



PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ
ESCOLA DE MEDICINA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS DA SAÚDE

ADRIANO FERRASA

**MODIFICAÇÕES TRANSCRICIONAIS NA REGULAÇÃO DE RETROELEMENTOS L1
EM DISTÚRBIOS DO NEURODESENVOLVIMENTO**

CURITIBA

2020

ADRIANO FERRASA

**MODIFICAÇÕES TRANSCRICIONAIS NA REGULAÇÃO DE RETROELEMENTOS L1 EM
DISTÚRBIOS DO NEURODESENVOLVIMENTO**

Tese apresentada ao Programa de Pós-Graduação em Ciências da Saúde, da Escola de Medicina da Pontifícia Universidade Católica do Paraná, como requisito parcial à obtenção do título de Doutor em Ciências da Saúde.

Orientador: Dr. Roberto Hirochi Herai

CURITIBA

2020

Dados da Catalogação na Publicação
Pontifícia Universidade Católica do Paraná
Sistema Integrado de Bibliotecas – SIBI/PUCPR
Biblioteca Central
Luci Eduarda Wielganczuk – CRB 9/1118

Ferrasa, Adriano
F378m Modificações transcricionais na regulação de retroelementos L1 em
2020 em distúrbios do neurodesenvolvimento / Adriano Ferrasa ; orientador: Roberto
Hirochi Herai. – 2020.
182 f. : il. ; 30 cm

Tese (doutorado) – Pontifícia Universidade Católica do Paraná, Curitiba,
2020
Bibliografia: f. 89-111

1. Transtornos do neurodesenvolvimento. 2. Retroelementos. 3. Elementos
de nucleotídeos longos e dispersos. 4. Expressão gênica. 5. Autofagia. I. Herai,
Roberto Hiroshi. II. Pontifícia Universidade Católica do Paraná. Programa de
Pós-Graduação em Ciências da Saúde. III. Título

CDD 20 ed. 616.8588



Pontifícia Universidade Católica do Paraná

Escola de Medicina

Programa de Pós-Graduação em Ciências da Saúde

ATA DA SESSÃO PÚBLICA DE EXAME DE TESE DO PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS DA SAÚDE EM NÍVEL DE DOUTORADO DA PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ.

Aos **26 dias do mês de novembro de 2020 às 13:30**, realizou-se a sessão aberta de Defesa de Tese **“Modificações Transcrpcionais Na Regulação De Retroelementos L1 Em Distúrbios Do Neurodesenvolvimento”** apresentado por **Adriano Ferrasa** para obtenção do título de Doutor; Área de concentração: Medicina e áreas afins.

A banca examinadora foi composta pelos seguintes membros:

MEMBROS DA BANCA	ASSINATURA
Prof. Dr. Roberto Hirochi Herai - Presidente	
Prof. Dr. Carlos Eduardo Coral de Oliveira (PUCPR)	
Prof. Dr. Cleber Machado de Souza (PUCPR)	
Prof. Dr. Michel Eduardo Beleza Yamagishi (EMBRAPA)	
Prof. Dr. Rafael Mazer Etto (UEPG)	

De acordo com as normas regimentais a Banca Examinadora deliberou sobre os conceitos a serem distribuídos e que foram os seguintes:

Prof. Dr. Roberto Hirochi Herai

Conceito: APROVADO

Prof. Dr. Carlos Eduardo Coral de Oliveira

Conceito: APROVADO

Prof. Dr. Cleber Machado de Souza

Conceito: APROVADO

Prof. Dr. Michel Eduardo Beleza Yamagishi

Conceito: APROVADO

Prof. Dr. Rafael Mazer Etto

Conceito: APROVADO

Parecer Final: APROVADO

Observações da Banca Examinadora:

Prof. Dr. Roberto Hirochi Herai
Presidente da Banca Examinadora

Profa. Dra. Cristina Pellegrino Baena
Coordenadora do PPGCS-PUCPR

AGRADECIMENTOS

Primeiramente, agradeço ao corpo docente e funcionários do Programa de Pós Graduação em Ciências da Saúde (PPGCS) da Escola de Medicina da PUCPR pela dedicação em oferecer disciplinas, cursos e eventos de alta qualidade e que proporcionaram trocas de informações e novos conhecimentos.

Agradeço a Fundação Araucária e a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa de doutorado.

Ao pesquisador Alysso Muotri, coordenador do Laboratório de Células Tronco da Universidade da Califórnia San Diego (UCSD, USA), pela colaboração essencial neste trabalho.

Agradeço aos meus colegas do Laboratório Experimental Multiusuário, especialmente do laboratório de bioinformática da PUCPR pelos momentos de descontração, discussões e aprendizados.

Ao meu orientador e grande amigo, Prof. Dr. Roberto Herai, sou muito grato por todas as discussões ricas de conhecimento, apoio nos momentos difíceis, paciência e principalmente, a confiança em mim depositada ao longo da elaboração e condução da tese.

Por fim, agradeço imensamente minha família, especialmente à minha querida esposa Ana, por todo o apoio durante a execução desse trabalho.

RESUMO

Introdução: L1 (*Long interspersed elements class 1*) são retroelementos genéticos móveis autônomos e ativos que se integram e alteram o conteúdo do genoma humano. Apresentam alta atividade em células cerebrais e estão envolvidos na modulação de expressão de genes relacionados com o desenvolvimento e plasticidade cerebral. Desta forma, a instabilidade na atividade de L1 pode ocasionar o surgimento de distúrbios do neurodesenvolvimento (DN), como transtornos do espectro do autismo (ASD), transtorno bipolar (BPD), síndrome de Rett (RTT) e esquizofrenia (SCZ). Entretanto, ainda não está claro o modo como a regulação de expressão de retroelementos L1 ocorre a nível transcricional em indivíduos portadores destes DN. **Objetivo:** Avaliar a expressão de genes que regulam a expressão e atividade de retroelementos do tipo L1 em dados de transcriptoma oriundos de amostras celulares de indivíduos com DN. **Metodologia:** Para identificar os genes envolvidos na regulação da atividade e expressão de L1, foi realizada uma revisão sistemática da literatura (RSL) para criar uma base de genes reguladores de L1. Em seguida, através de uma extensa busca em bases de dados de transcriptoma público, foram obtidos 729 amostras de diferentes tipos celulares cerebrais oriundas de pacientes com ASD, BPD, RTT e SCZ. Nestas amostras, foi realizada a análise de expressão de transcritos L1 com a utilização da base de anotações de L1 desenvolvida neste trabalho. Posteriormente, por meio de uma análise de genes diferencialmente expressos (GDE), foi explorado a presença e expressão de genes reguladores de L1. Os GDE foram também submetidos a análise de enriquecimento de vias relacionadas a processos biológicos, utilizando genes pré-classificados com base na variação de expressão entre amostras controle e de DN. **Resultados:** Para aprimorar a investigação de elementos genéticos móveis, foi criada uma base de anotações de L1s para o genoma humano, a qual contém uma ampla caracterização de retroelementos L1 completos, incluindo anotações de sequência baseadas em identidade e similaridade de cada uma das partes desta molécula. Esta base foi então utilizada para investigar a expressão de transcritos L1 em amostras celulares de DN. Em todas as amostras, incluindo controles e afetados, transcritos L1 apresentaram alta expressão. Posteriormente, com a utilização da base de reguladores L1, composta por 158 genes, foram observados diversos reguladores com alteração significativa de expressão entre as amostras de DN em relação aos respectivos controles. A alteração de expressão de L1, detectada em células cerebrais, sugere a ocorrência de uma alteração na atividade de retrotransposição nestas células. Ainda, para cada tipo celular e DN, foram identificados diversos processos biológicos nos quais os genes reguladores atuam, onde incluem-se vias associadas com processos de reparo de DNA, sinalização de interferon alfa/beta, senescência celular, biogênese de piRNAs e mitofagia. Também foi hipotetizado que o gene *AMBRA1* pode atuar como um novo regulador da expressão e atividade de L1 em células cerebrais humanas. Este gene está associado com o desenvolvimento do sistema nervoso central, neurogênese e também está relacionado com condições patológicas cerebrais. **Conclusão:** Este trabalho permitiu a construção de novas ferramentas para o estudo de retroelementos do tipo L1 por meio da base de anotações L1 e também da base de genes reguladores L1. Também, por meio da utilização destas ferramentas, foi demonstrado que a expressão de retroelementos L1 e seus respectivos

reguladores encontram-se alterados em células cerebrais de indivíduos com DN. Outra contribuição significativa deste trabalho foi a identificação do gene *AMBRA1*, que pela primeira vez, é descrito como um potencial regulador da atividade de L1. Por fim, os resultados obtidos servirão como base para outras pesquisas, de forma a expandir ainda mais o conhecimento sobre a regulação de L1s e desvendar as interações e correlações funcionais entre a biologia de L1s e de DN.

Palavras-chave: L1, retroelementos, distúrbios do neurodesenvolvimento, transcriptoma humano, expressão gênica, *AMBRA1*, autofagia.

ABSTRACT

Introduction: L1 (Long interspersed elements class 1) are active mobile genetic retroelements that integrate and alter the content of the human genome. They have high activity in brain cells and are involved in modulating the expression of genes related to brain development and plasticity. Thus, instability in L1 activity can cause neurodevelopmental disorders (DN), such as autism spectrum disorders (ASD), bipolar disorder (BPD), Rett syndrome (RTT) and schizophrenia (SCZ). However, it is still unclear how the regulation of expression of L1 retroelements occurs at the transcriptional level in individuals with these DN. **Objective:** To evaluate the expression of genes that regulate the expression and activity of L1 retroelements in transcriptome data from cell samples of individuals with DN. **Methods:** To identify the genes involved in the regulation of L1 activity and expression, a systematic literature review (RSL) was performed to create an atlas of L1 regulatory genes. Then, through an extensive search in public transcriptome databases, 729 samples of different brain cell types were obtained from patients with ASD, BPD, RTT and SCZ. In these samples, the expression analysis of L1 transcripts was performed using the L1 annotation base developed in this work. Subsequently, through an analysis of differentially expressed genes (DEG), the presence and expression of L1 regulatory genes were explored. The DEGs were also submitted to pathways enrichment analysis related to biological processes, using pre-classified genes based on the variation of expression between control and DN samples. **Results:** To improve the investigation of mobile genetic elements, an atlas of L1 annotations for the human genome was created, which contains a wide characterization of complete L1 retroelements, including sequence annotations based on the identity and similarity of each part of this molecule. This atlas was then used to investigate the expression of L1 transcripts in DN cell samples. In all samples, including controls and affected, L1 transcripts showed high expression. Subsequently, with the use of the L1 regulatory base, composed of 158 genes, several regulators were observed with significant alteration of expression between the DN samples in relation to the respective controls. The alteration of L1 expression detected in brain cells, suggests the occurrence of high retrotransposition activity in these cells. Still, for each cell type and DN, several biological processes were identified in which the regulatory genes act, including pathways associated with DNA repair processes, alpha/beta interferon signaling, cell senescence, piRNA biogenesis and mitophagy. It has also been hypothesized that the AMBRA1 gene may act as a new regulator of L1 expression and activity in human brain cells. This gene is associated with the development of the central nervous system, neurogenesis and is also related to pathological brain conditions. **Conclusion:** This work allowed the construction of new tools for the study of L1 type retroelements using the L1 annotation atlas and the L1 regulatory gene base. Also, using these tools, it was demonstrated that the expression of L1 retroelements and their respective regulators are altered in brain cells of individuals with DN. Another significant contribution of this work was the identification of the AMBRA1 gene, which, for the first time, is described as a potential regulator of L1 activity. Finally, the results obtained will serve as a basis for further

research, to expand knowledge about the regulation of L1s and to unravel the interactions and functional correlations between the biology of L1s and DN.

Keywords: L1, retroelements, neurodevelopmental disorders, human transcriptome, gene expression, *AMBRA1*, autophagy.

LISTA DE FIGURAS

Figura 1: Biogênese de L1.	18
Figura 2: Anotação de subfamílias de retroelementos L1 fornecidos pela ferramenta RepeatMasker.	20
Figura 3: Genes reguladores de L1.	28
Figura 4: Fluxograma de criação da base de anotações para retroelementos L1.	32
Figura 5: Diferenciação neural de iPSC não é afetada pela reprogramação celular.	34
Figura 6: Número de cópias por cromossomo de regiões e retroelementos L1 completos.	42
Figura 7: Distribuição por cromossomo de regiões pertencentes à subfamília L1HS entre genomas individuais e de referência.	44
Figura 8: Comparação entre a base de anotações de regiões L1 e anotações de L1 reportados pelo RepeatMasker.	46
Figura 9: Expressão de L1 total e L1HS em DN e tipos celulares.	48
Figura 10: Expressão de L1HS em amostras controle e de DN.	49
Figura 11: Reguladores L1 com expressão alterada na via da anemia de Fanconi.	51
Figura 12: Reguladores L1 com expressão alterada na via de reparo de DNA por excisão de nucleotídeos.	52
Figura 13: Reguladores L1 com expressão alterada na via de reparo de DNA por recombinação homóloga.	53
Figura 14: Reguladores L1 com expressão alterada na via da união de extremidade não homóloga.	54
Figura 15: Reguladores L1 com expressão alterada na via de senescência celular.	55
Figura 16: Reguladores L1 com expressão alterada na via de mitofagia.	56
Figura 17: Reguladores L1 com expressão alterada na via da biogênese de piRNAs.	57
Figura 18: Reguladores L1 com expressão alterada na via de sinalização de interferon alfa/beta.	58
Figura 19: Padrão de expressão de moléculas reguladoras de L1 em DN para células provenientes de tecidos cerebrais <i>post mortem</i>	59

Figura 20: Reguladores L1 com mesmo perfil de expressão em ASD, BPD, RTT e SCZ.	60
Figura 21: Reguladores L1 em ASD participantes da via da anemia de Fanconi.	61
Figura 22: Reguladores L1 em ASD participantes da via de reparo de DNA por excisão de nucleotídeos.	61
Figura 23: Reguladores L1 em BPD participantes da via da biogênese de piRNAs.....	62
Figura 24: Padrão de expressão de moléculas reguladores de L1 em iPSC de DN.....	63
Figura 25: Reguladores L1 em RTT participantes da via da anemia de Fanconi.	63
Figura 26: Reguladores L1 em ASD participantes da via da biogênese de piRNAs.....	64
Figura 27: Padrão de expressão de moléculas reguladores de L1 em DN para NPC. ...	65
Figura 28: Reguladores L1 em SCZ participantes da via de reparo de DNA por recombinação homóloga.	65
Figura 29: Reguladores L1 em ASD participantes da via de mitofagia.	66
Figura 30: Reguladores L1 em ASD participantes da via da biogênese de piRNAs.....	66
Figura 31: Padrão de expressão de moléculas reguladores de L1 em DN para neurônios provenientes de cultura <i>in vitro</i>	67
Figura 32: Reguladores L1 em RTT e SCZ participantes da via de reparo da união de extremidade não-homóloga.	68
Figura 33: Reguladores L1 em RTT e SCZ participantes da via de sinalização de interferon alfa/beta.	68
Figura 34: Reguladores L1 em ASD e SCZ participantes da via de senescência celular.	69
Figura 35: Reguladores L1 em SCZ participantes da via de mitofagia.	70
Figura 36: Reguladores L1 em RTT participantes da via da anemia da Fanconi.	70
Figura 37: Reguladores L1 em RTT participantes da via de mitofagia.	71
Figura 38: Reguladores L1 em RTT participantes da via da união de extremidade não- homóloga.	71
Figura 39: Via de autofagia em células humanas.	75
Figura 40: Correlação de expressão dos genes <i>AMBRA1</i> e <i>BECN1</i> em células cerebrais e não cerebrais.	76
Figura 41: Regulação de L1 via processo autofágico.	79

Figura 42: Fluxograma de identificação e seleção de artigos para a RSL.	114
--	-----

LISTA DE TABELAS

Tabela 1: Descrição de amostras com mutações pontuais no gene <i>MECP2</i>	33
Tabela 2: Estratégia de busca utilizada para obtenção de amostras de pacientes com DN.	36
Tabela 3: Amostras RNA-seq obtidas através de bases de dados públicas e em colaboração.	36
Tabela 4: Parâmetros de busca utilizados para levantamento de moléculas reguladoras de retroelementos L1.	113
Tabela 5: Genes reguladores de retroelementos L1.	129
Tabela 6: Contagem de regiões L1HS a nível alélico para o genoma individual Caucasiano.	173
Tabela 7: Expressão de genes reguladores de L1 em PO de ASD, BPD, RTT e SCZ.	174
Tabela 8: Expressão de genes reguladores de L1 em iPSC de ASD e RTT.	177
Tabela 9: Expressão de genes reguladores de L1 em NE de ASD, BPD, RTT e SCZ.	178
Tabela 10: Expressão de genes reguladores de L1 em NPC de ASD, RTT e SCZ.	181

LISTA DE SIGLAS

3'UTR	Região não traduzida 3', do inglês, <i>3' Untranslated Region</i>
5'UTR	Região não traduzida 5', do inglês, <i>5' Untranslated Region</i>
ASD	Transtornos do espectro do autismo, do inglês, <i>Autism Spectrum Disorders</i>
BPD	Transtorno bipolar, do inglês, <i>Bipolar Disorder</i>
CNV	Variações de número de cópias, do inglês, <i>Copy Number Variation</i>
DN	Distúrbios do neurodesenvolvimento
DNA	Ácido desoxirribonucleico, do inglês, <i>Desoxyribosenucleic Acid</i>
DT	Desenvolvimento típico
EB	Corpos embrionários, do inglês, <i>Embryoid Bodies</i>
EM	Maximização de expectativa, do inglês, <i>Expectation Maximization</i>
ESC	Célula-tronco embrionária, do inglês, <i>Embryonic Stem Cell</i>
FACS	Classificador de células por emissão de fluorescência, do inglês, <i>Fluorescence-Activated Cell Sorter</i>
FDR	Taxa de descoberta falsa, do inglês, <i>False Discovery Rate</i>
hg38	Genoma de referência humano montagem hg38
IFN	Interferon
iPSC	Células-tronco pluripotente induzida, do inglês, <i>Induced Pluripotent Stem Cells</i>
ISG	Genes estimulados por interferon, do inglês, <i>Interferon-Stimulated Gene</i>
KO	Nocaut, do inglês, <i>Knockout</i>
L1	Elemento nuclear intercalado longo tipo 1, do inglês, <i>Long Interspersed Nuclear Element 1</i>
L1CP-IND	base de anotações para L1 completos - genomas individuais
L1CP-REF	base de anotações para L1 completos - genoma de referência humano
L1HS	Elemento nuclear intercalado longo tipo 1 específico de humanos, do inglês, <i>Long Interspersed Nuclear Element 1, Human-Specific</i>
LINE-1	Elemento nuclear intercalado longo tipo 1, do inglês, <i>Long Interspersed Nuclear Element 1</i>
L1MA	Elemento nuclear intercalado longo tipo 1 de mamíferos
L1PB	Elemento nuclear intercalado longo tipo 1 de primatas
L1RG-IND	Base de anotações para regiões L1 - genomas individuais
L1RG-REF	Base de anotações para regiões L1 - genoma de referência humano
LTR	Longas repetições terminais, do inglês, <i>Long Terminal Repeat</i>

miRNA	MicroRNA
mRNA	RNA mensageiro, do inglês, <i>Messenger RNA</i>
NCBI	Centro nacional de informação biotecnológica, do inglês, <i>National Center for Biotechnology Information</i>
NE	Neurônio
Non-LTR	Não possuem longas repetições terminais, do inglês <i>non-Long Terminal Repeat</i>
NPC	Células precursoras neuronais, do inglês, <i>Neural Precursor Cells</i>
NSC	Células-tronco neurais, do inglês, <i>Neural Stem Cells</i>
nt	Nucleotídeos
ORF	Quadro aberto de leitura, do inglês, <i>Open Reading Frame</i>
P-Body	Corpos de processamento, do inglês, <i>Processing Bodies</i>
PO	Células oriundas de tecido cerebral <i>post mortem</i>
qPCR	Reação em cadeia da polimerase quantitativa, do inglês, <i>Quantitative Polymerase Chain Reaction</i>
RE	Retículo endoplasmático
RLE	Expressão relativa logarítmica, do inglês, <i>Relative Log Expression</i>
RNA	Ácido ribonucleico, do inglês, <i>Ribonucleic Acid</i>
RNP	Ribonucleoproteína
RTT	Síndrome de Rett, do inglês, <i>Rett syndrome</i>
SCZ	Esquizofrenia, do inglês, <i>Schizophrenia</i>
SG	Grânulos de estresse, do inglês, <i>Stress Granules</i>
SINE	Elemento nuclear intercalado curto, do inglês, <i>Short Interspersed Nuclear Element</i>
SNC	Sistema nervoso central
SVZ	Zona subventricular, do inglês, <i>Subventricular Zone</i>
TDAH	Transtorno do déficit de atenção/hiperatividade
TE	Elemento transponível, do inglês, <i>Transposable Element</i>
TPM	Transcritos por milhão, do inglês, <i>Transcripts per Million</i>
TPRT	Transcrição reversa no local alvo, do inglês, <i>Target Site Primed Reverse Transcription</i>
TSD	Duplicações no local alvo, do inglês, <i>Target-Site Duplication</i>
UCSD	Universidade da Califórnia San Diego

