heesom KJ, Fan J, Bessant C, Matthews DA. De novo derivation of proteomes from transcriptomes for transcript and protein identification. *Nat Methods*. dezembro de 2012;9(12):1207–11.
116. Castellana NE, Pham V, Arnott D, Lill JR, Bafna V. Template Proteogenomics: Sequencing Whole Proteins Using an Imperfect Database. *Mol Cell Proteomics*. junho de 2010;9(6):1260–70.
117. Helmy M, Sugiyama N, Tomita M, Ishihama Y. Mass spectrum sequential subtraction speeds up searching large peptide MS/MS spectra datasets against large nucleotide databases for proteogenomics. *Genes Cells*. agosto de 2012;17(8):633–44.
118. Ivanov MV, Lobas AA, Karpov DS, Moshkovskii SA, Gorshkov MV. Comparison of False Discovery Rate Control Strategies for Variant Peptide Identifications in Shotgun Proteogenomics. *J Proteome Res*. 5 de maio de 2017;16(5):1936–43.
119. Tanner S, Shen Z, Ng J, Florea L, Guigó R, Briggs SP, et al. Improving gene annotation using peptide mass spectrometry. *Genome Res*. 1º de fevereiro de 2007;17(2):231–9.
120. Tavares R, Scherer NM, Ferreira CG, Costa FF, Passetti F. Splice variants in the proteome: a promising and challenging field to targeted drug discovery. *Drug Discov Today*. março de 2015;20(3):353–60.
121. Tavares R, Wajnberg G, Scherer N de M, Pauletti BA, Cassoli JS, Ferreira CG, et al. Unveiling alternative splice diversity from human oligodendrocyte proteome data. *J Proteomics* [Internet]. maio de 2016 [citado 16 de junho de 2016]; Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S1874391916302160>

122. Chapman Brett, Bellgard Matthew. Plant Proteogenomics: Improvements to the Grapevine Genome Annotation. *PROTEOMICS*. 7 de novembro de 2017;17(21):1700197.
123. Fabre Bertrand, Korona Dagmara, Mata Clara I., Parsons Harriet T., Deery Michael J., Hertog Maarten L. A. T. M., et al. Spectral Libraries for SWATH-MS Assays for *Drosophila melanogaster* and *Solanum lycopersicum*. *PROTEOMICS*. 7 de novembro de 2017;17(21):1700216.
124. Proffitt JM, Glenn J, Cesnik AJ, Jadhav A, Shortreed MR, Smith LM, et al. Proteomics in non-human primates: utilizing RNA-Seq data to improve protein identification by mass spectrometry in vervet monkeys. *BMC Genomics* [Internet]. 13 de novembro de 2017 [citado 2 de dezembro de 2017];18. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5683380/>
125. Palomino-Alonso M, Lachén-Montes M, González-Morales A, Ausín K, Pérez-Mediavilla A, Fernández-Irigoyen J, et al. Network-Driven Proteogenomics Unveils an Aging-Related Imbalance in the Olfactory I κ B α -NF κ B p65 Complex Functionality in Tg2576 Alzheimer's Disease Mouse Model. *Int J Mol Sci*. 27 de outubro de 2017;18(11):2260.
126. Karolchik D, Baertsch R, Diekhans M, Furey TS, Hinrichs A, Lu YT, et al. The UCSC Genome Browser Database. *Nucleic Acids Res*. 1º de janeiro de 2003;31(1):51–4.
127. Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*. 1º de janeiro de 2006;34(Database issue):D173-180.
128. Hermjakob H, Apweiler R. The Proteomics Identifications Database (PRIDE) and the ProteomExchange Consortium: making proteomics data accessible. *Expert Rev Proteomics*. fevereiro de 2006;3(1):1–3.
129. Martins-de-Souza D, Carvalho PC, Schmitt A, Junqueira M, Nogueira FCS, Turck CW, et al. Deciphering the Human Brain Proteome: Characterization of the Anterior Temporal Lobe and Corpus Callosum As Part of the Chromosome 15-centric Human Proteome Project. *J Proteome Res*. 3 de janeiro de 2014;13(1):147–57.