SUMÁRIO

1	INTRODUÇÃO.....	16
1.1	O RETROELEMENTO L1	17
1.1.1	<i>Anotação de retroelementos L1 no genoma humano.....</i>	<i>19</i>
1.2	ATIVIDADE DE L1 NO CÉREBRO.....	21
1.2.1	<i>Retrotransposição em distúrbios do neurodesenvolvimento</i>	<i>22</i>
1.3	REGULAÇÃO DE RETROELEMENTOS L1	24
2	JUSTIFICATIVA	29
3	OBJETIVOS	30
3.1	OBJETIVO GERAL	30
3.2	OBJETIVOS ESPECÍFICOS.....	30
4	MÉTODOS.....	31
4.1	BASE DE ANOTAÇÕES PARA RETROELEMENTOS L1	31
4.2	PREPARAÇÃO DE AMOSTRAS DE NEURÔNIOS DERIVADOS DE CÉLULAS iPSC DE SÍNDROME DE RETT..	33
4.3	OBTENÇÃO DE AMOSTRAS DE TRANSCRIPTOMA PÚBLICO (RNA-SEQ) DE PACIENTES COM DISTÚRBIOS DO NEURODESENVOLVIMENTO.....	35
4.4	ANÁLISE DE EXPRESSÃO DE BIBLIOTECAS DE TRANSCRIPTOMA	37
4.5	ANÁLISE DE ENRIQUECIMENTO RELACIONADO A PROCESSOS BIOLÓGICOS	38
4.6	ANÁLISE DE EXPRESSÃO DE RETROELEMENTOS L1	39
5	RESULTADOS E DISCUSSÃO	41
5.1	BASE DE ANOTAÇÃO PARA RETROELEMENTOS L1	41
5.2	ANÁLISE DE EXPRESSÃO DE RETROELEMENTOS L1 EM DISTÚRBIOS DO NEURODESENVOLVIMENTO	47
5.3	EXPRESSÃO DE GENES REGULADORES DE L1 EM DISTÚRBIOS DO NEURODESENVOLVIMENTO	50
5.3.1	<i>Expressão de reguladores L1 em tecidos cerebrais post mortem de DN.....</i>	<i>58</i>
5.3.2	<i>Expressão de reguladores L1 em iPSC de DN.....</i>	<i>62</i>
5.3.3	<i>Expressão de reguladores L1 em NPC de DN</i>	<i>64</i>
5.3.4	<i>Expressão de reguladores L1 em neurônios de DN</i>	<i>66</i>
5.3.5	<i>Associação entre genes reguladores de L1 e DN</i>	<i>72</i>
5.4	REGULAÇÃO DE RETROELEMENTOS L1 E VIA DE AUTOFAGIA.....	74
6	CONCLUSÕES.....	81
7	PERSPECTIVAS	84

8	ATIVIDADES COMPLEMENTARES.....	86
8.1	ARTIGOS PUBLICADOS.....	86
8.2	ARTIGOS SUBMETIDOS OU EM AVALIAÇÃO	87
	REFERÊNCIAS	89
	APÊNDICE A - REVISÃO SISTEMÁTICA DA LITERATURA PARA LEVANTAMENTO DE GENES PARTICIPANTES DO PROCESSO DE REGULAÇÃO DE RETROELEMENTOS L1.....	112
	APÊNDICE B - DISTRIBUIÇÃO DE REGIÕES L1HS A NÍVEL ALÉLICO PARA O GENOMA INDIVIDUAL CAUCASIANO	173
	APÊNDICE C - EXPRESSÃO DE GENES REGULADORES DE L1 EM DISTÚRBIOS DO NEURODESENVOLVIMENTO.....	174

1 INTRODUÇÃO

Elementos transponíveis (TEs) são sequências repetitivas no DNA (ácido desoxiribonucleico), capazes de modificar o tamanho do genoma de eucariotos, desde plantas, microorganismos e até mamíferos (BECK et al., 2010; KAZAZIAN, 2000; LANDER et al., 2001). Compõem cerca de 45% a 60% do genoma humano, e algumas sequências mantêm a habilidade de se mobilizar através deste (DE KONING et al., 2011; LANDER et al., 2001).

TEs possuem diferentes estruturas e formas de mobilização (GOODIER; KAZAZIAN, 2008). DNA *transposons* (classe II) utilizam o mecanismo de “cortar e colar”, sendo sua sequência retirada do genoma e inserida em uma nova localização, utilizando para isso a atividade da enzima transposase (MUÑOZ-LÓPEZ; GARCÍA-PÉREZ, 2010). Em humanos, DNA *transposons* constituem cerca de 3% do genoma e não são mais ativos (LANDER et al., 2001). Embora não seja reportada a atividade de DNA *transposons* na maioria dos mamíferos (RICHARDSON et al., 2015), morcegos contêm DNA *transposons* ativos em seus genomas, sugerindo que sua atividade continua a gerar diversidade genética em seu genoma (RAY et al., 2008). De maneira contrastante, os RNA *transposons* (classe I), também chamados de *retrotransposons* (ou retroelementos), são altamente ativos no genoma de mamíferos, sendo a classe mais abundante no genoma dos organismos eucarióticos (HAVECKER; GAO; VOYTAS, 2004). Utilizam o mecanismo “copiar e colar”, e são capazes de se auto copiar e se reintegrar em outro local no genoma através de transcrição reversa (GILBERT; MORAN, 2002).

Segundo sua organização estrutural, os retroelementos podem ser classificados em dois grupos, os que possuem repetições terminais longas (do inglês, *Long Terminal Repeat* - LTR) e os que não possuem LTR (*non*-LTR). O grupo de *non*-LTR é subdividido ainda com base na capacidade de elementos se mobilizarem independentemente (autônomos) ou apenas com a maquinaria codificada por outro retroelemento (não-autônomo). O elemento nuclear intercalado longo do tipo 1 (do inglês, *Long Interspersed Nuclear Element* 1 - LINE-1 ou L1) é o único retroelemento *non*-LTR autônomo em humanos e compõe aproximadamente 21 % do genoma nuclear destes (LANDER et al., 2001). Além de promover sua própria mobilização, L1s, através da sua maquinaria,

mobilizam elementos nucleares intercalados curtos (do inglês, *Short Interspersed Nuclear Element* - SINE) como elementos Alu, SVA (SINE-VNTR-Alu) e mRNAs (RNA mensageiro), resultando em pseudogenes processados (DEWANNIEUX; ESNAULT; HEIDMANN, 2003; ESNAULT; MAESTRE; HEIDMANN, 2000; RAIZ et al., 2012).

1.1 O retroelemento L1

Retroelementos L1 possuem tamanho aproximado de 6000 nucleotídeos (nt) e contêm uma região 5'UTR (região não traduzida 5') que possui um promotor RNA-polimerase II na orientação senso (SWERGOLD, 1990), dois quadros abertos de leitura (do inglês, *Open Reading Frame* - ORF) não sobrepostos denominados ORF1 e ORF2 (DOMBROSKI et al., 1991), e uma região 3' UTR terminando com uma cauda poli-A (Figura 1). ORF1 codifica uma proteína de aproximadamente 40 kDa (ORF1p) (MATHIAS et al., 1991) que apresenta atividades de ligação ao RNA e chaperona (MARTIN, 2006). Possui ainda, uma região C-terminal, um motivo de ligação ao RNA e um domínio N-terminal (OSTERTAG; KAZAZIAN, 2001a). A ORF2 codifica para uma proteína de 150 kDa (ORF2p) com três domínios conservados: um domínio N-terminal endonuclease (EN); domínio de transcriptase reversa (do inglês, *Reverse Transcriptase* - RT); e um domínio C-terminal (DOUCET et al., 2010). O retroelemento ainda apresenta uma região promotora na orientação antisenso da região 5'UTR do L1 (SPEEK, 2001), onde também situa-se uma ORF encontrada somente em primatas e denominada ORF0, cuja função ainda não é conhecida (DENLI et al., 2015).

Embora existam muitas lacunas no entendimento do mecanismo utilizado para mobilização de L1s através do genoma (retrotransposição), sabe-se que a atividade do L1 se inicia com a ativação de seu promotor interno contido na região 5'UTR. Com isto, o mRNA L1 é transcrito e exportado do núcleo para o citoplasma onde suas próprias proteínas, ORF1p e ORF2p ligam-se *in cis* ao L1 exportado. Isto resulta na formação de complexos compostos por proteínas ligadas ao ácido ribonucleico ou ribonucleoproteínas (RNP) (DOUCET et al., 2010; LEIBOLD et al., 1990), os quais podem se localizar no citoplasma juntamente com grânulos de *stress* (do inglês, *Stress Granules* - SGs) (DOUCET et al., 2010; GOODIER et al., 2007) ou reentrar no núcleo. Quando o complexo RNP retorna ao núcleo, a endonuclease codificada pelo L1 corta

uma das fitas do DNA genômico expondo um grupo 3' hidroxila, o qual é utilizado como um iniciador da transcrição reversa através da transcriptase codificada pelo L1 (do inglês, *Target Site Primed Reverse Transcription* - TPRT) (MATHIAS et al., 1991). A inserção do L1 tipicamente apresenta cauda poli(A), ocorre em regiões genômicas com padrões de nucleotídeos (motivos) ricos em timidina (como a sequência 5'-TTAAAA-3'), (JURKA, 1997) e pode apresentar duplicações em local alvo (do inglês, *Target-Site Duplication* - TSD), que possuem entre 7 a 20 nt (LEVIN; MORAN, 2011) (Figura 1).

Aproximadamente 40% das inserções de L1 no genoma nuclear são inserções idênticas a sequência do L1 que a originou (GILBERT et al., 2005), sendo a região 5'UTR a região que apresenta a maior variação de sequência em novas inserções (MYERS et al., 2002).

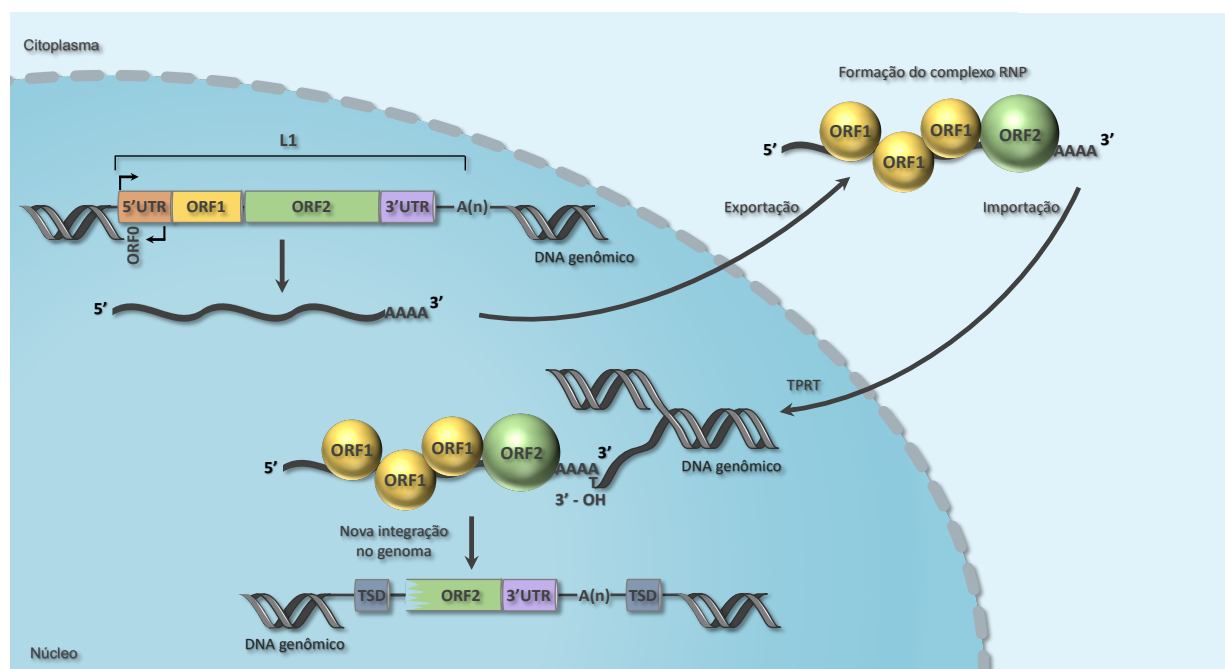


Figura 1: Biogênese de L1.

Os retroelementos L1 completos, consistem em uma região 5'UTR com função promotora, dois ORFs (do inglês, *Open Reading Frame*) denominados ORF1 e ORF2, um ORF0 na orientação antisense com função não reportada, e uma região 3'UTR terminando com uma cauda poli(A). As posições aproximadas de iniciação da transcrição na orientação senso e a iniciação da transcrição na orientação antisense são indicadas com setas na parte superior e inferior da região 5' UTR, respectivamente. A polimerase II transcreve o elemento em RNA, que é transportado para o ribossomo para produzir as proteínas ORF1p e ORF2p. Estas proteínas ligam-se ao transcrito L1 com preferência *cis*, formando o complexo de ribonucleoproteínas (RNP). Este complexo pode reentrar no núcleo da célula, provavelmente quando o envelope nuclear é quebrado durante a divisão celular. A endonuclease codificada pela ORF2, faz um corte em uma região rica em timidina e começa a transcrever reversamente o transcrito L1, processo denominado TPRT (do inglês, *Target Site Primed Reverse Transcription*). Inserções de L1 podem

apresentar duplicações em local alvo (do inglês, *Target-Site Duplication* - TSD), que possuem entre 7 a 20 nt. Se a transcrição reversa é ineficiente, pode gerar um retroelemento incompleto. Se a transcrição reversa for eficiente e o elemento recém-inserido contém todas as características funcionais, este novo retroelemento L1 será capaz de realizar a retrotransposição.

Fonte: o autor.

Devido a inserções de retroelementos L1 possuindo regiões incompletas ou contendo inversões, deleções ou mutações pontuais em sua sequência, somente entre 80 a 100 cópias em média são capazes de realizar a retrotransposição em cada indivíduo humano (BECK et al., 2010; LUTZ et al., 2003; OSTERTAG; KAZAZIAN, 2001b). Estes retroelementos ativos formam um subgrupo de elementos chamados “Ta” e são pertencentes a subfamília mais recente dos retroelementos L1 que é referida como L1PA1 ou L1HS, específica de *Homo sapiens* (BOISSINOT; CHEVRET; FURANO, 2000; SASSAMAN et al., 1997).

A vasta maioria das inserções de retroelementos L1 são encontradas em homozigose nos seres humanos (TANG et al., 2017), e representam inserções ancestrais que se iniciaram a pelo menos 170 milhões de anos (KHAN; SMIT; BOISSINOT, 2005). Para o mapeamento de *loci* no genoma humano onde L1s estão inseridos, é necessária uma anotação que permita classificá-los através das subfamílias a que cada elemento pertence.

1.1.1 Anotação de retroelementos L1 no genoma humano

Anotações decorrentes de inserções de L1 são comumente realizadas utilizando-se a ferramenta RepeatMasker (SMIT; HUBLEY; GREEN, 2018). Para isto, ela emprega sequências de L1 disponíveis em bibliotecas como RepBase (BAO; KOJIMA; KOHANY, 2015) e Dfam (WHEELER et al., 2013). Para o genoma humano, as anotações reportadas pelo RepeatMasker compreendem dezoito subfamílias L1, sendo a L1HS correspondente às inserções mais recentes ocorridas no genoma nuclear, e a L1PA17 correspondente às inserções mais antigas (Figura 2A). Quando compara-se as contagens de inserções da subfamília mais recente L1HS, com as inserções de subfamílias mais antigas como L1PA5 ou L1PA16, observa-se que a maior parte das inserções de L1 ocorridas no genoma humano advém destas subfamílias mais antigas (Figura 2A). Ao analisar a subfamília L1HS, pode-se notar que aproximadamente 81% das inserções no genoma correspondem a retroelementos incompletos, sendo que

apenas 304 retroelementos correspondem ao comprimento médio aproximado de 6000 nt esperado para um L1HS (Figura 2B).

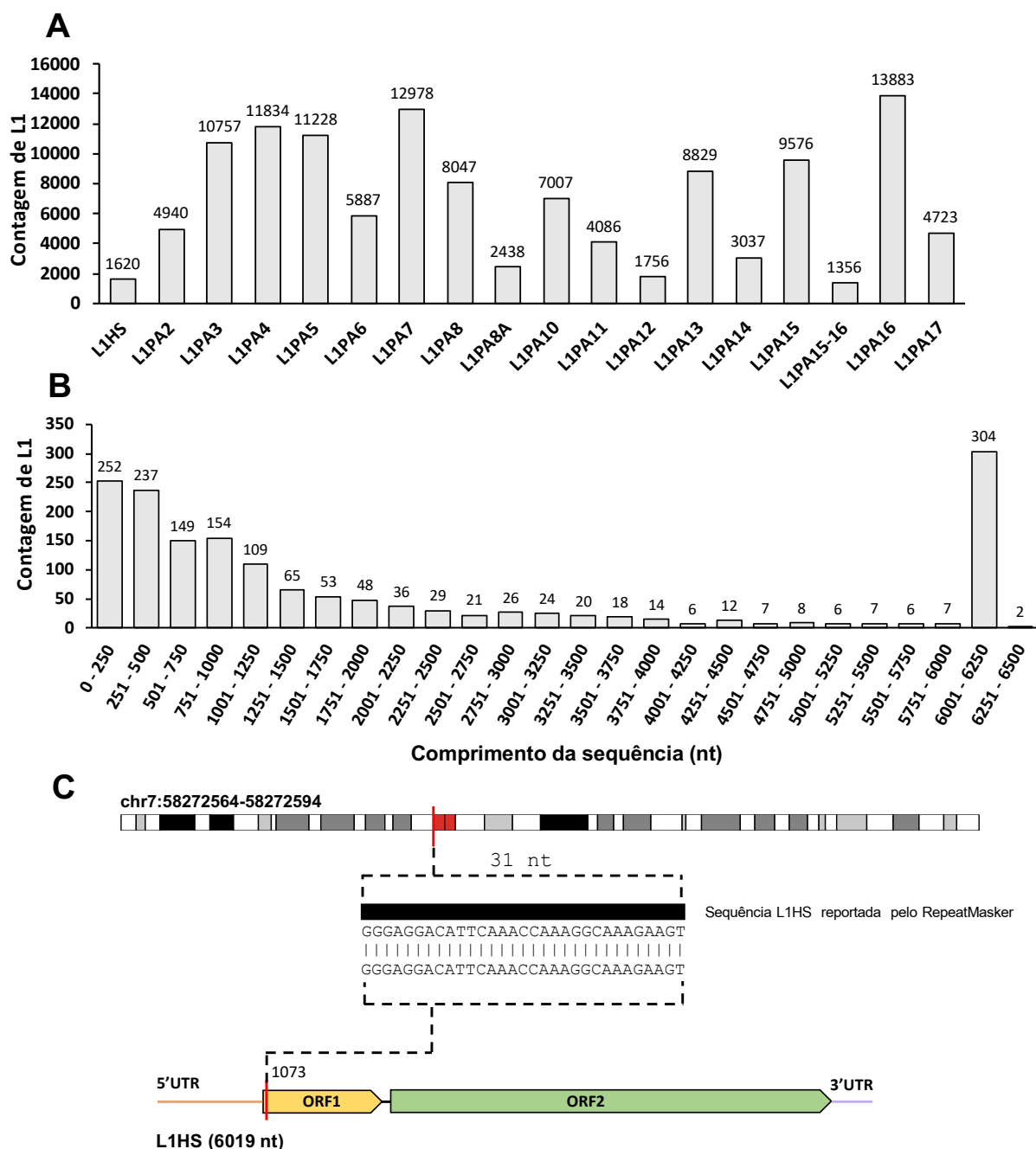


Figura 2: Anotação de subfamílias de retroelementos L1 fornecidos pela ferramenta RepeatMasker.
A. Número de cópias de inserções de retroelementos L1 por subfamília. As subfamílias L1 são representadas em ordem cronológica de inserções, sendo os retroelementos L1HS contendo as inserções mais recentes e os retroelementos L1PA17 as inserções mais antigas. **B.** Distribuição de retroelementos

da subfamília L1HS por intervalos de comprimento em nucleotídeos (nt) da sequência anotada como L1. **C.** Sequência de 31 nt relatada por RepeatMasker e atribuída como membro da subfamília L1HS. A sequência está no *locus* chr7: 58272564-58272594 (hg38) e pertence a uma ORF1, posição 1073 de um retroelemento L1HS.

Fonte: o autor.

As anotações reportadas pelo RepeatMasker fornecem um conjunto amplo da localização de retroelementos L1 no genoma humano. Porém, algumas limitações podem ser observadas, como anotações de *loci* genômicos com sequências menores que 100 nt reportadas como L1HS. Em um dos casos, há uma região genômica extremamente curta e repetitiva no genoma, de 31 nt de comprimento, que é anotada como um elemento L1HS (Figura 2C). Essas pequenas sequências anotadas podem levar a erros de interpretação ao se analisar a expressão das ORFs presentes em L1, uma vez que apenas alguns retroelementos específicos, com aproximadamente 6000 nt de comprimento, são capazes de se transcrever e se expressar como uma molécula de RNA com potencial relevância biológica (SMIT; HUBLEY; GREEN, 2018). Detalhes sobre as regiões L1 (5'UTR, ORF1, íntron, ORF2, 3'UTR) e também de L1s completos (contêm todas as regiões de uma L1), não são fornecidos por meio das anotações do RepeatMasker. Para o genoma humano, outra limitação que pode-se ressaltar, são anotações de L1s que existem apenas para o genoma de referência, porém nenhum dado relacionado com um genoma diploide individual. O problema disso é o fato de que o genoma de referência humano é representado como uma sequência consenso haplóide, sendo originado a partir da sequência do genoma de múltiplos indivíduos (SCHNEIDER et al., 2017) o que não permite uma avaliação de inserções de L1 a nível alélico. Assim, a disponibilidade de recursos que melhor caracterizam e anotam elementos L1 com informações detalhadas, incluindo anotações de sequências baseadas em similaridade, podem aprimorar estudos *in silico*, *in vitro* e *in vivo* que busquem compreender o papel dos retroelementos L1 no genoma humano, e seus impactos na determinação de fenótipos moleculares e celulares.

1.2 Atividade de L1 no cérebro

O progresso constante das técnicas de bioinformática bem como no sequenciamento de DNA e análise genética, tornou possível identificar variantes genéticas que geram genótipos distintos entre células cerebrais humanas. Essas

variantes incluem polimorfismos de um único nucleotídeo, variações no número de cópias (CNVs) e inserções de retroelementos (BAILLIE et al., 2011; ERWIN et al., 2016; EVRONY et al., 2012; MCCONNELL et al., 2013; UPTON et al., 2015), as quais contribuem para a formação do mosaicismo de células cerebrais. O mosaicismo é definido como a existência de duas ou mais células com genótipos diferentes em um indivíduo (HAZEN et al., 2016; MACIA et al., 2017; RICHARDSON et al., 2017).

Vários estudos mostram que os retroelementos contribuem para a variação individual em genomas de células cerebrais, promovendo assim sua diversidade e heterogeneidade. Desta forma, a manutenção da atividade destes retroelementos pode ser essencial para uma função cerebral saudável, enquanto que falhas em sua regulação podem desempenhar um papel no desenvolvimento de doenças neurológicas (ERWIN; MARCHETTO; GAGE, 2014; SINGER et al., 2010).

1.2.1 Retrotransposição em distúrbios do neurodesenvolvimento

Retroelementos podem ser ativados por um ambiente celular anormal e, assim, gerar inserções somáticas com o potencial de influenciar o desenvolvimento de doenças como os distúrbios do neurodesenvolvimento (DN).

Os DN são condições caracterizadas por prejuízos na cognição, comunicação, comportamento e/ou habilidades motoras resultantes do desenvolvimento cerebral anormal (MULLIN et al., 2013). Deficiência intelectual, distúrbios da comunicação, transtornos do espectro do autismo (do inglês, *Autism Spectrum Disorder* - ASD), transtorno do déficit de atenção com hiperatividade (TDAH), síndrome de Rett (RTT), esquizofrenia (do inglês, *Schizophrenia* - SCZ) e transtorno bipolar (do inglês, *Bipolar Disorder* - BPD), se enquadram no grupo de DN (AMERICAN PSYCHIATRIC ASSOCIATION, 2013; MARANGONI; DE CHIARA; FAEDDA, 2015; NEUL; ZOGHBI, 2004; RAPOPORT; GIEDD; GOGTAY, 2012).

Uma relevante associação entre L1 e DN é observada na síndrome de Rett. A RTT é um DN causado por mutações no gene *MECP2* (*Methyl-CpG Binding Protein 2*) (AMIR et al., 1999) e tem como principais atributos autismo, perda da fala, torção das mãos, ansiedade e eventual deterioração da capacidade motora. O gene *MECP2* regula globalmente os níveis de metilação do DNA e reprime a transcrição e retrotransposição

de L1 (MUOTRI et al., 2010; YU, 2001). Já foi observado que células neuroepiteliais de camundongos *knockout* (KO) para o gene *MECP2* expressam até 4 vezes mais transcritos L1 e também exibem um aumento de até 3,5 vezes no número de novas inserções comparadas a amostras de camundongos controle (MUOTRI et al., 2010). Além disso, também já foi demonstrado que em células precursoras neuronais (do inglês, *Neural Precursor Cells* – NPC) diferenciadas a partir de fibroblastos de pacientes com RTT há um aumento de até 2,5 vezes na quantidade de eventos de retrotransposição em comparação com células controle oriundas de pessoas com desenvolvimento típico (DT) (MUOTRI et al., 2010). Em outro trabalho, também foi observado que neurônios provenientes de amostras *post mortem* de cérebros de pacientes com RTT apresentaram maior número de inserções de L1 no DNA genômico em comparação com amostras de pacientes controle DT. Tais dados permitiram hipotetizar que neurônios da RTT provavelmente foram submetidos a uma maior quantidade de eventos de retrotransposição (ZHAO et al., 2019).

Retroelementos L1, de uma forma geral, têm influência em genes que são fatores de risco dentro do ASD (JACOB-HIRSCH et al., 2018; WILLIAMS et al., 2013). ASD representa um grupo heterogêneo de doenças que compartilham um conjunto de características comuns, onde incluem-se déficits na comunicação social e desenvolvimento da linguagem que são acompanhadas por interesses altamente restritos, comportamentos estereotipados ou ambos (VOLKMAR; STATE; KLIN, 2009). Já foi observado que em tecidos da região do cerebelo de pacientes com ASD há aumento de estresse oxidativo, desregulação epigenética (JAMES et al., 2013) e também expressão elevada dos transcritos ORF1 e ORF2 de L1 comparada a controles de DT (SHPYLEVA et al., 2018). Também foi observado que a ligação da proteína MECP2 na região 5'UTR de L1 foi significativamente menor no cerebelo de pacientes diagnosticados com ASD, contribuindo para o aumento da expressão de L1s (SHPYLEVA et al., 2018).

Existem ainda evidências de que a redução da metilação de L1 está associada ao desenvolvimento de diferentes patologias (BABUSHOK; KAZAZIAN, 2007). Em um estudo envolvendo amostras de sangue periférico de 92 pacientes com SCZ e 99 pacientes com BPD, foi demonstrado que os níveis de metilação em três locais da região 5'UTR de L1s é menor quando comparado aos respectivos controles, sugerindo a

possibilidade de uma maior atividade de L1s nestes distúrbios (LI et al., 2018). Neste sentido, um estudo complementar revelou um aumento significativo na expressão da ORF2 de L1 em neurônios isolados do córtex pré-frontal de pacientes com SCZ quando comparados com os respectivos neurônios controle (BUNDO et al., 2014). Esse mesmo resultado foi replicado em cultura de neurônios de pacientes afetados pela deleção 22q11, uma alteração cromossômica encontrada em pacientes com SCZ (BASSETT et al., 2003). Por fim, foi observado um aumento consistente no número de cópias de L1 no tecido de córtex pré-frontal de amostras de SCZ para humanos e camundongos (BUNDO et al., 2014).

1.3 Regulação de retroelementos L1

A retrotransposição pode ocorrer em células de linhagem germinativa, durante a embriogênese, bem como durante o desenvolvimento do sistema nervoso, e pode envolver células-tronco neurais (do inglês, *neural stem cells* - NSC), NPC e neurônios pós-mitóticos (KUBO et al., 2006; MACIA et al., 2017; MUOTRI et al., 2005). Dependendo do estágio de desenvolvimento e tipo celular, a inserção de um retroelemento L1 pode ser expandido às células descendentes dela, podendo ocasionar efeitos potencialmente significativos na fisiologia cerebral (CRICHTON et al., 2014). Assim, mecanismos que restringem a expressão e a atividade de L1s são particularmente importantes tanto em células germinativas quanto em células adultas.

Esses mecanismos de regulação da expressão e atividade de L1s são mantidos através da atuação de 158 genes reguladores de L1 conhecidos e catalogados por este trabalho (Apêndice A). Estas genes atuam na regulação de L1 a nível pré-transcricional e pós-transcricional (Figura 3). A nível pré-transcricional, as modificações das histonas e a metilação do DNA, por exemplo, desempenham papéis importantes na regulação destes retroelementos (CUELLAR et al., 2017; GARCIA-PEREZ; WIDMANN; ADAMS, 2016; KARIMI et al., 2011). A repressão de L1 por metilação é promovida pela proteína com atividade de histona deacetilase e expressa em neurônios, SIRT6 (*Sirtuin 6*) (VAN METER et al., 2014). Outras proteínas, presentes neste nível de regulação e que atuam reprimindo a transcrição de L1, são as metiltransferases DNMT1 (*DNA Methyltransferase*

1), DNMT3A (*DNA Methyltransferase 3 Alpha*), DNMT3 (*DNA Methyltransferase 3 Beta*) (LIANG et al., 2002) e a proteína de ligação a grupos metis, MECP2.

Proteínas associadas a cromatina também são responsáveis pela regulação de retroelementos L1. A proteína CHAF1 (*Chromatin assembly factor 1*) é uma chaperona capaz de reunir histonas H3 e H4 em DNA recém-sintetizado, tanto durante a fase de replicação quanto na fase de reparo do DNA (CLÉMOT et al., 2018). Estudos mostraram que embriões de camundongos são protegidos da atividade de retroelementos por modificações repressivas de histonas mediadas por CHAF1 (HATANAKA et al., 2015). Também foi observado em células-tronco embrionárias de camundongos, que o acúmulo da marca de histona repressiva H3K9me3, mediado pelas metiltransferases SUV39H1 e SUV39H2, leva à repressão transcricional de L1 (BULUT-KARSLIOGLU et al., 2014).

Também a nível pré-transcricional, a clivagem mediada pela EN produzida pelo L1 no DNA alvo da inserção, pode ocasionar a quebra da fita dupla (DSB) e a ativação de vias de sinalização de DSB e vias de reparo por recombinação não-homóloga (NHEJ). Células humanas HeLa, com níveis reduzidos da proteína ATM (*Ataxia Telangiectasia Mutated*), uma quinase ativada após a DSB, atenuaram a atividade de L1 (GASIOR et al., 2006; WALLACE et al., 2013). Além disso, também foi reportado que em linhagens celulares DT40 (*Gallus gallus*) com *knockout* para genes pertencentes a via NHEJ, como *DCLRE1C* (*DNA Cross-Link Repair 1C*), *LIG4* (*DNA Ligase 4*) e *XRCC6* (*X-Ray Repair Cross Complementing 6*), foi observado uma diminuição na atividade de L1 transfectados de humanos, associando esta via a regulação de L1s (SUZUKI et al., 2009).

A nível pós-transcricional há diferentes fatores, como os envolvidos na biogênese de moléculas de microRNA (miRNA), que atuam diretamente na regulação da atividade de retroelementos L1 (GOODIER et al., 2012; ZHANG et al., 2014; ZHAO et al., 2013). Desta forma, o complexo microprocessador (Drosha / DGCR8), um importante complexo nuclear implicado na biogênese de miRNAs, é capaz de ligar-se *in vivo* ao transcrito de L1 reduzindo sua expressão e limitando a atividade destes retrotransposons (HERAS et al., 2013, 2014). miRNAs também estão envolvidos diretamente com a regulação de retroelementos (ROBERTS; CARDIN; BORCHERT, 2014), como o mir-128 que liga-se ao transcrito de L1, reprimindo assim sua integração em células HeLa e iPSC (HAMDORF et al., 2015).

Outros tipos de genes que também participam da regulação da expressão e atividade de L1 são aquelas envolvidas na restrição de atividade viral. Neste contexto, a resposta imune inata é uma das primeiras linhas de defesa contra uma ampla gama de infecções virais. Ela envolve fatores celulares com atividades antivirais, entre os quais a via de resposta ao interferon (IFN) desempenha um papel central (IVASHKIV; DONLIN, 2014; MACMICKING, 2012). Essa via leva à ativação de genes estimulados por IFN (ISG), atuando como efetores e reforçando a própria sinalização de IFN. Vários ISGs, como MOV10 (*Mov10 RISC Complex RNA Helicase*), ZAP (*Zinc Finger CCCH-Type Containing, Antiviral 1*) ou RNASEL (*Ribonuclease L*), limitam a replicação de L1 restringindo o acúmulo de transcritos L1. O aumento de expressão da proteína MOV10 reduz os níveis de transcritos L1 a nível pós-transcricional, enquanto que sua inibição aumenta os níveis de transcritos de L1 reduzindo também sua atividade (LI et al., 2013). Em cultura celular de células HeLa e 293T, foi observado que a proteína MOV10 se co-localiza com a ORF1p de L1 em grânulos citoplasmáticos, e liga-se ao transcrito L1 (ARJAN-ODEDRA et al., 2012; GOODIER et al., 2012). Assim como a proteína MOV10, a proteína ZAP é um membro da família de proteínas poli (ADP-ribose) polimerase (PARP), que também se co-localiza com a ORF1p e liga-se ao transcrito L1 em grânulos de estresse, reduzindo a atividade deste retroelemento (MOLDOVAN; MORAN, 2015)(GOODIER et al., 2015). Outro gene induzido por interferon, a *RNASEL*, é traduzido para uma endoribonuclease que liga e cliva regiões de fita simples de RNAs virais e celulares, e após ativação prolongada, induz autofagia e apoptose e a morte de células que contêm vírus (CASTELLI et al., 1998; SILVERMAN et al., 1983). Ele restringe a atividade de L1 em células humanas cultivadas e causa perda do transcrito de L1 (ZHANG et al., 2014).

Também a nível pós-transcricional, as vias de autofagia, embora pouco exploradas dentro da regulação de L1, vem a apresentar um papel importante na remoção de transcritos L1 e regulação de sua atividade. A macroautofagia (ou autofagia) é um processo catabólico pelo qual as células submetem os componentes citosólicos à degradação e reciclagem nos lisossomos, muitas vezes em resposta ao estresse metabólico ou à privação de nutrientes (BENTO et al., 2016). Tais compostos citosólicos incluem agregados de proteínas, mitocôndrias danificadas, peroxissomos

desnecessários, excesso de ribossomos, retículo endoplasmático e endossomos, corpos lipídicos, patógenos intracelulares (SHAID et al., 2013; ZAFFAGNINI; MARTENS, 2016) e também transcritos provenientes de retroelementos (GUO et al., 2014).

A ligação entre o sistema de autofagia e a degradação de retroelementos L1 é exposto no trabalho de Guo e colegas (GUO et al., 2014). Foi demonstrado que a autofagia pode direcionar especificamente o transcrito de L1 para a degradação, prevenindo assim a reinserção destes no genoma (GUO et al., 2014). Por meio de abordagens bioquímicas e baseadas em microscopia utilizando células HeLa e 293T, o trabalho demonstrou que elementos L1 e também elementos Alus co-localizam-se e co-purificam-se com autofagossomos, levando à sua subsequente degradação. Ainda, a degradação por autofagia de transcritos provenientes de L1s e Alus pode restringir o grau de inserção genômica destes. Além disso, foi demonstrado que o transcrito L1 se co-localiza com corpos de processamento (*P-Body*) e grânulos de estresse (SG) e, em concordância com descobertas recentes de leveduras, vermes e mamíferos (BUCHAN et al., 2013; ZHANG et al., 2009; ZHAO; TIAN; ZHANG, 2009), esses grânulos de RNA citoplasmático são direcionados para degradação autofágica pelos receptores de autofagia CALCOCO2 (*Calcium Binding And Coiled-Coil Domain 2*) e SQSTM1 (*Sequestosome 1*), respectivamente. Por fim, em tecidos de ovários, córtex pré-frontal e tecido muscular de camundongos heterozigotos para o gene *Becn1*^{+/-} (*Beclin 1*), um gene necessário para a biogênese de autofagossomos, foram observados níveis aumentados de transcritos L1 bem como um significativo acúmulo de inserções genômicas (GUO et al., 2014).

2 JUSTIFICATIVA

Embora anotações para retroelementos L1 atualmente disponíveis forneçam um conjunto amplo de inserções L1 no genoma humano, são observadas limitações, como a falta de informação sobre regiões individuais de L1 e anotações de elementos L1 que possuem todas as suas regiões (L1 completo). Neste sentido, este trabalho vem a apresentar uma base de anotações de retroelementos L1 para o genoma de referência humano e genomas individuais diploides de um Caucasiano e um Asiático, contendo informações sobre regiões L1 além de elementos L1 completos.

Também, sabe-se que a atividade de L1 no cérebro pode ocorrer em diversos tipos celulares, incluindo NPC e neurônios pós-mitóticos. O estresse ambiental, bem como a alteração da expressão de moléculas reguladoras de L1, podem resultar em altos níveis de atividade de L1 e levar ao desenvolvimento de DN. Isto direciona a investigação a expressão de transcritos L1 em células provenientes de ASD, RTT, BPD e SCZ, observando também a expressão destes retroelementos em diferentes tipos celulares.

Ainda, existe pouco conhecimento a respeito da regulação de L1 em células cerebrais de DN. Para preencher esta lacuna, é necessário o levantamento de um conjunto de genes reguladores de L1 atualmente conhecidos e a análise do perfil de expressão destes em diferentes tipos celulares provenientes de DN. Também é necessário explorar nesta análise, o contexto de processos biológicos envolvidos na regulação de L1.

Por fim, no contexto de processos biológicos envolvidos na regulação de L1 em células cerebrais, a autofagia vem a desempenhar papel importante na sobrevivência destas células, como em neurônios, que permanecem em um estado quiescente e são particularmente dependentes de vias robustas de controle de qualidade para manter uma homeostase celular adequada. A ligação entre a remoção de transcritos provenientes de L1s e a via de autofagia ainda é pouco compreendida, o que leva a um aprofundamento na cadeia de interações que ocorrem para a remoção de L1s através desta via. Com isso, é apresentada uma nova molécula que pode atuar na regulação da expressão e atividade de L1s em células cerebrais humanas.

3 OBJETIVOS

3.1 Objetivo Geral

Avaliar a expressão de genes que regulam a expressão e atividade de retroelementos do tipo L1 em dados de transcriptoma oriundos de amostras celulares de indivíduos com distúrbios do neurodesenvolvimento.

3.2 Objetivos Específicos

- a) Criar um algoritmo para detecção e anotação de retroelementos L1 completos e parciais em genomas.
- b) Criar uma base de dados contendo anotações de retroelementos L1 completos e parciais em genomas humanos.
- c) Criar um catálogo de genes reguladores da atividade e expressão de retroelementos L1.
- d) Analisar a variação de expressão de retroelementos L1 em distúrbios do neurodesenvolvimento.
- e) Analisar a variação de expressão de genes reguladores de L1 em distúrbios do neurodesenvolvimento.
- f) Analisar a regulação de retroelementos L1 relacionada ao processo biológico da autofagia.

4 MÉTODOS

4.1 Base de anotações para retroelementos L1

A base de anotação de L1 é composta de anotações das regiões do L1 (5'UTR, ORF1, íntron, ORF2, 3'UTR) e por anotações de retroelementos L1 completos (contendo todas as 5 regiões), ambas distribuídas por subfamílias.

Na construção das anotações de L1 (Figura 4), foram utilizadas as sequências de consenso das subfamílias L1 descritas por Khan e colegas (KHAN; SMIT; BOISSINOT, 2005). A abordagem utilizando sequências de consenso em vez de homologia para a construção da base de anotações, se deve ao fato de considerarmos apenas os elementos L1 dispostos no genoma humano, e também pelo alto grau de conservação das sequências ORF1 e ORF2, essenciais para a caracterização de elementos L1 (KHAN; SMIT; BOISSINOT, 2005; SOOKDEO; HEPP; BOISSINOT, 2018; WAGSTAFF; BARNERSOI; ROY-ENGEL, 2011).

Para a construção da base de anotações contendo regiões L1, estas regiões foram inicialmente extraídas para cada uma das dezoito subfamílias L1 (L1HS, L1PA2, L1PA3, L1PA4, L1PA5, L1PA6, L1PA7, L1PA8, L1PA8A, L1PA10, L1PA11, L1PA12, L1PA13B, L1PA13A, L1PA14, L1PA15A, L1PA15B, L1PA16), e alinhadas contra um genoma humano com a utilização do *software* Bowtie2 (LANGMEAD; SALZBERG, 2012) (modo *end-to-end*, parâmetros: -a -D 20 -R 3 -N 0 -L20 -iS, 1,0,50). Os genomas utilizados para este processo foram o genoma de referência humano, versão hg38 (montagem GCA_000001405.27), o genoma individual diploide de um indivíduo Caucasiano (LEVY et al., 2007) (montagem GCA_000252825.1) e o genoma individual de um indivíduo Asiático (AHN et al., 2009) (montagem KOREF_20090224).

Após o processo de alinhamento, foi executado o cálculo de similaridade¹ e identidade² de sequência para a parametrização das regiões alinhadas. Através desta etapa, temos uma base de anotações de regiões L1 onde ocorrem sobreposições entre regiões de subfamílias L1 distintas. Para remover essas sobreposições, foi considerado o nível de similaridade e identidade entre as sequências. Assim, para uma região

¹ Valor numérico que representa a relação entre o comprimento (nt) entre duas sequências distintas.

² Valor numérico que representa a relação entre o número de nucleotídeos idênticos entre duas sequências distintas.

genômica em que ocorre a sobreposição de uma mesma região L1, mas de subfamílias L1 distintas, a região com maior identidade e similaridade é escolhida para a base de anotações. A base de anotações não sobreposta é então filtrada, criando anotações de regiões com similaridade mínima de 90% e identidade mínima de 70% (sequências L1 conservadas) e anotações de regiões com similaridade mínima de 90% e identidade mínima de 99% (sequências L1 altamente conservadas) (Base de regiões L1 - Figura 4).