130. Zhang P, Karani R, Turner RL, Dufresne C, Ferri S, Van Eyk JE, et al. The proteome of normal human retrobulbar optic nerve and sclera. *PROTEOMICS*. 1º de outubro de 2016;16(19):2592–6.
131. Dammalli M, Dey G, Madugundu AK, Kumar M, Rodrigues B, Gowda H, et al. Proteomic Analysis of the Human Olfactory Bulb. *OMICS J Integr Biol*. agosto de 2017;21(8):440–53.
132. Passeti F, De Oliveira P, Farah J, Dobroff V, Garcia M, Pereira C, et al. Ternary Matrix Adapter and Retrieval Method for Molecular Biological Information [Internet]. WO/2008/138087, 2008 [citado 27 de março de 2018]. Disponível em: <https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2008138087>

133. SILVA RTD. Desenvolvimento de uma abordagem computacional para a tradução in silico de variantes de splicing detectadas no transcriptoma humano. :82.
134. Kent WJ. BLAT—The BLAST-Like Alignment Tool. *Genome Res.* 1º de abril de 2002;12(4):656–64.
135. EMBOSS Credits [Internet]. [citado 18 de agosto de 2017]. Disponível em: <http://emboss.sourceforge.net/credits/>
136. Carvalho PC, Lima DB, Leprevost FV, Santos MDM, Fischer JSG, Aquino PF, et al. Integrated analysis of shotgun proteomic data with PatternLab for proteomics 4.0. *Nat Protoc.* 10 de dezembro de 2015;11(1):102–17.
137. Zhang S-R, Shan Y-C, Jiang H, Liu J-H, Zhou Y, Zhang L-H, et al. The Null-Test for peptide identification algorithm in Shotgun proteomics. *J Proteomics.* 23 de junho de 2017;163:118–25.
138. Eng JK, Jahan TA, Hoopmann MR. Comet: an open-source MS/MS sequence database search tool. *Proteomics.* janeiro de 2013;13(1):22–4.
139. Eng JK, Hoopmann MR, Jahan TA, Egertson JD, Noble WS, MacCoss MJ. A deeper look into Comet – implementation and features. *J Am Soc Mass Spectrom.* novembro de 2015;26(11):1865–74.
140. Carvalho PC, Fischer JSG, Xu T, Cociorva D, Balbuena TS, Valente RH, et al. Search engine processor: Filtering and organizing peptide spectrum matches. *PROTEOMICS.* abril de 2012;12(7):944–9.
141. What is the canonical sequence? Are all isoforms described in one entry? [Internet]. [citado 6 de maio de 2018]. Disponível em: https://www.uniprot.org/help/canonical_and_isoforms
142. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The Human Genome Browser at UCSC. *Genome Res.* 1º de junho de 2002;12(6):996–1006.
143. Xiong Y, Lei Q-Y, Zhao S, Guan K-L. Regulation of glycolysis and gluconeogenesis by acetylation of PKM and PEPCK. *Cold Spring Harb Symp Quant Biol.* 2011;76:285–9.
144. Noguchi T, Inoue H, Tanaka T. The M1- and M2-type isozymes of rat pyruvate kinase are produced from the same gene by alternative RNA splicing. *J Biol Chem.* 15 de outubro de 1986;261(29):13807–12.
145. Konno M, Ishii H, Koseki J, Tanuma N, Nishida N, Kawamoto K, et al. Pyruvate kinase M2, but not M1, allele maintains immature metabolic states of murine embryonic stem cells. *Regen Ther.* 1º de junho de 2015;1:63–71.
146. Mazurek S, Boschek CB, Hugo F, Eigenbrodt E. Pyruvate kinase type M2 and its role in tumor growth and spreading. *Semin Cancer Biol.* 1º de agosto de 2005;15(4):300–8.

147. Bomsztyk K, Denisenko O, Ostrowski J. hnRNP K: One protein multiple processes. *BioEssays*. 1º de junho de 2004;26(6):629–38.
148. Moumen A, Masterson P, O'Connor MJ, Jackson SP. hnRNP K: an HDM2 target and transcriptional coactivator of p53 in response to DNA damage. *Cell*. 16 de dezembro de 2005;123(6):1065–78.