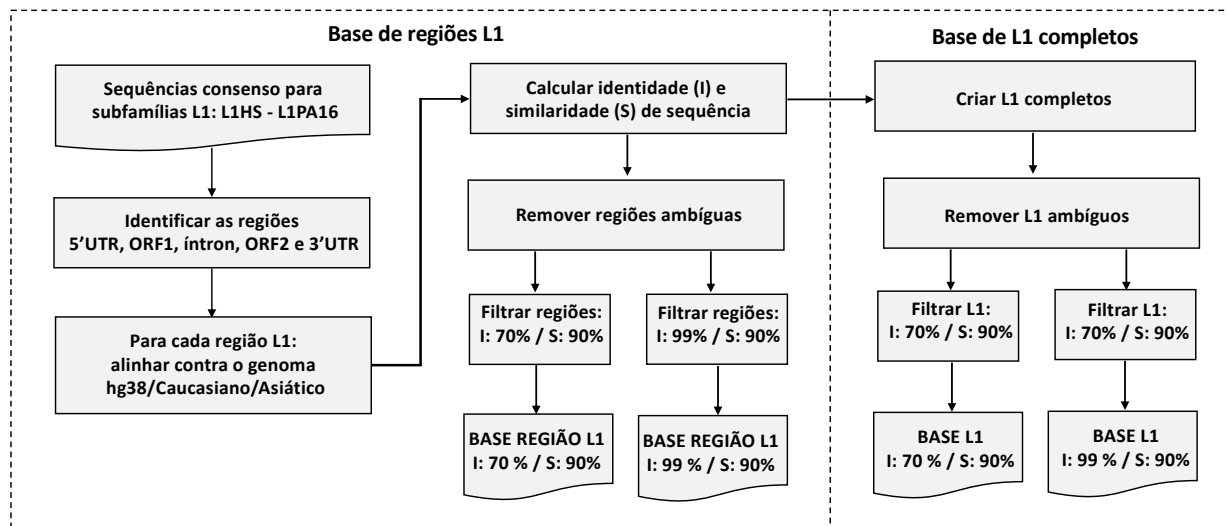


Figura 4: Fluxograma de criação da base de anotações para retroelementos L1.

A base de anotações criada, fornece entradas de regiões para subfamílias L1 e retroelementos L1 completos para o genoma de referência humano (hg38) e os genomas individuais Caucasiano e Asiático. Todas as bases de anotações foram criadas considerando sequências L1 conservadas (similaridade mínima de 90% e identidade mínima de 70%) e altamente conservadas (similaridade mínima de 90% e identidade mínima de 99%).

Fonte: o autor.

Para a criação da base de anotações contendo retroelementos L1 completos, as regiões L1 com informações sobre similaridade e identidade de sequência, identificadas na etapa anterior, foram ordenadas de acordo com suas posições genômicas, permitindo lacunas de no máximo 1 nt entre duas regiões adjacentes. Após, foi realizada a remoção de elementos sobrepostos considerando os mesmos critérios da etapa anterior. A base de anotações não sobreposta de elementos L1 completos é então filtrada, criando anotações com sequências L1 conservadas e sequências L1 altamente conservadas.

4.2 Preparação de amostras de neurônios derivados de células iPSC de síndrome de Rett

Para esta pesquisa foram preparadas 6 amostras de neurônios derivados de iPSC de dois indivíduos com RTT e seus respectivos controles DT (Tabela 1). A geração destas amostras e dados, foram preparadas pelo orientador e colaboradores durante sua pesquisa de pós-doutoramento no Laboratório de Células Tronco da Universidade da Califórnia San Diego (UCSD, USA).

Para criar neurônios funcionais para extração de RNA, foram geradas iPSC a partir de fibroblastos de pacientes do sexo masculino com RTT (Q83X e N126I) e dos pais destes pacientes (WT83 e WT126), respectivamente. Estas células foram reprogramadas e diferenciadas em NPC, conforme protocolo descrito anteriormente por Marchetto e colaboradores (MARCHETTO et al., 2010).

Tabela 1: Descrição de amostras com mutações pontuais no gene *MECP2*.

As amostras Q83X e N126I são linhagens celulares derivadas de pacientes do sexo masculino portadores de RTT.

Amostra	Descrição
WT83	Amostra controle (pai do paciente Q83X)
WT126	Amostra controle (pai do paciente N126I)
Q83X	mutação <i>nonsense</i> de uma glutamina (Q) para um stop códon (X) na posição 83 do aminoácido, (NP_004983:p.Gln83*).
N126I	mutação <i>missense</i> de uma asparagina (N) para uma isoleucina (I) na posição 126 do aminoácido, (NP_004983:p.Asn126Ile).

A diferenciação neuronal é iniciada pela geração de corpos embrionários (EB) com meios de diferenciação neuronal. Para o enriquecimento de NPC, as rosetas foram coletadas manualmente a partir de EB previamente plaqueados (Figura 5A).

As rosetas foram dissociadas e revestidas para formar uma população homogênea de NPC, que são positivas para marcadores neurais iniciais, como a proteína Nestina (*Nestin*) (Figura 5B). As NPC continuaram a se proliferar em meios neurais suplementados com FGF2 (*Fibroblast Growth Factor 2*) (Figura 5C-5D) e a diferenciação neuronal foi induzida após a retirada do FGF2. Em 2 semanas, uma rede neuronal se torna visível (Figura 5E-5F) e também neurônios eletrofisiologicamente maduros (Figura 5G-5H). Nesta fase, os neurônios são positivos para marcadores

neuronal, como os genes *TUJ-1* e *MAP2*. Durante as próximas 2 a 3 semanas seguintes, as células foram então mantidas em meios de diferenciação até serem utilizadas para experimentos adicionais.

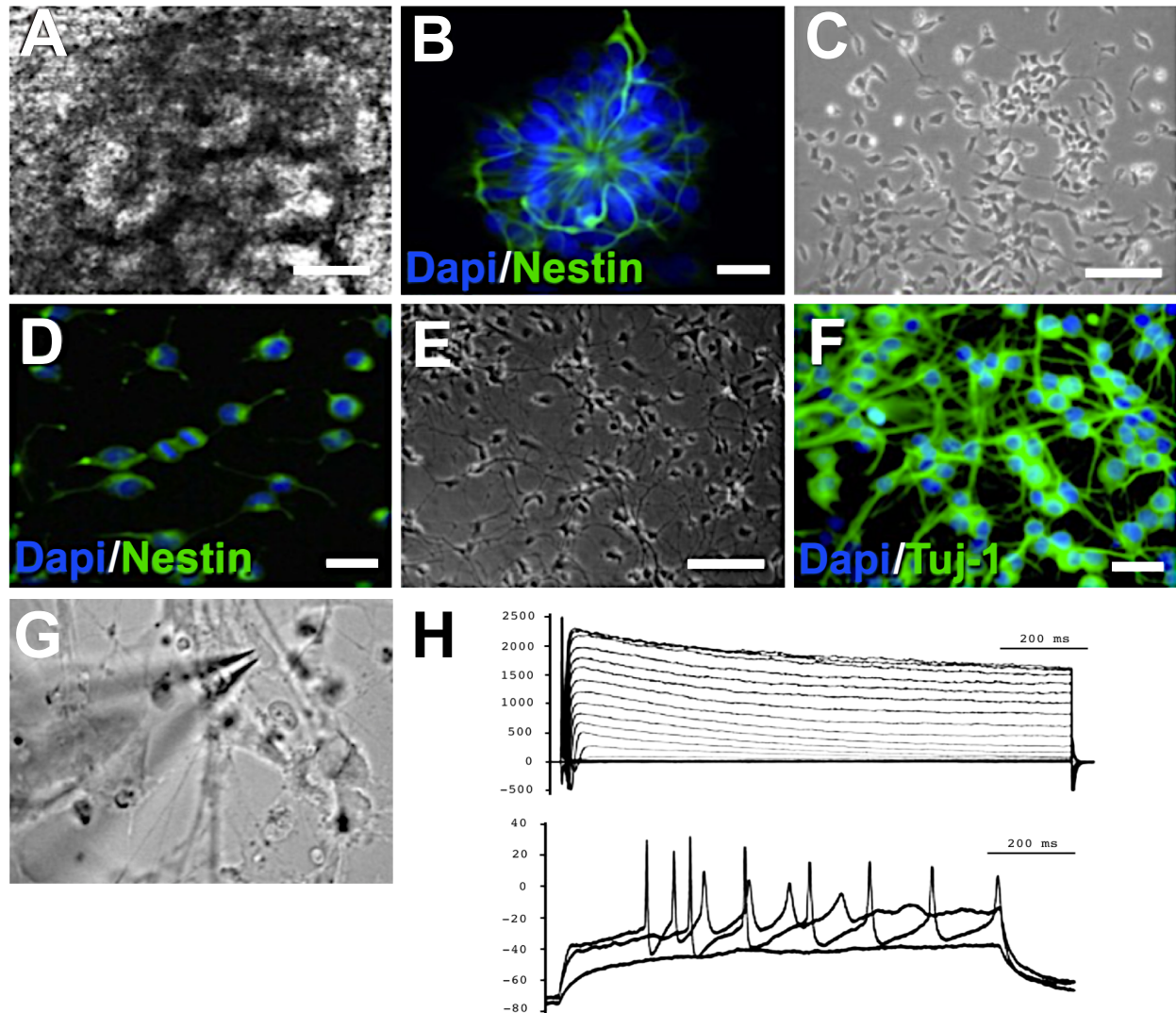


Figura 5: Diferenciação neural de iPSC não é afetada pela reprogramação celular.

A e B. Rosetas derivadas de EB. **C e D.** Aspecto de células precursoras neurais. **E e F.** Neurônios derivados de NPC. **G.** Imagem representativa do neurônio eletrofisiologicamente maduro. **H.** Superior: correntes transitórias de Na^+ e correntes sustentadas de K^+ em resposta a despolarizações; Inferior: respostas sub e supra-limiar a injeções somáticas de corrente em células agrupadas.

Fonte: o autor por meio de colaboração científica com a UCSD.

Para as células diferenciadas, foi realizado o enriquecimento de neurônios através de seleção por FACS (do inglês, *Fluorescence-Activated Cell Sorter*) usando neurônios

transduzidos com um vetor sinapsina::EGFP (do inglês, *Enhanced Green Fluorescent Protein*) derivado do HIV-1 (Invitrogen Inc, Carlsbad, CA, USA).

O RNA foi coletado das culturas enriquecidas com neurônios usando a extração com TRIzol e as bibliotecas foram preparadas com os kits de preparação de amostras de RNA TruSeq Stranded Total RNA (Illumina Inc, San Diego, CA, USA). Usando *barcoding*, o RNA mensageiro de diferentes amostras foi multiplexado em 3 amostras por *lane*, usando o sequenciamento em larga escala (HGS – *High throughput sequencing*) do equipamento Illumina HiSeq 2000, gerando bibliotecas de RNA-seq de pelo menos 40 milhões de *reads paired-end* de 100 bp por amostra.

A escolha destas amostras deve-se ao fato de apresentarem um caso mais grave da doença, onde temos a perda da proteína MECP2 (Q83X), e outro mais brando, onde a proteína MECP2 sofre uma mutação *missense* (N126I). Como a proteína MECP2 é um regulador da atividade de L1, estas amostras fornecem o meio para um estudo dirigido sobre a atividade e regulação de L1s nesta síndrome.

4.3 Obtenção de amostras de transcriptoma público (RNA-seq) de pacientes com distúrbios do neurodesenvolvimento

Para o levantamento de amostras de pacientes com DN, foi realizada uma busca por dados de transcriptoma (RNA-seq) disponíveis através do banco de dados público BioProject, disponibilizado pelo *National Center for Biotechnology Information* (NCBI). O banco BioProject armazena diversas categorias de dados biológicos produzidos pela comunidade de pesquisa. Os tipos de dados incluem por exemplo, genomas completos e parciais, transcriptomas, epigenoma, dados de variantes genéticas e dados fenotípicos (BARRETT et al., 2012). A busca por dados de transcriptoma considerou os parâmetros de busca expostos na Tabela 2, e também amostras com um número mínimo de 30 milhões de *reads*. Como resultado, obtivemos um total de 717 amostras. Para esta pesquisa, foram totalizadas 729 amostras, considerando aquelas descritas no item 4.2 (Tabela 3).

Tabela 2: Estratégia de busca utilizada para obtenção de amostras de pacientes com DN.

Base de Dados	Parâmetros de Busca
BioProject (NCBI)	"Autism"[All Fields] OR "Autistic"[All Fields] OR "Schizophrenia"[All Fields] OR "Bipolar Disorder"[All Fields] OR "Mental Retardation"[All Fields] OR "Developmental disorder"[All Fields] OR "Developmental dyslexia"[All Fields] OR "Developmental delay"[All Fields] OR "Specific language impairment"[All Fields] OR "Autistic spectrum disorder"[All Fields] OR "Intellectual disability"[All Fields] OR "Rett syndrome"[All Fields] OR "ASD"[All Fields] AND ("transcriptome gene expression"[Filter] AND "org human"[Filter])

As amostras selecionadas referentes a DN compreendem amostras de ASD, BPD, RTT e SCZ, devido ao fato de que nestas DN já foram reportadas atividade ou expressão de L1s. Para avaliar a expressão de reguladores e retroelementos L1 as amostras, foram selecionadas considerando diferentes tipos celulares cerebrais, contemplando células-tronco embrionária (ESC), iPSC, NPC, neurônios (NE) e células cerebrais de tecidos *post mortem* (PO), estas últimas oriundas das regiões: córtex pré-frontal dorsolateral, córtex pré-frontal, lóbulo frontal e lóbulo temporal.

Tabela 3: Amostras RNA-seq obtidas através de bases de dados públicas e em colaboração.

DN: tipo de transtorno estudado: ASD (transtornos do espectro do autismo), BPD (transtorno bipolar), RTT (síndrome de Rett) ou SCZ (esquizofrenia); BioProject ID: identificador da biblioteca de amostras; Amostras CT: número de amostras do tipo controle para a biblioteca; Amostras DN: número de amostras que são caracterizadas para o DN estudado. Tipo celular: tipo celular ou origem do tecido extraído para amostras: ESC (célula-tronco embrionária), iPSC (célula-tronco pluripotente induzida), NPC (célula precursora neuronal), NE (neurônio) e PO (tecido *post mortem*).

(continua)				
DN	BioProject ID	Amostras CT	Amostras DN	Tipo Celular
ASD	PRJNA398545	39	13	PO
ASD	PRJNA143369	6	6	PO
ASD	PRJNA316853	9	9	PO
ASD	PRJNA222268 PRJNA254971	38	34	PO
ASD	PRJNA261255	28	20	NPC
ASD	PRJNA280163	11	17	iPSC
ASD	PRJNA280163	8	17	NE
ASD	PRJNA280163	9	21	NPC
BPD	PRJNA231202	31	31	PO
BPD	PRJNA253946	4	6	NE

DN	BioProject ID	Amostras CT	Amostras DN	(conclusão)
				Tipo Celular
BPD	PRJNA314463	24	30	PO
RTT(Q83X)	-	3	3	NE
RTT(N126I)	-	3	3	NE
RTT	PRJNA302685	3	3	PO
RTT	PRJNA224073	3	13	iPSC
RTT	PRJNA224073	3	-	ESC
RTT	PRJNA419983	5	4	iPSC
RTT	PRJNA419983	4	4	NE
RTT	PRJNA419983	5	4	NPC
RTT	PRJNA419983	2	2	PO
RTT	PRJNA527289	8	8	PO
SCZ	PRJNA358689	4	4	NE
SCZ	PRJNA291180	6	10	NE
SCZ	PRJNA291180	6	11	NPC
SCZ	PRJNA248182	3	6	NE
SCZ	PRJNA343829	19	19	PO
SCZ	PRJNA314463	24	28	PO
SCZ	PRJNA417295	24	24	NE
SCZ	PRJNA417295	24	23	NPC

4.4 Análise de expressão de bibliotecas de transcriptoma

Para análise de cada amostra, foram inicialmente descartadas *reads* de baixa qualidade (escore PHRED inferior a 20), e eliminadas sequências de adaptadores utilizando o NGS-QC Toolkit (PATEL; JAIN, 2012). As *reads* de alta qualidade foram mapeadas contra o genoma de referência humano (hg38) com o *software* de alinhamento de transcritos STAR (*Spliced Transcripts Alignment to a Reference*) (DOBIN et al., 2013) versão 2.5.4b. Foram utilizados alguns parâmetros adicionais para o STAR permitindo posterior utilização dos dados alinhados na análise de expressão de L1s, os quais requerem um processo de alinhamento mais exaustivo pois suas

sequências repetitivas estão dispersas em milhares de *loci* no genoma. O alinhamento considerou *reads* mapeadas em até 1×10^9 *loci* no genoma humano (`--outFilterMultimapNmax 1000000000 --winAnchorMultimapNmax 1000000000`), e alinhamentos com o melhor score como primários (`--outSAMprimaryFlag AllBestScore`).

Após a análise dos alinhamentos, foi realizada a contagem da quantidade absoluta de *reads* alinhadas para cada uma das amostras. A contagem considera *reads* alinhadas unicamente e foi realizada com o uso do *software* featureCounts versão 1.6.3 (LIAO; SMYTH; SHI, 2014), utilizando parâmetros padrão e anotação para o genoma de referência humano (hg38).

A análise de significância estatística da expressão de transcriptoma foi realizada com o pacote DESeq2 (LOVE; HUBER; ANDERS, 2014) versão 1.16.1 utilizando o *software* R (TEAM, 2016) versão 3.4.1. Para esta análise, as amostras foram normalizadas através do método RLE (*Relative Log Expression*) e a partir dos dados normalizados, foi realizado o cálculo do *fold-change* em escala *log* na base 2 e a significância estatística (p) calculada através do teste de Wald (WALD, 1945). O *fold-change* indica a variação relativa de expressão dos transcritos entre duas amostras comparadas. Estes valores são controlados por erros estatísticos do tipo 1 (probabilidade de falso positivos), calculados pelo método conhecido por *False Discovery Rate* (FDR) e ajustados com o uso de testes múltiplos de Benjamin-Hochberg (BENJAMINI; HOCHBERG, 1995). O nível de significância estatística foi considerado para $p < 0,05$.

4.5 Análise de enriquecimento relacionado a processos biológicos

A análise de enriquecimento relacionado a processos biológicos foi realizado através da base de dados STRING versão 11 (SZKLARCZYK et al., 2019). Para esta análise, foram utilizadas as moléculas catalogadas nesta pesquisa, que estão descritas na literatura como reguladores da atividade de L1. Como parâmetros, foi utilizado o escore de confiança (0,7) e para a fonte de dados, interações reportadas em bases de dados e experimentos.

De forma complementar, a análise de enriquecimento de processos biológicos foi realizada também com o uso da ferramenta IPA (*Ingenuity Pathway Analysis*) (Qiagen, Velencia, CA, USA), considerando as listas de expressão dos reguladores L1 e o *background* fornecido pelos nossos dados.

4.6 Análise de expressão de retroelementos L1

A expressão de retroelementos L1 foi calculada através do algoritmo de contagem de elementos repetitivos intercalados SQuIRE (YANG et al., 2019b). Os arquivos resultantes do alinhamento de cada amostra contra o genoma de referência humano (item 4.4) foram utilizados como entrada para o algoritmo de contagem. A fim de possibilitar uma maior detecção de *reads* alinhadas unicamente, foi considerado para o cálculo, *reads* únicos com 80% de seu comprimento alinhando-se a uma sequência de flanco de um L1. Para *reads* provenientes de sequenciamento *paired-end*, cada extremidade é avaliada individualmente para um alinhamento único antes de identificar seu par. Se uma extremidade possui um alinhamento múltiplo e a outra possuir um alinhamento único, o par será considerado único e a outra extremidade do par alinhando a múltiplos *loci* é descartado.

Após, o algoritmo atribui frações de uma *read* a cada L1 como uma função da probabilidade de que um L1 específico seja a origem dessa *read*. Para *reads* alinhadas unicamente, essa probabilidade é de 100 %, ou seja, a contagem é igual a 1. As *reads* únicas e as frações de *reads* alinhadas a múltiplos L1s são somadas e normalizadas para o comprimento do transcrito de cada L1. Por último, para refinar a estimativa de *reads* alinhadas a múltiplos L1, as contagens totais das *reads* normalizadas passam por um algoritmo de maximização de expectativa (do inglês, *Expectation Maximization* - EM). O algoritmo EM alterna entre calcular a distribuição fracionária de cada *read* múltipla para cada instância de L1 mapeada e estimar as abundâncias relativas de todos os transcritos de L1, até que as abundâncias relativas estimadas venham a convergir (L1s com contagem igual ou superior a 10 variam menos de 1%).

A base de dados contendo uma anotação de retroelementos L1 no genoma humano (item 4.1) foi utilizada pelo algoritmo de contagem para a determinação do

locus de cada um deles, e que apresentam-se de forma completa (com as regiões 5'UTR, ORF1, íntron, ORF2 e 3'UTR).

Para cada amostra de RNA-seq, foi gerada uma tabela com a contagem de cada subfamília de retroelementos L1. Posteriormente, foi aplicada análise de expressão de transcritos com significância estatística (item 4.4).

5 RESULTADOS E DISCUSSÃO

Nesta seção são apresentados e discutidos os resultados obtidos nesta pesquisa. Na seção 5.1 são apresentadas as características da base de anotações detalhadas de retroelementos L1. A partir destas anotações, na seção 5.2, é realizada a análise da expressão de L1s em DN avaliadas em diferentes tipos celulares. Na seção 5.3, é analisado detalhadamente, o perfil de moléculas reguladoras de L1 nesses mesmos DN e tipos celulares. Por fim, na seção 5.4 é explorado o contexto do processo celular da autofagia na restrição da atividade de L1s, onde é descrito a identificação da molécula AMBRA1 como um novo regulador da regulação e atividade de L1.

5.1 Base de anotação para retroelementos L1

Considerando o genoma de referência humano (hg38), foram criadas duas bases com anotações de L1: a primeira, chamada L1RG-REF, contém as anotações por subfamília das regiões do L1 (5'UTR, ORF1, íntron, ORF2, 3'UTR); a segunda, chamada L1CP-REF, contém retroelementos L1 completos (contendo todas as 5 regiões) distribuídos por subfamília. Ambas as anotações não apresentam sobreposição entre os retroelementos e possuem versões com regiões/retroelementos conservados (90% de similaridade e 70% de identidade) e regiões de retroelementos L1 altamente conservados (90% de similaridade e 99% de identidade). As anotações L1 são descritas por 12 atributos que contêm as seguintes informações: nome do cromossomo, posição inicial e final do retroelemento no cromossomo, nome da subfamília L1, nome da região da subfamília L1, sentido da inserção, número de *mismatches*, deleções e inserções, comprimento do *locus*, identidade e similaridade da sequência.

Considerando a base de anotações L1RG-REF, contendo anotações de regiões L1, observa-se uma alta concentração de regiões L1 no cromossomo X, corroborando com dados previamente relatados na literatura (BELANCIO; HEDGES; DEININGER, 2008; CHEN et al., 2005) (Figura 6A-6E). O mesmo fato é notado na base de anotações de retroelementos L1 completos, a L1CP-REF X (Figura 6F). Além disso, observa-se que a região 5'UTR possui a menor quantidade de elementos identificados, quando comparada às demais regiões de um retroelemento L1 completo (Figura 6A). Esta

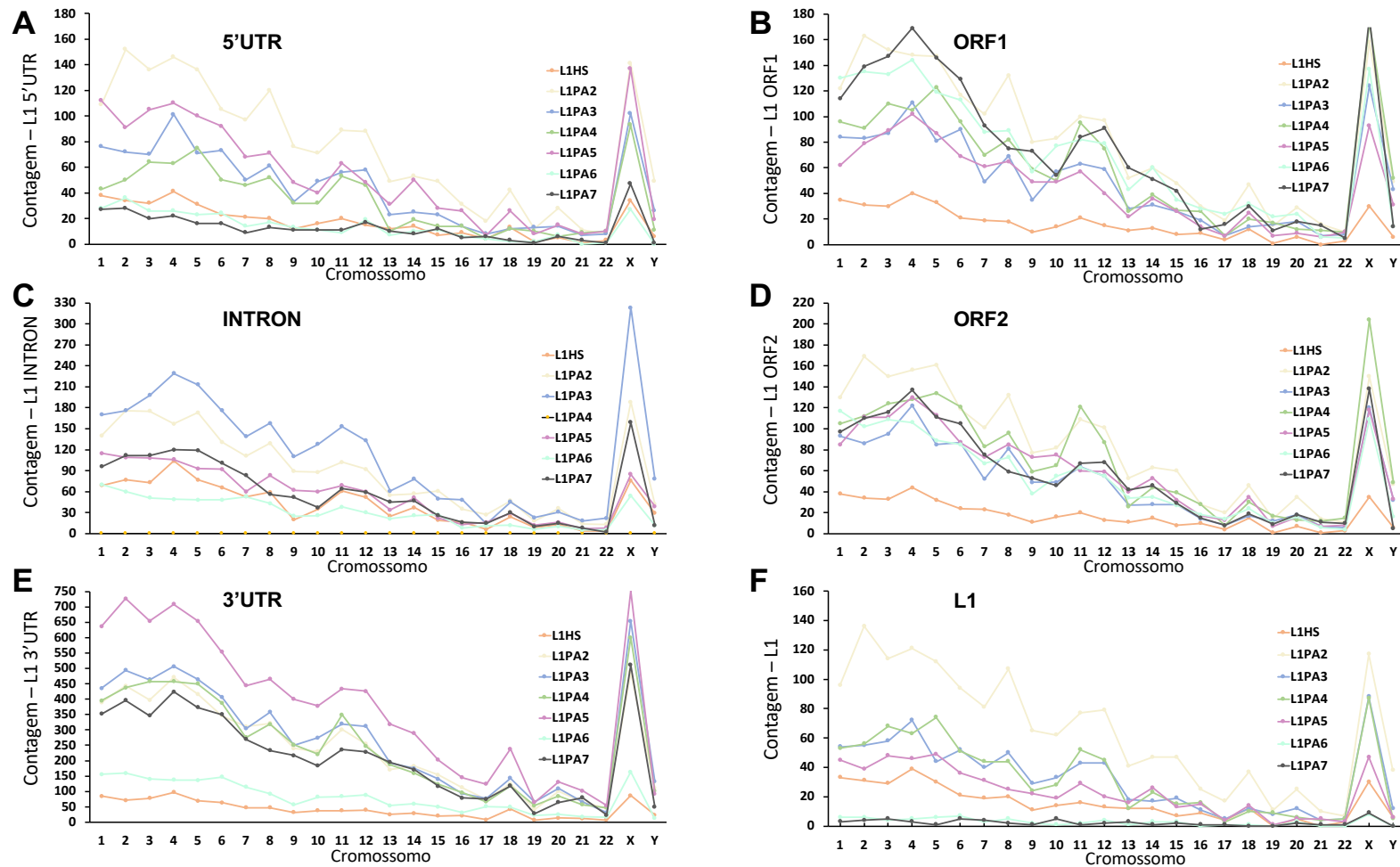


Figura 6: Número de cópias por cromossomo de regiões e retroelementos L1 completos.

Quantidade, por cromossomo, de regiões e retroelementos L1 completos divididos em subfamílias (L1HS a L1PA7), considerando sequências conservadas. **A.** Quantidade, por cromossomo, de regiões 5'UTR divididos por subfamília L1. **B.** Quantidade, por cromossomo, de regiões ORF1 divididos por subfamília L1. **C.** Quantidade, por cromossomo, de regiões íntron divididos por subfamília L1. **D.** Quantidade, por cromossomo, de regiões ORF2 divididos por subfamília L1. **E.** Quantidade, por cromossomo, de regiões 3'UTR divididos por subfamília L1. **F.** Quantidade, por cromossomo, de retroelementos L1 completos divididos por subfamília L1.

Fonte: o autor.

observação é consistente com a afirmação de que a região 5'UTR tem a maior variação a nível de sequência de nucleotídeos entre as subfamílias L1 (BOISSINOT; CHEVRET; FURANO, 2000; KHAN; SMIT; BOISSINOT, 2005; WAGSTAFF; BARNERSOI; ROY-ENGEL, 2011).

Complementarmente às bases de anotações L1CP-REF e L1RG-REF criadas sobre o genoma de referência humano, foi realizada uma análise adicional da distribuição de L1 em genomas individuais para verificar se seguem as mesmas características de distribuição encontradas para o genoma de referência humano. Desta forma, foram criadas anotações para o genoma diploide de um indivíduo Caucasiano e outro Asiático. Para cada genoma, foram criadas as bases com anotações por subfamília de regiões L1, chamada L1RG-IND (equivalente a L1RG-REF) e também de retroelementos L1 completos, chamada L1CP-IND (equivalente a L1CP-REF).

Foram comparadas as anotações para a base L1RG-REF com aquelas anotadas pela base L1RG-IND para os genomas Caucasiano e Asiático. Para a base L1RG-IND (genoma Asiático), a distribuição das regiões pelos cromossomos diploides é 98% semelhante. A mesma comparação feita para a base L1RG-IND (genoma Caucasiano) mostra menor similaridade, 83% (Figura 7). O cromossomo 22 expõe a principal diferença quando comparado ao genoma de referência humano: 70% semelhante para o genoma Asiático e 45% semelhante para o genoma Caucasiano. Ainda, para este genoma, a nível alélico, observa-se variações de menos de 1% para cada cromossomo referente a inserções de L1 (Anexo B).

Também foi investigado a distribuição de regiões anotadas contendo elementos L1HS da base L1RG-REF e que ocorrem em *loci* codificantes de proteínas (Figura 8A-8B). Foram encontradas 439 sequências conservadas e 284 sequências altamente conservadas. Destas, várias inserções estão em *loci* intrônicos de genes que são fatores de risco para autismo, como *STAG1* (*Stromal Antigen 1*) (LEHALLE et al., 2017) e *CDH13* (*Cadherin 13*) (KUSHIMA et al., 2018). Embora as inserções de L1 em *loci* intrônicos de genes não tenham um impacto direto na proteína traduzida, elas podem gerar transcrições aberrantes que podem levar ao desenvolvimento destas doenças (KAER et al., 2011; RODRÍGUEZ-MARTÍN et al., 2016).

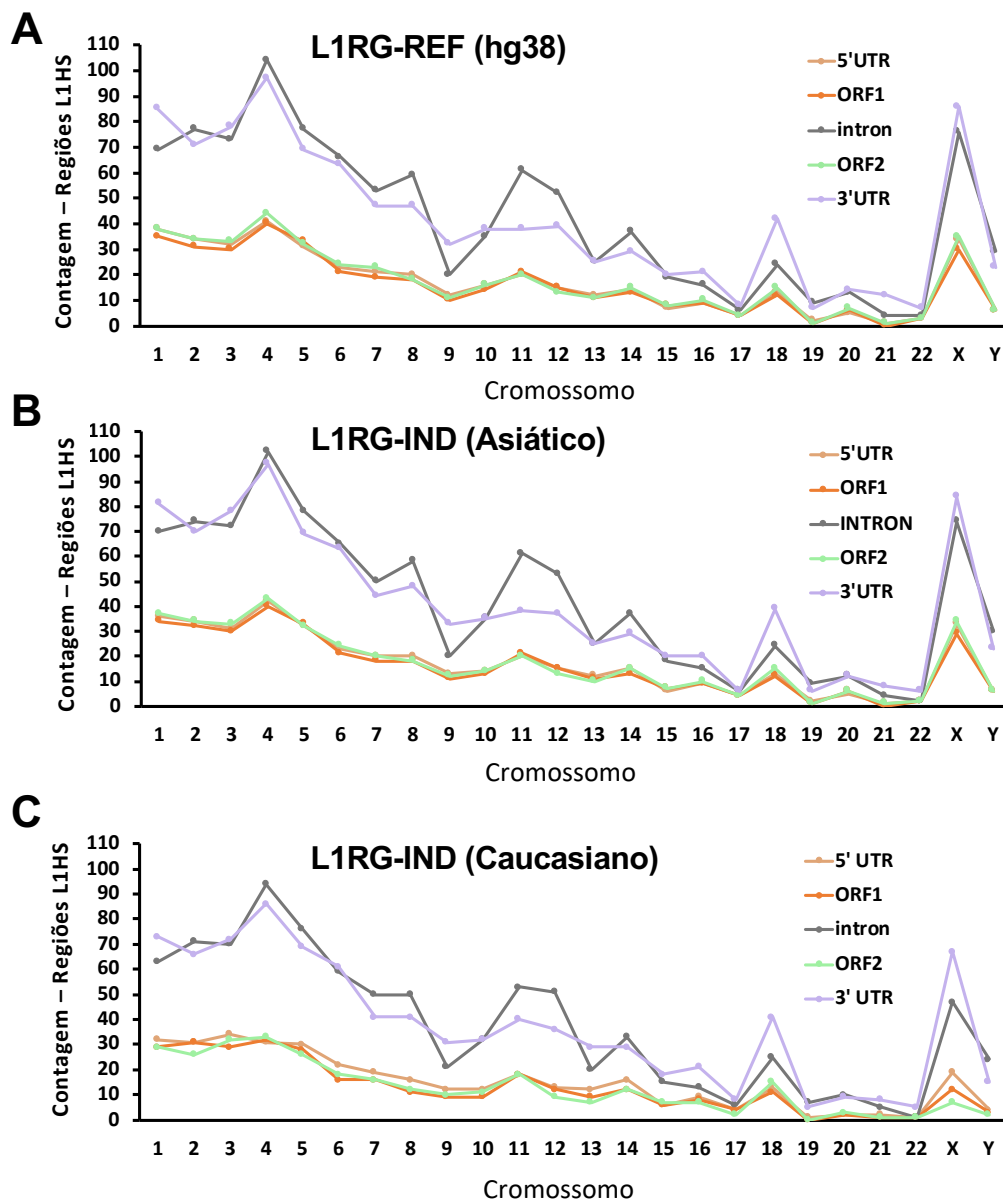


Figura 7: Distribuição por cromossomo de regiões pertencentes à subfamília L1HS entre genomas individuais e de referência.

Os gráficos apresentam anotações de regiões L1 por cromossomo, considerando sequências conservadas. **A.** Distribuição das regiões de retroelementos L1HS para o genoma de referência hg38. **B.** Distribuição das regiões de L1HS para o genoma individual Asiático. **C.** Distribuição das regiões L1HS para o genoma diploide individual Caucasiano.

Fonte: o autor.

Além das análises anteriores, também foram comparadas as anotações de regiões L1 para o genoma de referência (base L1RG-REF), com os retroelementos L1 anotados pelo RepeatMasker, que é atualmente a principal base de anotações utilizada

como referência para estudos com retroelementos e sequências repetitivas (Figura 8C-8D). O critério para considerar que um L1 está em ambas as bases de anotações é a existência de pelo menos um nucleotídeo sobreposto e a sequência anotada possuir a mesma orientação na sequência de DNA. Considerando sequências conservadas, foram identificadas 227 regiões L1 que não são relatadas pelo RepeatMasker (Figura 8C). Esses 227 elementos são caracterizados como pertencentes a várias famílias extintas, como oriundos de primatas (L1PB) e outros mamíferos (L1MA). Como exemplo, para regiões L1 que apresentam maior variação em sua sequência de nucleotídeos, como a região 5'UTR, observa-se através da nossa anotação uma sequência pertencente à subfamília L1PA5 (hg38, chrX: 152,006,773-152,007,798), possuindo identidade de 94,2% e similaridade de 99,6%. Esta mesma sequência é reportada pelo RepeatMasker como um retroelemento L1 diferente e pertencente a mamíferos, o L1MA4A. Em outro caso, uma região relatada pela nossa anotação como uma ORF2 (hg38, chr2: 199.073.688-199.077.484) pertencente à subfamília L1PA7 (identidade de 94% e similaridade de 99,3%), é reportada pelo RepeatMasker com duas anotações distintas: uma sequência L1MD2 e outra sequência repetitiva simples (AT)_n. Para sequências altamente conservadas, foram reportadas apenas 2 regiões não informadas pelo RepeatMasker (Figura 8D). Uma delas, uma região 3'UTR (hg38, chr10:37.226.120-37.226.329) exposta pela nossa anotação como pertencente a subfamília L1PA5 (identidade de 99,0% e similaridade de 99,0%), é reportada pelo RepeatMasker como pertencente a um retroelemento L1MB8.

As bases de anotações L1 construídas, L1RG-REF, L1CP-REF, L1RG-IND e L1CP-IND, contêm elementos L1 caracterizados e anotados com informações detalhadas, incluindo anotações de sequência baseadas em similaridade. Através do uso destas anotações de L1 em humanos, foi realizada a análise de expressão de retroelementos L1 em DN, exposta na próxima seção.

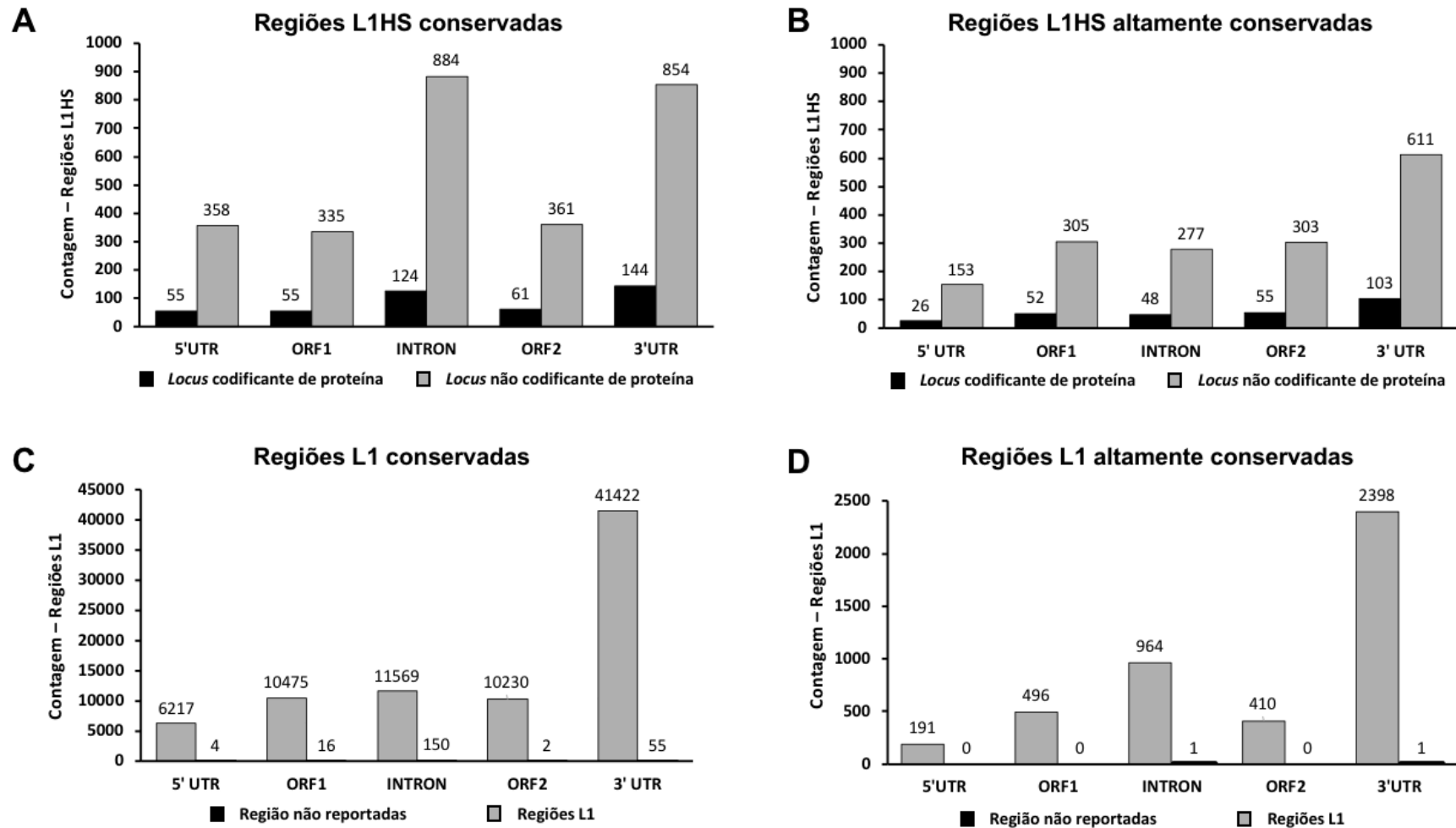


Figura 8: Comparação entre a base de anotações de regiões L1 e anotações de L1 reportados pelo RepeatMasker.

A. Distribuição de regiões L1HS em *loci* genômicos codificantes de proteínas considerando sequências conservadas. **B.** Distribuição de regiões L1HS em *loci* genômicos codificantes de proteínas considerando sequências altamente conservadas. **C.** Interseção entre regiões L1 anotadas e elementos relatados pelo RepeatMasker considerando sequências conservadas. **D.** Interseção entre regiões L1 anotadas e elementos relatados pelo RepeatMasker considerando sequências altamente conservadas.

Fonte: o autor.

5.2 Análise de expressão de retroelementos L1 em distúrbios do neurodesenvolvimento

A análise de expressão de L1 realizada neste trabalho forneceu um panorama geral da expressão destes retroelementos em diferentes tipos celulares (PO, NE, iPSC, NPC e ESC) em 327 amostras de DN (ASD, BPD, RTT e SCZ) (Figura 9). A expressão de L1 é caracterizada através da subfamília que contém retroelementos capazes de realizar a retrotransposição (atividade) em humanos, L1HS, e também da expressão de L1 total, que considera a expressão de todas as subfamílias L1 reunidas (L1HS-L1PA7). Por fim, foram considerados somente *loci* de L1 que possuem sequência conservada.

A partir da análise dos dados, observou-se alta expressão de transcritos L1 em todos os tipos celulares considerados (Figura 9A). Os valores de média (μ) e desvio padrão (σ) para transcritos L1 em amostras de PO, NE, iPSC, NPC e ESC são, respectivamente: ($\mu = 35137$, $\sigma = 4530$), ($\mu = 33745$, $\sigma = 2361$), ($\mu = 37998$, $\sigma = 9082$), ($\mu = 33507$, $\sigma = 1953$) e ($\mu = 32412$, $\sigma = 1094$). Também foram observados elevados níveis de expressão de transcritos L1HS em células cerebrais: ($\mu = 3000$, $\sigma = 1747$) em PO, ($\mu = 3152$, $\sigma = 1113$) em NE, ($\mu = 4176$, $\sigma = 2639$) em iPSC, ($\mu = 2576$, $\sigma = 1021$) em NPC e ($\mu = 4271$, $\sigma = 341.0$) em ESC (Figura 9B).