149. Expert-Bezançon A, Le Caer JP, Marie J. Heterogeneous nuclear ribonucleoprotein (hnRNP) K is a component of an intronic splicing enhancer complex that activates the splicing of the alternative exon 6A from chicken beta-tropomyosin pre-mRNA. *J Biol Chem*. 10 de maio de 2002;277(19):16614–23.
150. Au PYB, You J, Caluseriu O, Schwartzenruber J, Majewski J, Bernier FP, et al. GeneMatcher Aids in the Identification of a New Malformation Syndrome with Intellectual Disability, Unique Facial Dysmorphisms, and Skeletal and Connective Tissue Abnormalities Caused by De Novo Variants in HNRNPK. *Hum Mutat*. outubro de 2015;36(10):1009–14.
151. Hamajima N, Matsuda K, Sakata S, Tamaki N, Sasaki M, Nonaka M. A novel gene family defined by human dihydropyrimidinase and three related proteins with differential tissue distribution. *Gene*. 1º de janeiro de 1996;180(1):157–63.
152. Minturn JE, Fryer HJ, Geschwind DH, Hockfield S. TOAD-64, a gene expressed early in neuronal differentiation in the rat, is related to unc-33, a *C. elegans* gene involved in axon outgrowth. *J Neurosci Off J Soc Neurosci*. outubro de 1995;15(10):6757–66.
153. Goshima Y, Nakamura F, Strittmatter P, Strittmatter SM. Collapsin-induced growth cone collapse mediated by an intracellular protein related to UNC-33. *Nature*. 10 de agosto de 1995;376(6540):509–14.
154. Swanson DA, Steel JM, Valle D. Identification and Characterization of the Human Ortholog of Rat STXBP1, a Protein Implicated in Vesicle Trafficking and Neurotransmitter Release. *Genomics*. 15 de março de 1998;48(3):373–6.
155. Saitsu H, Kato M, Mizuguchi T, Hamada K, Osaka H, Tohyama J, et al. *De novo* mutations in the gene encoding STXBP1 (MUNC18-1) cause early infantile epileptic encephalopathy. *Nat Genet*. junho de 2008;40(6):782–8.
156. Grone BP, Marchese M, Hamling KR, Kumar MG, Krasniak CS, Sicca F, et al. Epilepsy, Behavioral Abnormalities, and Physiological Comorbidities in Syntaxin-Binding Protein 1 (STXBP1) Mutant Zebrafish. *PLOS ONE*. 3 de outubro de 2016;11(3):e0151148.
157. Behan Á, Byrne C, Dunn MJ, Cagney G, Cotter DR. Proteomic analysis of membrane microdomain-associated proteins in the dorsolateral prefrontal cortex in schizophrenia and bipolar disorder reveals alterations in LAMP, STXBP1 and BASP1 protein expression. *Mol Psychiatry*. junho de 2009;14(6):601–13.
158. Furuno T, Ito A, Koma Y-I, Watabe K, Yokozaki H, Bienenstock J, et al. The spermatogenic Ig superfamily/synaptic cell adhesion molecule mast-cell adhesion molecule promotes interaction with nerves. *J Immunol Baltim Md 1950*. 1º de junho de 2005;174(11):6934–42.

159. Hung AY, Haass C, Nitsch RM, Qiu WQ, Citron M, Wurtman RJ, et al. Activation of protein kinase C inhibits cellular production of the amyloid beta-protein. *J Biol Chem*. 5 de novembro de 1993;268(31):22959–62.
160. Yamasaki T, Takahashi A, Pan J, Yamaguchi N, Yokoyama KK. Phosphorylation of Activation Transcription Factor-2 at Serine 121 by Protein Kinase C Controls c-Jun-mediated Activation of Transcription. *J Biol Chem*. 27 de março de 2009;284(13):8567–81.
161. Hamacher M, Meyer HE. HUPO Brain Proteome Project: aims and needs in proteomics. *Expert Rev Proteomics*. 1º de janeiro de 2005;2(1):1–3.