O aumento de expressão de transcritos L1 está associado ao aumento da atividade de L1 (BENITEZ-GUIJARRO et al., 2018; MUOTRI et al., 2010; THOMAS et al., 2017). Desta forma, nossas observações sobre a expressão de L1 são suportadas através da análise da atividade de L1 realizadas por outros estudos. Coufal e colegas observaram número de cópias de L1 significativamente mais altos em tecidos provenientes de cérebros humanos adultos saudáveis em comparação com tecidos de fígado e coração destes mesmos indivíduos (COUFAL, 2009). Também demonstraram que em NPC isoladas de células-tronco do cérebro fetal humano e de ESC, através do uso da técnica de reação em cadeia da polimerase quantitativa (qPCR), um aumento de aproximadamente 80 cópias de ORF2 de L1 por célula com base em amostras controle (COUFAL, 2009). Um estudo anterior corrobora com estes dados sobre retrotransposição de L1 em NPC. Muotri e colegas, através de um ensaio baseado em plasmídeo, observou que L1s, quando expressos em células multipotentes como NPC, são capazes de

retrotransposição (MUOTRI et al., 2005). Complementando estes estudos, a atividade de L1s também foi observada em células pluripotentes, como em iPSC e ESC. Nestas células, a atividade de L1 é facilitada, devido a baixa metilação do DNA quando comparada a células diferenciadas (MARCHETTO et al., 2013; WISSING et al., 2012).

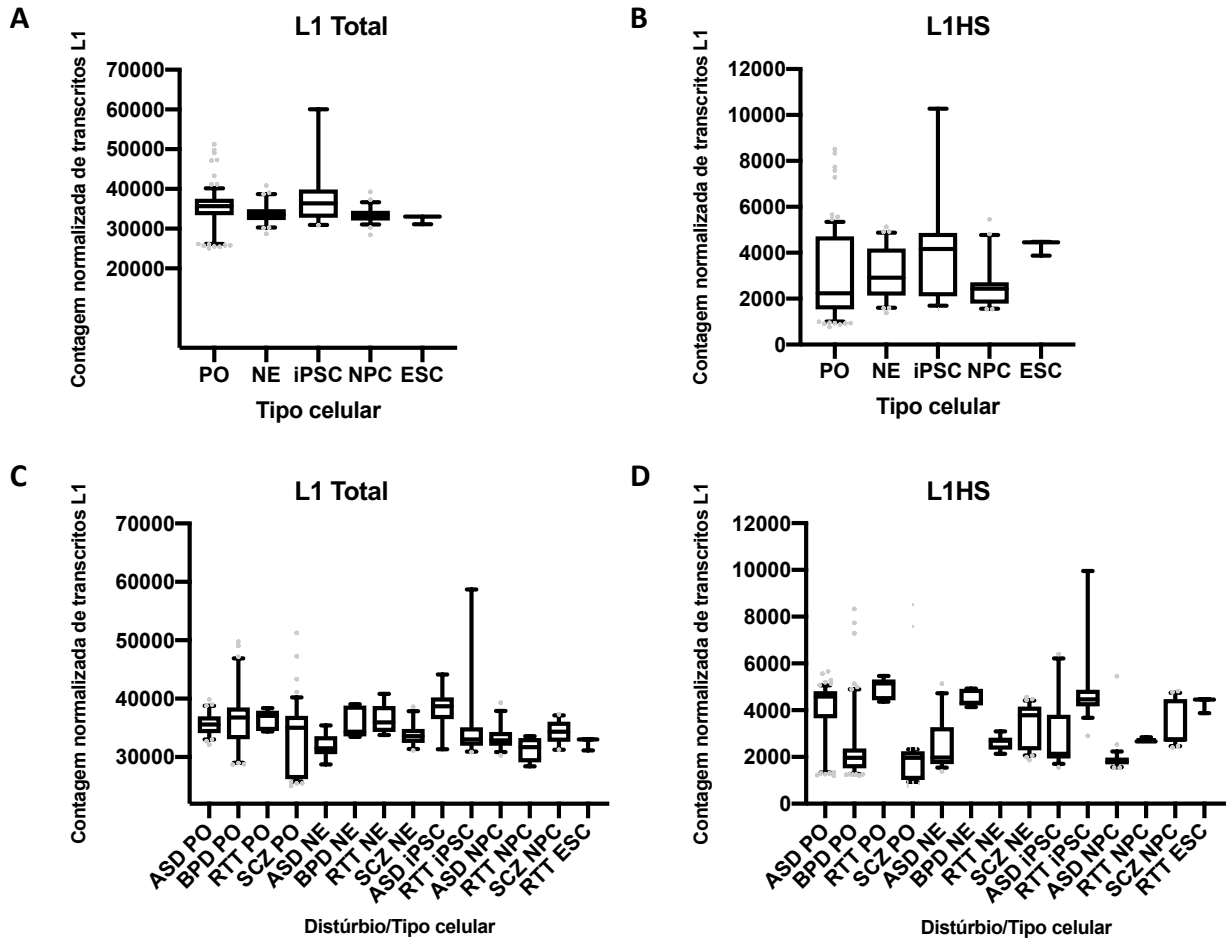


Figura 9: Expressão de L1 total e L1HS em DN e tipos celulares.

O gráfico de barras mostra o percentil 25-75 (retângulo), a mediana (linha horizontal dentro do retângulo), e o percentil 5-95 (*whisker*) da expressão normalizada (RLE) de L1 total (subfamílias L1HS-L1PA7) e L1HS para DN (ASD, BPD, RTT e SCZ) e tipo celulares (PO, NE, iPSC, NPC e ESC). **A.** Expressão de transcritos de L1 total por tipo celular. **B.** Expressão de transcritos de L1HS por tipo celular. **C.** Expressão de transcritos de L1 total por DN e tipo celular. **D.** Expressão de transcritos de L1HS por DN e tipo celular.

Fonte: o autor.

Considerando a expressão de L1 total (subfamílias L1HS-L1PA7), observa-se uma variação média de 6% para a expressão de transcritos L1 entre todos os distúrbios analisados (Figura 9C). Além disso, para a expressão de transcritos da subfamília de L1, a L1HS, é notada uma variação média de 33% entre os distúrbios (Figura 9D). Para os

mesmos DN, foi realizada uma comparação com os respectivos controles para tecidos *post mortem* (Figura 10A) e neurônios em cultura *in vitro* (Figura 10B). Embora não sejam observadas variações significativas da expressão, eventos relacionados a atividade de L1 em DN podem ocorrer com maior frequência devido a desregulações epigenéticas, como a diminuição de metilação do DNA (AMIR et al., 1999; LI et al., 2018; MISIAK et al., 2015). De fato, através do uso da técnica qPCR, Bundo e colegas observaram números aumentados de cópias ORF2 de L1 em neurônios do córtex pré-frontal de pacientes com SCA (BUNDO et al., 2014). Além disso, um número aumentado de cópias L1 foi encontrado em neurônios oriundos de cultura *in vitro* de pacientes com SCZ contendo a deleção 22q11 (BUNDO et al., 2014). Outro estudo, mostrou que em ASD e RTT, a atividade de L1HS estava aumentada cerca de 8,5 vezes em cérebros de pacientes com esses DN comparada a controles (JACOB-HIRSCH et al., 2018).

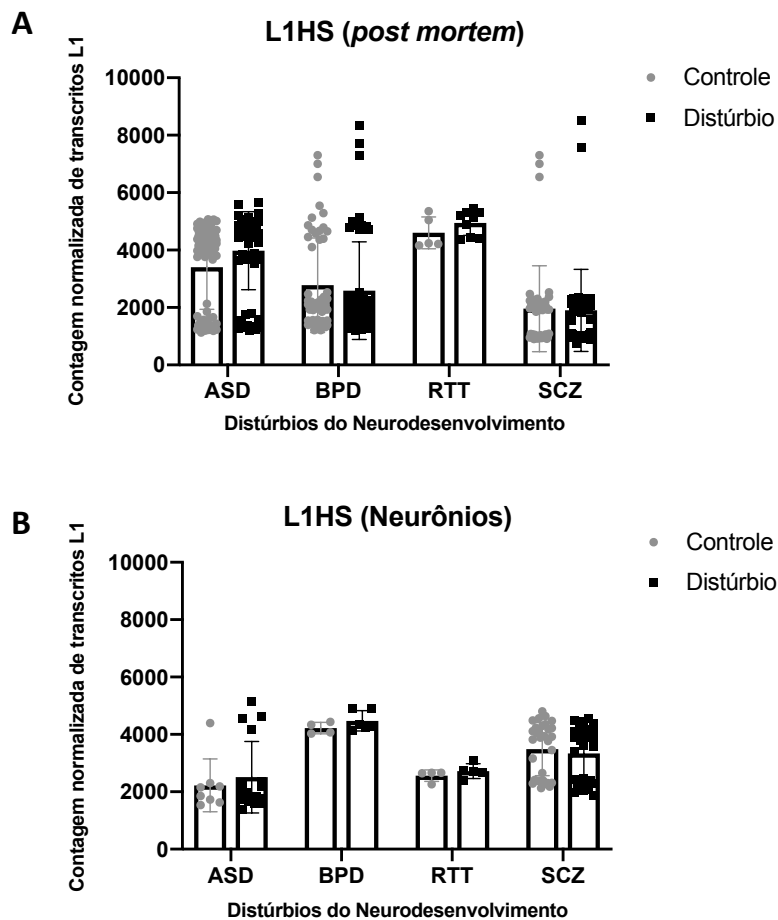


Figura 10: Expressão de L1HS em amostras controle e de DN.

Os gráficos mostram valores de expressão de transcritos L1HS para amostras controle e amostras de DN (ASD, BPD, RTT e SCZ). **A.** Expressão de transcritos de L1HS em células provenientes de tecidos *post*

mortem para amostras controle e amostras de DN. **B.** Expressão de transcritos de L1HS em amostras de neurônios em cultura *in vitro* para amostras controle e amostras de DN.

Fonte: o autor.

Os dados gerados demonstraram alta expressão de transcritos L1 em todos os tipos celulares cerebrais e distúrbios do neurodesenvolvimento. Através da desregulação de moduladores da expressão e atividade de L1, podem ocorrer novas inserções destes retroelementos no genoma nuclear, bem como a mobilização de outros retroelementos, o que pode levar ao aumento da instabilidade genômica. A próxima seção explora o perfil de expressão de genes relacionados ao controle da expressão e atividade de retroelementos L1, fornecendo um perfil da regulação de L1s em diferentes DN e tipos celulares.

5.3 Expressão de genes reguladores de L1 em distúrbios do neurodesenvolvimento

Nesta seção é explorado o perfil de expressão de genes reguladores de L1, agrupados por DN e avaliados em diferentes tipos celulares e etapas de desenvolvimento celular (Figura 19 a 38 e Apêndice C – Tabela 7 a 10). É importante frisar que os genes reguladores foram identificados a partir de uma extensa revisão sistemática da literatura, que considerou todos os trabalhos publicados no tema, incluindo análise de sinônimos e termos associados a área para garantir qualidade na busca. A partir disso, para cada conjunto de genes reguladores de L1 reportados neste trabalho, foi realizada uma análise de enriquecimento de vias metabólicas. Nesta análise, foram levantados processos biológicos contendo reguladores L1 com expressão alterada em DN.

Uma das vias encontradas é a via de reparo de DNA que está associada com a ocorrência da anemia de Fanconi (Figura 11). A anemia de Fanconi é um distúrbio hereditário que gera instabilidade genômica (MOLDOVAN; D'ANDREA, 2009), e a via descrita é necessária para o reparo eficiente de DNA danificado, especialmente de ligações cruzadas entre cadeias de DNA (*crosslinking*) (KANEHISA et al., 2019). Acredita-se que a via da anemia de Fanconi coordene um mecanismo molecular que

envolve elementos de vias clássicas de reparo do DNA, como recombinação homóloga e reparo por excisão de nucleotídeos (MOLDOVAN; D'ANDREA, 2009).

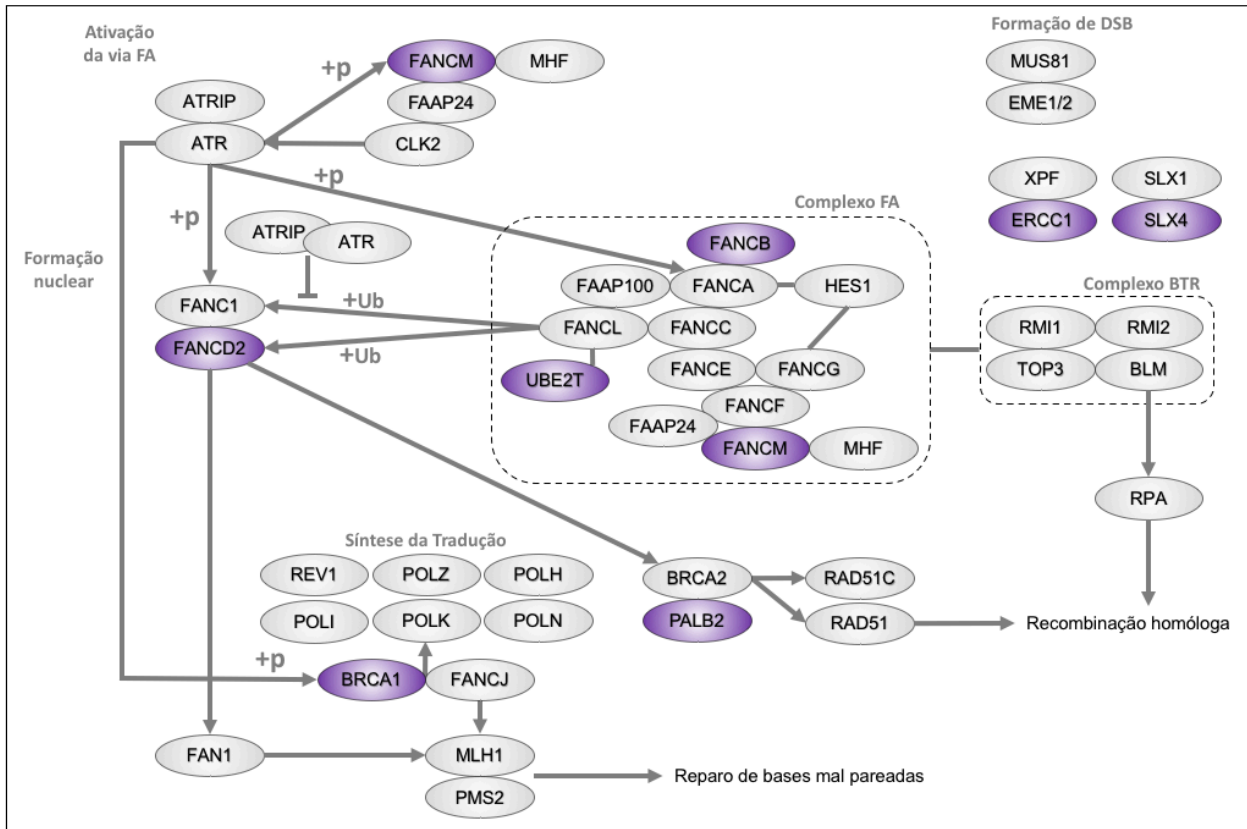


Figura 11: Reguladores L1 com expressão alterada na via da anemia de Fanconi.

O DNA danificado, é diretamente reconhecido pela proteína FANCM e outras proteínas associadas como a CLK2 (*CDC Like Kinase 2*), que recrutam o complexo central da via (Complexo FA). Este complexo monoubiquitina FANCD2 e FANCI (*FA Complementation Group I*), e estes interagem com uma série de proteínas de reparo de DNA, como BRCA2 (*BRCA2 DNA Repair Associated*) e FAN1 (*FANCD2 And FANCI Associated Nuclease 1*), ativando outras vias de reparo como a via de recombinação homóloga. Moléculas em roxo apresentam reguladores L1 com expressão de transcritos alterada em DN.

Fonte: adaptado de (KANEHISA et al., 2019).

Os genes reguladores de L1 com expressão significativamente alterados nesta via são: *UBE2T* (*Ubiquitin Conjugating Enzyme E2 T*), *FANCB* (*FA Complementation Group B*) e *FANCM* (*FA Complementation Group M*), que participam do complexo central desta via; *FANCD2* (*FA Complementation Group D2*), necessário para a ativação de outras moléculas, como a *PALB2* (*Partner And Localizer Of BRCA2*), dentro da cascata de eventos necessários para a ativação de outras vias de reparo de DNA como a recombinação homóloga; e os genes *ERCC1* (*ERCC Excision Repair 1, Endonuclease Non-Catalytic Subunit*) e *SLX4* (*SLX4 Structure-Specific Endonuclease Subunit*),

envolvidos no reparo de quebra de fita dupla de DNA. Os perfis de expressão para estes genes, considerando tipos celulares e DN são expostos nas Figuras 21, 25 e 36.

Outro processo biológico encontrado e também relacionado ao reparo de DNA, é a via de reparo por excisão de nucleotídeos (Figura 12). O papel desta via é atuar no reconhecimento e reparo a danos no DNA (em células eucarióticas, fragmentos de DNA de 24-32 nt) ocasionados por diferentes fontes como as causadas por compostos, carcinógenos ambientais e exposição à luz ultravioleta (PETRUSEVA; EVDOKIMOV; LAVRIK, 2014).

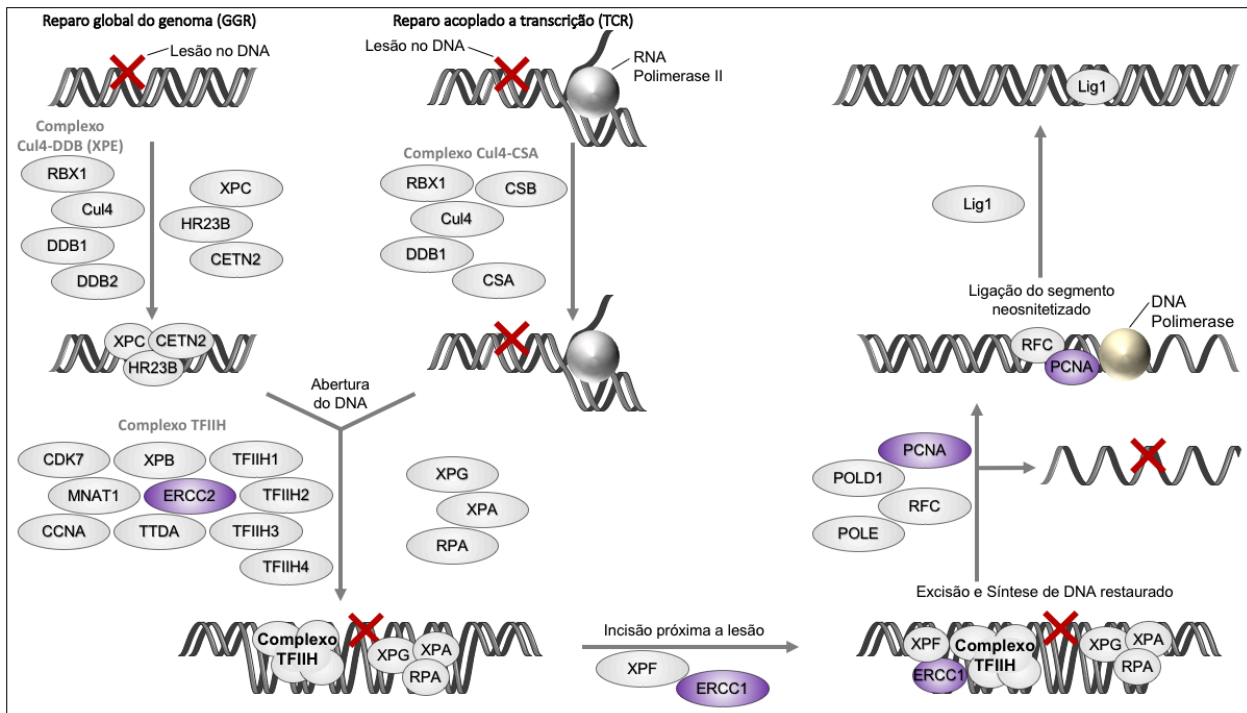


Figura 12: Reguladores L1 com expressão alterada na via de reparo de DNA por excisão de nucleotídeos.

A via de reparo do DNA por excisão de nucleotídeos envolve duas sub-vias diferentes: reparo acoplado à transcrição (TCR) e reparo global do genoma (GGR). O GGR pode ocorrer em qualquer parte do genoma, enquanto o TCR é responsável pelo reparo acelerado de lesões na fita transcrita de genes ativos pela RNA polimerase II (RNAP II). O TCR é iniciado pela RNA polimerase estacionada sobre a lesão com a ajuda de fatores específicos como CSA (*ERCC Excision Repair 8, CSA Ubiquitin Ligase Complex Subunit*) e CSB (*ERCC Excision Repair 6, Chromatin Remodeling Factor*). No GGR, a primeira etapa do reconhecimento de danos, envolve o complexo XPC-HR23B junto com o complexo XPE. Após, o complexo de helicase TFIIH e XPC-HR23B são essenciais para abertura do DNA em torno do dano. A proteína RPA (*Replication Protein A1*), estabiliza o complexo de DNA aberto e está envolvida no posicionamento das endonucleases XPG (*XPG-Complementing Protein*) e ERCC1-XPF responsáveis pelas incisões de DNA ao redor da lesão. A próxima etapa envolve a síntese de DNA através da atividade de POLD1 (*DNA Polymerase Delta 1, Catalytic Subunit*), POLE (*DNA Polymerase Epsilon, Catalytic Subunit*) e seus cofatores PCNA e RFC (*Replication Factor C*). Moléculas em roxo apresentam reguladores L1 com expressão de transcritos alterada em DN.

Fonte: adaptado de (KANEHISA et al., 2019).

São reportados nesta via genes com expressão significativamente alterada em DN como o gene *ERCC2* (*ERCC Excision Repair 2*, *TFIIF Core Complex Helicase Subunit*), componente essencial do complexo TFIIF, *ERCC1* (*ERCC Excision Repair 1*, *Endonuclease Non-Catalytic Subunit*), que juntamente com a endonuclease XPF (*DNA Repair Endonuclease XPF*), são necessários para clivar o DNA ao redor da lesão, e *PCNA* (*Proliferating Cell Nuclear Antigen*), envolvido na síntese do DNA (perfis de expressão destes genes são expostos na Figura 22).

A próxima via relacionada ao reparo de DNA e com reguladores L1 alterados é a via de reparo por recombinação homóloga (Figura 13). A recombinação homóloga é essencial para o reparo preciso de quebras de fita dupla de DNA, sendo indispensável para a integridade do genoma (KANEHISA et al., 2019; SUNG; KLEIN, 2006).

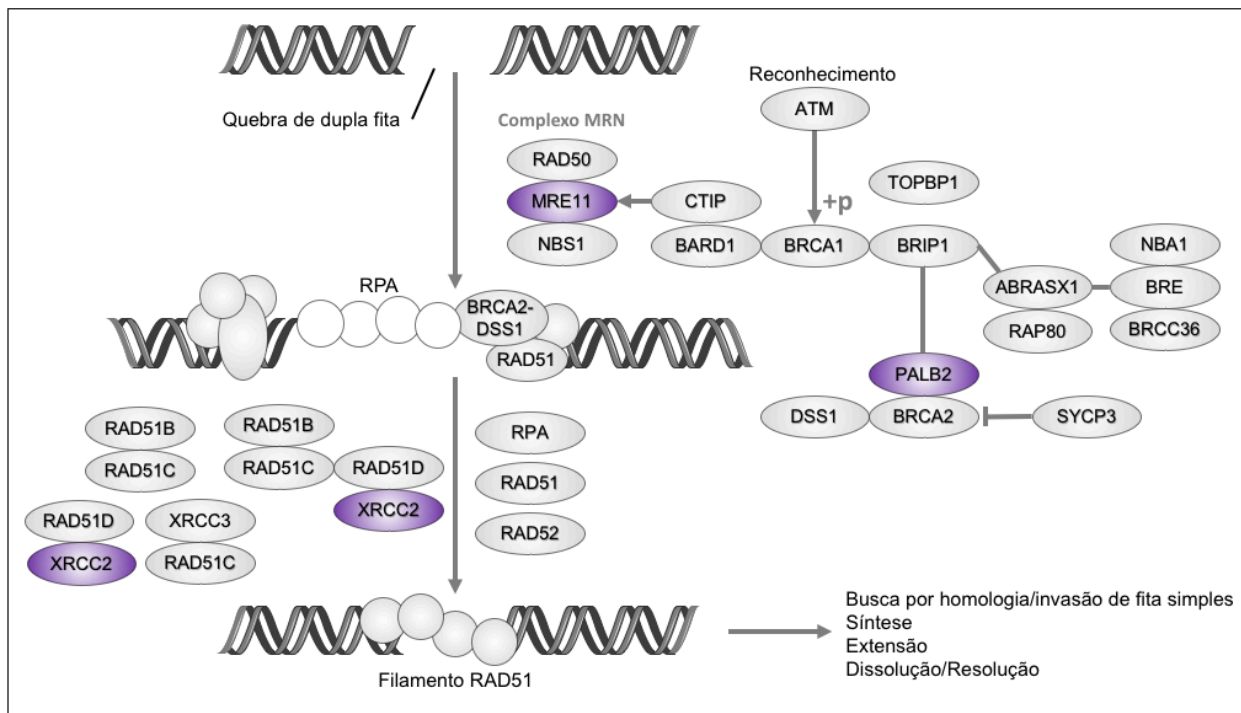


Figura 13: Reguladores L1 com expressão alterada na via de reparo de DNA por recombinação homóloga.

A escolha da via de reparo adequada é mediada por BRCA1 (*BRCA1 DNA Repair Associated*). Inicialmente, as pontas da quebra de fita dupla são reconhecidas. Após, ocorre a ressecção destes filamentos através do complexo MRN formado pelas moléculas MRE11, RAD50 (*RAD50 Double Strand Break Repair Protein*) e NBS1 (*Nibrin*). Moléculas, como a proteína CTIP (*RB Binding Protein 8, Endonuclease*), criam saliências no DNA, que são revestidas pela proteína RPA (*Replication Protein A1*) (círculos preenchidos em branco). Durante a recombinação homóloga canônica, a RPA é removida por RAD51 (círculos preenchidos em cinza). A formação do filamento RAD51 é regulado positivamente por BRCA2, PALB2 e os parálogos de RAD51 como as proteínas XRCC2 e XRCC3 (*X-Ray Repair Cross Complementing 3*). Demais etapas envolvem o processo de invasão da fita homóloga de DNA a ser

reparada e processos de dissolução ou resolução. Moléculas em roxo apresentam reguladores L1 com expressão de transcritos alterada em DN.
Fonte: adaptado de (KANEHISA et al., 2019).

O gene MRE11 (*MRE11 Homolog, Double Strand Break Repair Nuclease*), participante do complexo MRN, e os genes essenciais para o processo de reparo, PALB2 e XRCC2 (*X-Ray Repair Cross Complementing 2*) (BUISSON; MASSON, 2012; JOHNSON; LIU; JASIN, 1999), encontram-se com expressão alterada nas amostras de DN (Figura 28).

Por fim, a última via exposta com expressão alterada de genes reguladores L1 e relacionada ao reparo de DNA, é a via da união de extremidade não-homóloga (Figura 14). Esta via tem o papel de reparar quebras de fita dupla de DNA as quais levam à instabilidade ou diversidade genômica (BÉTERMIER; BERTRAND; LOPEZ, 2014).

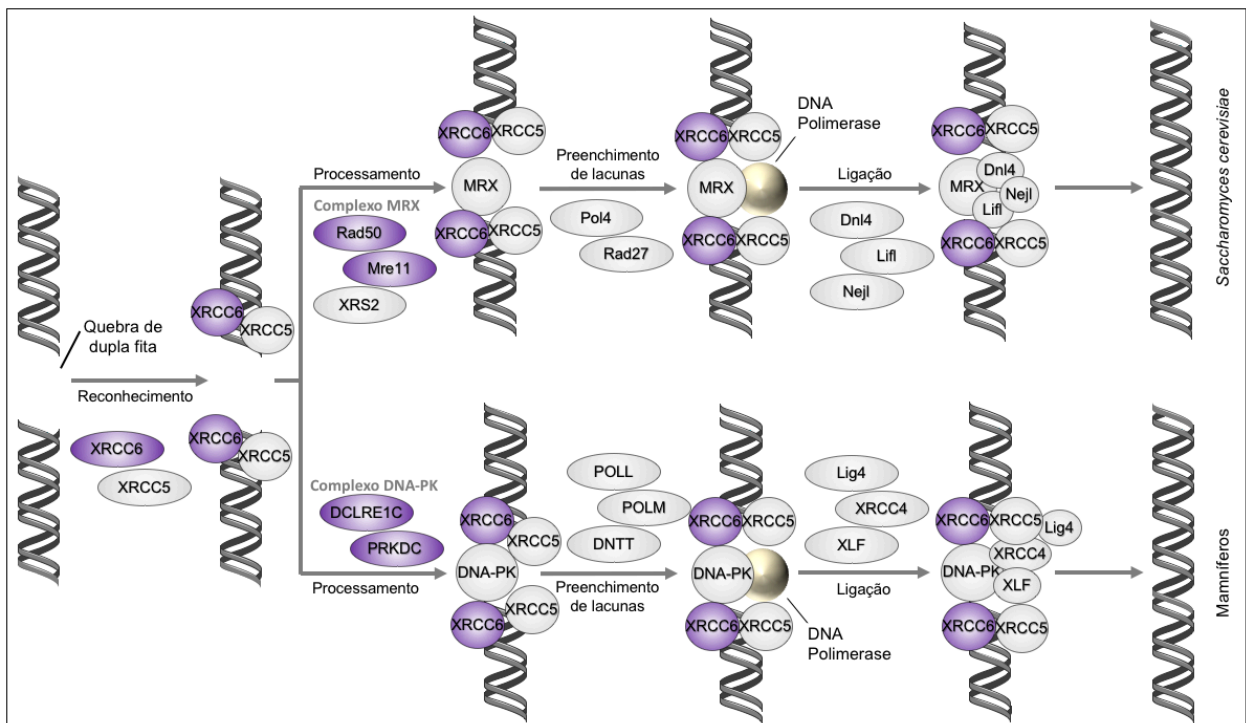


Figura 14: Reguladores L1 com expressão alterada na via da união de extremidade não homóloga. A via da união de extremidade não-homóloga repara quebras de fita dupla de DNA em todos os estágios do ciclo celular. Envolve a ligação de XRCC5 (*X-Ray Repair Cross Complementing 5*), e XRCC6 às extremidades de DNA de fita dupla, recrutamento de DNA-PKs (complexo MRX em levedura), processamento de extremidades e recrutamento do complexo DNA ligase IV (LIG4) - XRCC4 (*X-Ray Repair Cross Complementing 4*). Moléculas em roxo apresentam reguladores L1 com expressão de transcritos alterada em DN.
Fonte: adaptado de (KANEHISA et al., 2019).

Nesta via, os genes reguladores de L1 com expressão significativamente alterada são *XRCC6* (*X-Ray Repair Cross Complementing 6*), *DCLRE1C* (*DNA Cross-Link Repair 1C*), *PRKDC* (*Protein Kinase, DNA-Activated, Catalytic Subunit*), *RAD50* e *MRE11* (perfil de expressão destes genes são expostos na Figura 32 e 38).

Além das diversas vias de reparo de DNA, a análise revelou a via de senescência celular (Figura 15) associada com os reguladores de L1 alterados nos dados de DN. A senescência celular é um estado de parada celular irreversível e pode ser desencadeada por uma série de fatores, como encurtamento do telômero, irradiação, dano ao DNA e estresse oxidativo (KANEHISA et al., 2019).

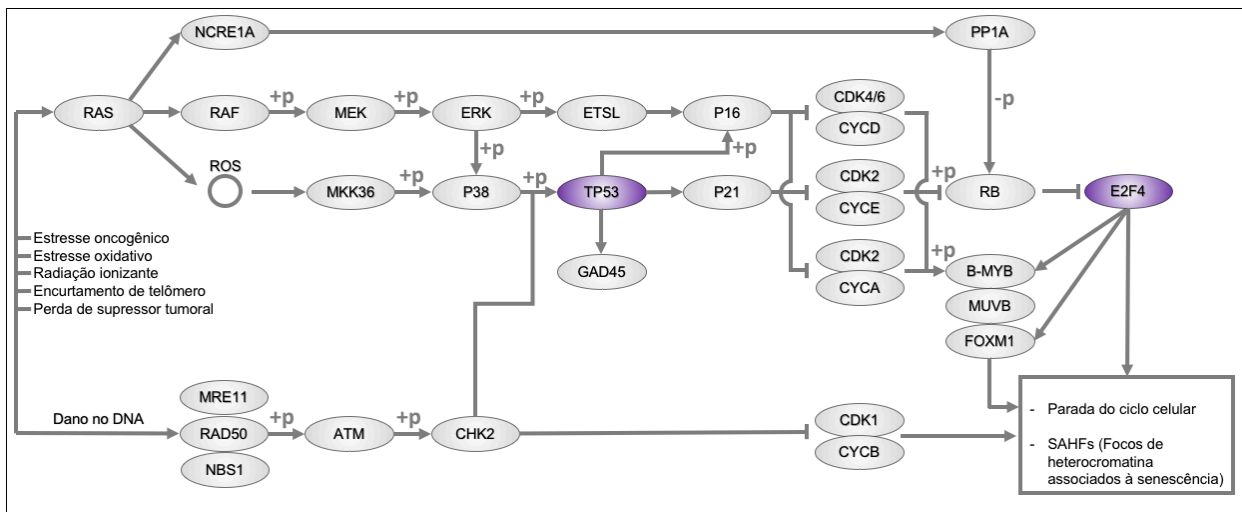


Figura 15: Reguladores L1 com expressão alterada na via de senescência celular.

As proteínas TP53 e RB (*RB Transcriptional Corepressor 1*), são críticos na indução da senescência. A proteína TP53 pode iniciar o processo de senescência ativando RB através de P21 (*Cyclin Dependent Kinase Inhibitor 1A*) ou P16 (*Cyclin Dependent Kinase Inhibitor 2A*). A proteína RB ativa a senescência por inibir a transcrição de genes alvo de E2F como o E2F4 (BEN-PORATH; WEINBERG, 2005). Moléculas em roxo apresentam reguladores L1 com expressão de transcritos alterada em DN.

Fonte: adaptado de (KANEHISA et al., 2019).

Os genes essenciais para esta via e com alteração significativa de expressão observada em DN são TP53 (*Tumor Protein P53*) e o fator de transcrição E2F4 (*E2F Transcription Factor 4*) (perfil de expressão destes genes são expostos na Figura 34).

Outro processo biológico com alteração significativa de expressão de genes nas amostras de DN comparada com controles é a via de mitofagia. Nesta via são reportados 5 genes apresentando expressão alterada nas análises realizadas (Figura 16). Esta via é ativada quando ocorre a disfunção mitocondrial, que pode ser causada por subprodutos deletérios do processo de fosforilação oxidativa, chamados de espécies reativas ao

oxigênio (do inglês, *Reactive Oxygen Species* - ROS). Se o dano for excessivo para ser reparado, as mitocôndrias são seletivamente reconhecidas e direcionadas para degradação por um modo específico de autofagia, denominado mitofagia.

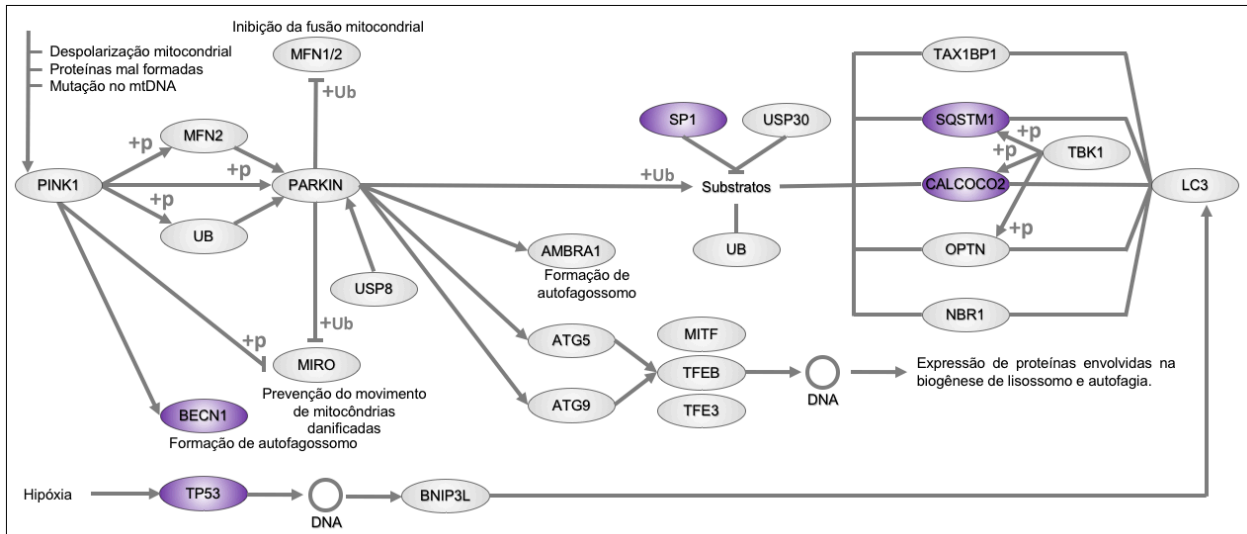


Figura 16: Reguladores L1 com expressão alterada na via de mitofagia.

A indução da mitofagia (através de fatores como a despolarização da membrana mitocondrial) envolve a quinase PINK1 (*PTEN Induced Kinase 1*) e a ligase E3 Parkin. PINK1 serve como sensor para a despolarização mitocondrial e recruta a proteína PARKIN (*Parkin RBR E3 Ubiquitin Protein Ligase*), seguido pela ubiquitinação de receptores de mitofagia como CALCOCO2. Moléculas em roxo apresentam reguladores L1 com expressão de transcritos alterada em DN.

Fonte: adaptado de (KANEHISA et al., 2019).

Os genes *BECN1*, *TP53*, *SP1* (*Sp1 Transcription Factor*), *SQSTM1* e *CALCOCO2*, essenciais para a ativação da via e remoção de mitocôndrias danificadas, são reportados com alteração significativa na expressão de transcritos em DN (Figuras 29, 35 e 37).

A análise realizada também identificou a via da biogênese de piRNAs com reguladores de L1 alterados (Figura 17). Os piRNAs são RNAs curtos e não codificantes, com tamanho variando entre 26 a 30 nt (em vertebrados e *D. melanogaster*) (KIRINO; MOURELATOS, 2007; SAITO et al., 2007), e que possuem funções como restringir a atividade de vários tipos de retroelementos (ARAVIN et al., 2007; GHILDIYAL; ZAMORE, 2009; MALONE; HANNON, 2009), e também atuam na regulação de mRNAs, possivelmente através do recrutamento da proteína CCR4 (*Carbon Catabolite Repressor, protein 4*) (ROUGET et al., 2010).

Os genes participantes desta via e com expressão significativamente alterada em DN são: *PLD6* (*Phospholipase D Family Member 6*), *HENMT1* (*HEN Methyltransferase 1*), *PIWIL2* (*Piwi Like RNA-Mediated Gene Silencing 2*), *PIWIL4* (*Piwi Like RNA-Mediated*

Gene Silencing 4), *TDRD1* (Tudor Domain Containing 1), *TDRD9* (Tudor Domain Containing 9), *MOV10L1* (Mov10 Like RISC Complex RNA Helicase 1), *MAEL* (Maelstrom Spermatogenic Transposon Silencer) e *HSP90AA1* (Heat Shock Protein 90 Alpha Family Class A Member 1) (Figuras 23, 26 e 30).

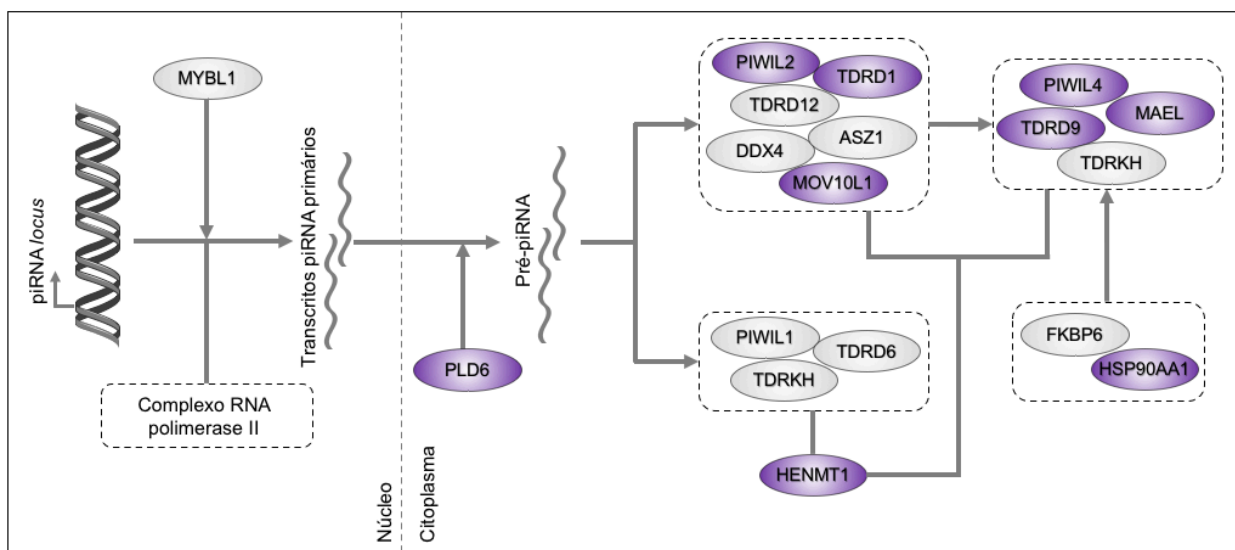


Figura 17: Reguladores L1 com expressão alterada na via da biogênese de piRNAs.

Em *Mus musculus*, o fator de transcrição MYBL1 (*MYB Proto-Oncogene Like 1*) se liga a uma região promotora canônica e inicia a transcrição do piRNA primário. Ao mesmo tempo, MYBL1 também inicia a expressão de genes da via de piRNA (por exemplo, *PIWIL2*, *PLD6*). Os transcritos primários possuem cap 5' e cauda poli(A). Após, os transcritos primários são exportados do núcleo, e processados por PLD6 e demais moléculas envolvidas na biogênese de piRNAs. Moléculas em roxo apresentam reguladores L1 com expressão de transcritos alterada em DN.

Fonte: adaptado de (FABREGAT et al., 2018; WEICK; MISKA, 2014).

Por fim, a via de sinalização de interferon alfa/beta (IAB) é reportada com expressão alterada de reguladores de L1 em DN (Figura 18). Esta via é responsável por iniciar uma cascata de eventos em resposta à infecções virais (TEIJARO, 2016).

Os genes induzidos por interferons *RNASEL* e *SAMHD1* (*SAM And HD Domain Containing Deoxynucleoside Triphosphate Triphosphohydrolase 1*) apresentam alteração significativa de expressão nas DN analisadas (Figura 33).

Nas diversas vias analisadas, foram observados reguladores de L1 com expressão alterada e exercendo papéis essenciais nas atividades destas vias. Nos próximos itens, são apresentados, de forma individual, o perfil de expressão destes genes reguladores de L1 e também seu papel na regulação de retroelementos L1.

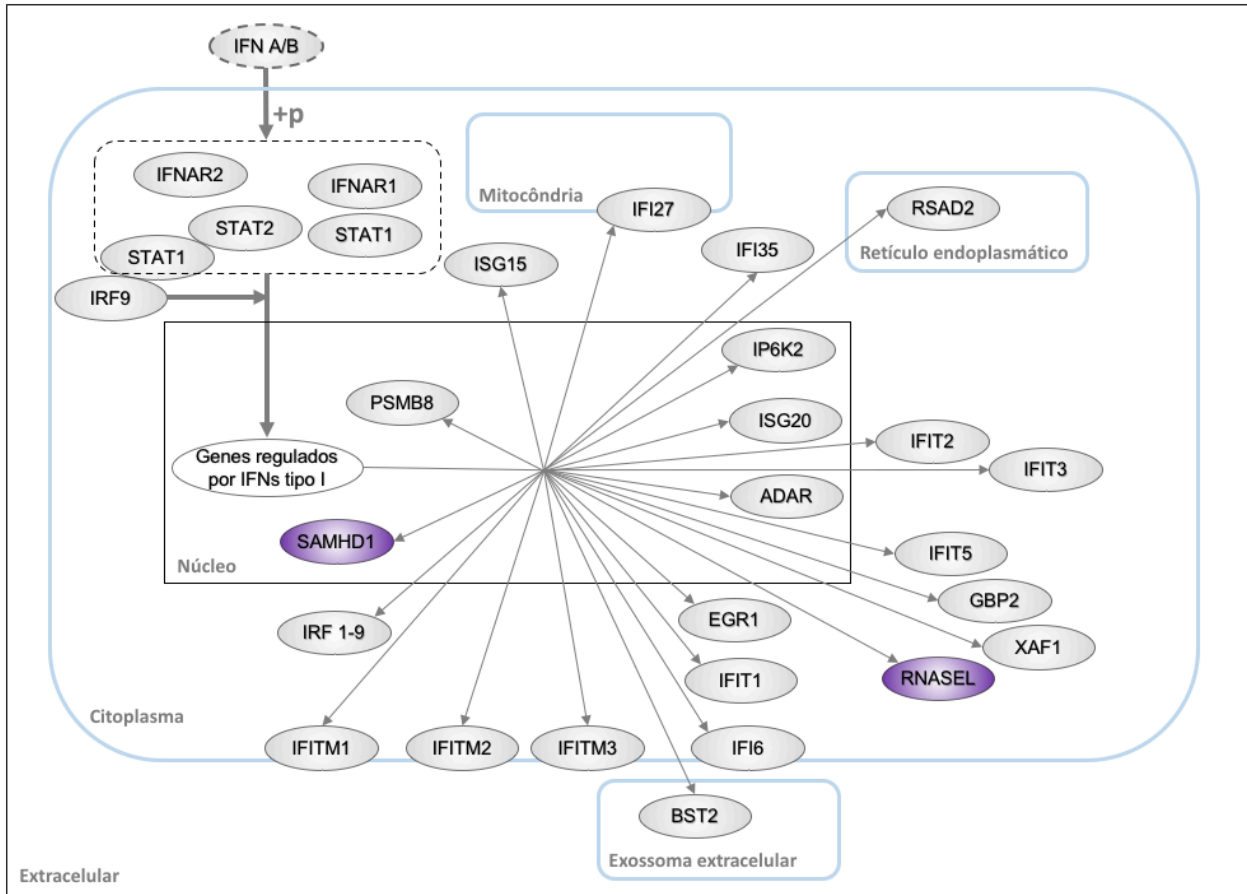


Figura 18: Reguladores L1 com expressão alterada na via de sinalização de interferon alfa/beta.

A estimulação do complexo de interferons (IFN) tipo I (IFNA/B), leva a ativação do complexo transcrricional de ISGF3, que compreende moléculas como STAT1 (*Signal Transducer And Activator Of Transcription 1*), STAT2 (*Signal Transducer And Activator Of Transcription 1*) e IRF9 (*Interferon Regulatory Factor 9*). Após, este complexo ativa vários genes regulados por interferons como *RNASEL* e *SAMHD1*. Moléculas em roxo apresentam reguladores L1 com expressão de transcritos alterada em DN.

Fonte: adaptado de (FABREGAT et al., 2018).

5.3.1 Expressão de reguladores L1 em tecidos cerebrais *post mortem* de DN

O perfil de expressão dos reguladores de L1 é exibido através do diagrama de Venn (Figura 19). Nele, as interseções indicam os reguladores que apresentam o mesmo perfil de expressão entre diferentes DN (expressão aumentada ou reduzida em amostras para um determinado DN quando comparado com os respectivos controles). Observa-se dois genes que apresentam o mesmo padrão de expressão compartilhado entre ASD, SCZ, BPD e RTT em células provenientes de tecidos cerebrais *post mortem* (Figura 19 e Apêndice C – Tabela 7).



Figura 19: Padrão de expressão de moléculas reguladoras de L1 em DN para células provenientes de tecidos cerebrais *post mortem*.

O Diagrama de Venn representa o número de reguladores L1 com um mesmo perfil de expressão (aumentada/reduzida) em amostras de ASD, SCZ, BPD e RTT em células provenientes de tecidos cerebral *post mortem*. Variação de expressão (*fold-change*) entre amostras controle e de DN considerando $p < 0,05$. Fonte: o autor.

O primeiro gene, *CALCOCO2*, é um receptor de autofagia sendo responsável pelo início do processo de degradação de transcritos L1 co-localizados em corpos de processamento (*P-Body*) (GUO et al., 2014). Em todos os DN, é notada expressão significativamente aumentada de transcritos deste gene quando comparado aos respectivos controles (Figura 20).

Também é observada expressão significativamente aumentada nas amostras de DN para o segundo gene, o fator de transcrição *SOX2* (*SRY-Box 2*), o qual é associado a repressão da atividade e expressão de L1s (Figura 20). Um estudo observou que a redução na expressão de *SOX2* durante os estágios iniciais da diferenciação neuronal, leva a aumentos na expressão e atividade de L1 (MUOTRI et al., 2005). Cabe ressaltar que esta restrição de L1 ocorre também durante a neurogênese em idade adulta (MUOTRI et al., 2005).

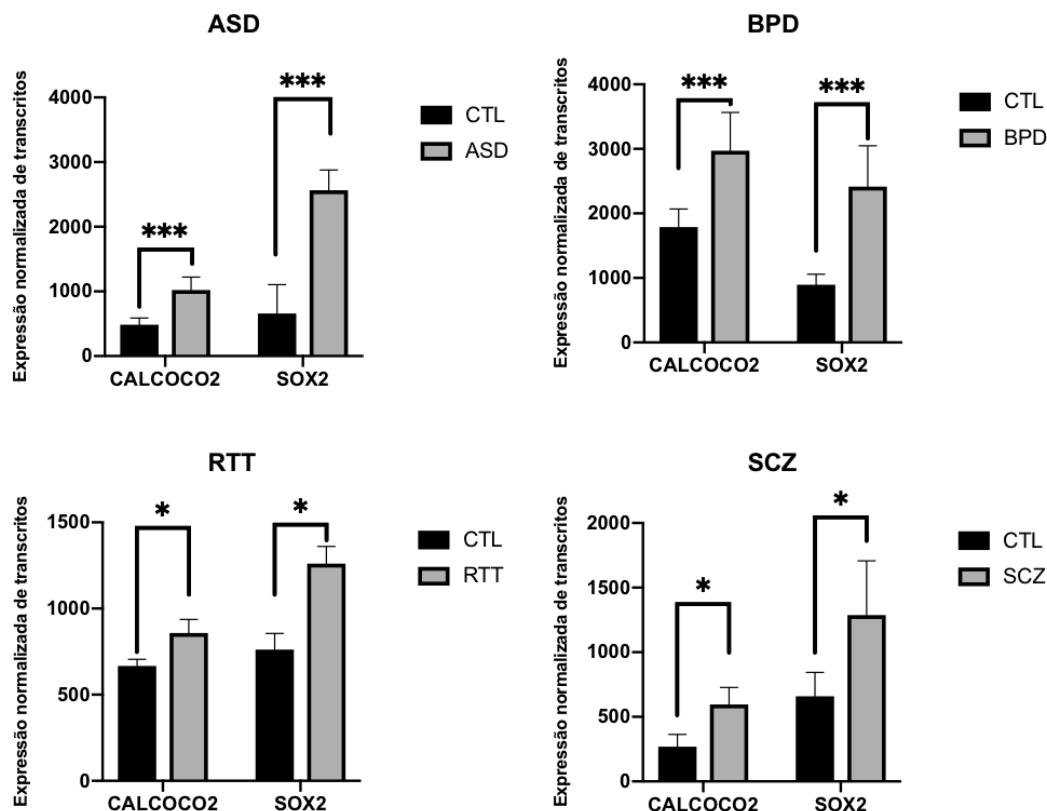


Figura 20: Reguladores L1 com mesmo perfil de expressão em ASD, BPD, RTT e SCZ.

Os gráficos mostram valores de expressão normalizada (RLE) de transcritos para amostras controle e amostras de ASD, BPD, RTT e SCZ para os genes *CALCOCO2* e *SOX2*. Os gráficos de barras mostram a média $\pm$ SD (desvio padrão). *P*-value: * $p < 0,05$; *** $p < 0,01$; **** $p < 0,001$.

Fonte: o autor.

Através da análise de enriquecimento de vias com 49 reguladores de L1 apresentando perfis de expressão alterados em ASD, observa-se duas vias enriquecidas: via da anemia de Fanconi (Figura 11) e a via de reparo de DNA por excisão de nucleotídeos (Figura 12). Na via da anemia de Fanconi, são observados os seguintes reguladores L1 com expressão significativamente alterada em ASD: *UBE2T*, *FANCB*, *SLX4*, *PALB2* (Figura 21). Todos estes reguladores atuam reprimindo a expressão de L1 (LIU et al., 2018; MITA et al., 2020).

Para a via de reparo de DNA por excisão de nucleotídeos, são reportados os reguladores L1, *ERCC2* e *ERCC1* com expressão significativamente reduzida em ASD (os mesmos também participam da via da anemia de Fanconi exposta anteriormente), e o regulador de L1 *PCNA*, apresentando expressão significativamente aumentada em ASD (Figura 22). *ERCC1* e *ERCC2* atuam de forma a reprimir a expressão e atividade

de L1 (GASIOR; ROY-ENGEL; DEININGER, 2008). De forma antagônica, *PCNA* coopera de forma positiva na atividade de L1 (TAYLOR et al., 2013).

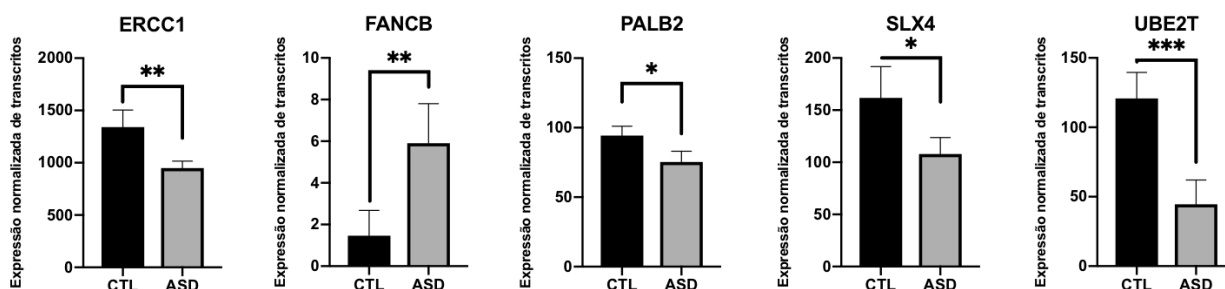


Figura 21: Reguladores L1 em ASD participantes da via da anemia de Fanconi.

Os gráficos mostram valores de expressão normalizada (RLE) de transcritos para amostras controle e amostras de ASD para os genes *ERCC1*, *FANCB*, *PALB2*, *SLX4* e *UBE2T*, participantes da via da anemia de Fanconi. Os gráficos de barras mostram a média $\pm$ SD (desvio padrão). *P-value*: '*' $p < 0,05$; '**' $p < 0,01$; '***' $p < 0,001$.

Fonte: o autor.

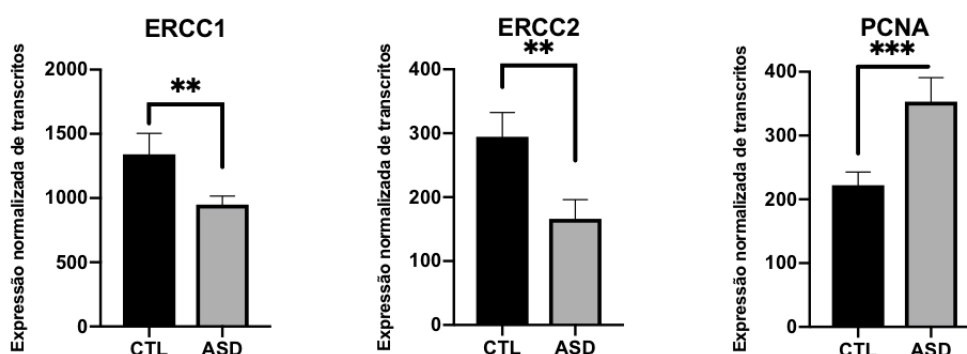


Figura 22: Reguladores L1 em ASD participantes da via de reparo de DNA por excisão de nucleotídeos.

Os gráficos mostram valores de expressão normalizada (RLE) de transcritos para amostras controle e amostras de ASD para os genes *ERCC1*, *ERCC2* e *PCNA*, participantes da via de reparo de DNA por excisão de nucleotídeos. Os gráficos de barras mostram a média $\pm$ SD (desvio padrão). *P-value*: '*' $p < 0,05$; '**' $p < 0,01$; '***' $p < 0,001$.

Fonte: o autor.

São observados também, reguladores L1 com expressão significativamente aumentada em BPD e que atuam na via da biogênese de piRNAs (Figura 17). Os reguladores de L1 encontrados nesta via restringem sua expressão e atividade: *HSP90AA1*, *PIWIL2*, *PIWIL4* e *TDRD9* (Figura 23). Foi observado que em camundongos *knockout* para o gene *Hsp90aa1*, que é codificado para a proteína Hsp90 α , apresenta aumento da proteína ORF1p de L1 (ICHIYANAGI et al., 2014). Outro estudo também observou que o gene *PIWIL2* reprime a atividade de L1s em iPSC de humanos e oócitos de camundongos (LIM et al., 2013; MARCHETTO et al., 2013; WATANABE et al., 2008).

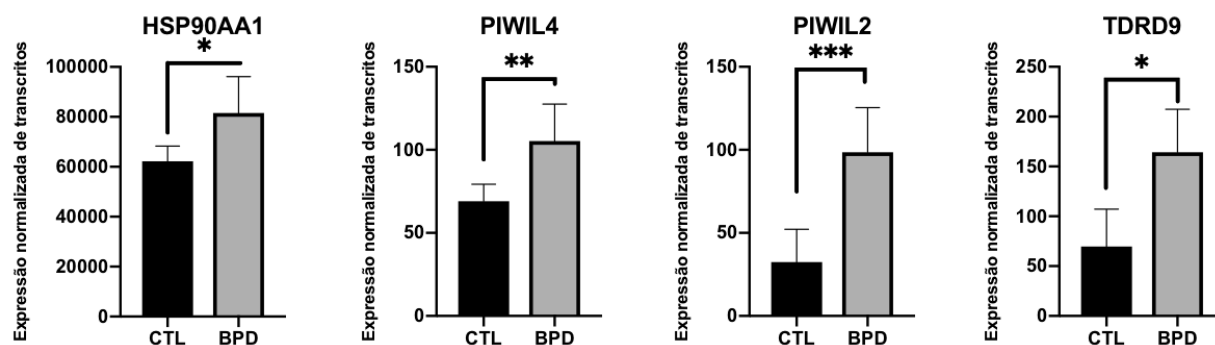


Figura 23: Reguladores L1 em BPD participantes da via da biogênese de piRNAs.

Os gráficos mostram valores de expressão normalizada (RLE) de transcritos para amostras controle e amostras de BPD para os genes *HSP90AA1*, *PIWIL4*, *PIWIL2* e *TDRD9*, participantes da via da biogênese de piRNAs. Os gráficos de barras mostram a média $\pm$ SD (desvio padrão). *P*-value: '*' $p < 0,05$; '**' $p < 0,01$; '***' $p < 0,001$.

Fonte: o autor.

5.3.2 Expressão de reguladores L1 em iPSC de DN

Para reguladores de L1 em ASD e RTT em células iPSC não foram observados reguladores em comum. Para reguladores de L1 em RTT, seis genes apresentam expressão significativamente alterada (Figura 24 e Apêndice C – Tabela 8). Um deles, o gene *TEX19* (*Testis Expressed 19*) apresenta expressão aumentada nas amostras analisadas, é específico de mamíferos e possui expressão restrita a células-tronco pluripotentes e em células germinativas (KUNTZ et al., 2008). Também, em camundongos, *TEX19* reprime a expressão de L1s em ESC a nível pós-transcricional (OLLINGER et al., 2008).

A análise também identificou os reguladores de L1, *UBE2T* e *FANCD2*, com perfil de expressão significativamente aumentada em RTT (Figura 25), e que também fazem parte da via da anemia de Fanconi (Figura 11). A literatura reporta que em células germinativas de camundongos, *Fancd2* co-ocupa o promotor de L1, reprimindo sua expressão e atividade. Evidências sugerem que essa repressão provavelmente envolve marcas de histonas H2A/H4R3me2 catalisadas por *Fancd2* e pelo gene *Prmt5* (*Protein Arginine Methyltransferase 5*) (NIE et al., 2020). Além disso, a proteína FANCD2 também é importante para ativação do gene *SLX4* (BRÉGNARD et al., 2016), outro regulador de L1, e ambos atuam reprimindo a atividade de L1 (NIE et al., 2020).

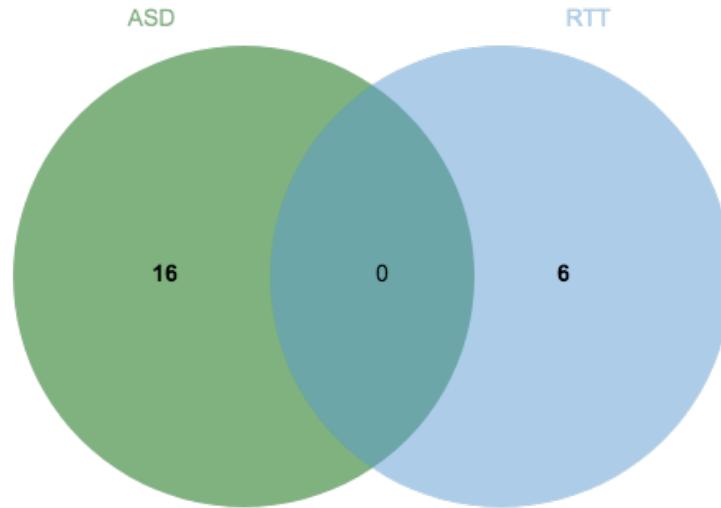


Figura 24: Padrão de expressão de moléculas reguladores de L1 em iPSC de DN.

O Diagrama de Venn representa o número de reguladores L1 com um mesmo perfil de expressão (aumentada/reduzida) em amostras de ASD, RTT em iPSC. Variação de expressão (*fold-change*) entre amostras controle e de DN considerando $p < 0,05$.

Fonte: o autor.

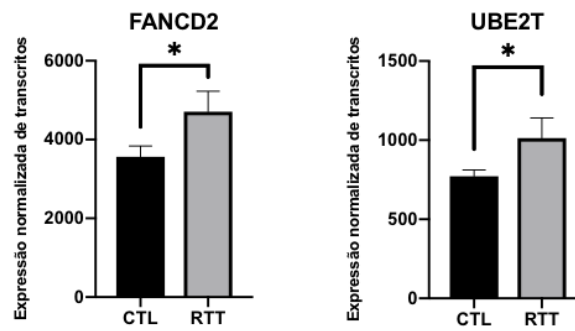


Figura 25: Reguladores L1 em RTT participantes da via da anemia de Fanconi.

Os gráficos mostram valores de expressão normalizada (RLE) de transcritos para amostras controle e amostras de RTT para os genes *FANCD2* e *UBE2T*, participantes da via da anemia de Fanconi. Os gráficos de barras mostram a média $\pm$ SD (desvio padrão). *P*-value: '*' $p < 0,05$; '**' $p < 0,01$; '***' $p < 0,001$.

Fonte: o autor.

Para reguladores de L1 com perfis de expressão significativamente alterados em ASD, são observados 16 genes com expressão alterada (Figura 24). Através da análise de enriquecimento de vias, os genes com expressão significativamente aumentada, *HENMT1*, *MAEL*, *MOV10L1* e *TDRD1* (Figura 26), participam do processo de biogênese de piRNA (Figura 17). Estes genes são reportados como repressores da atividade de L1 (LIM et al., 2015; SHOJI et al., 2009; SOPER et al., 2008).

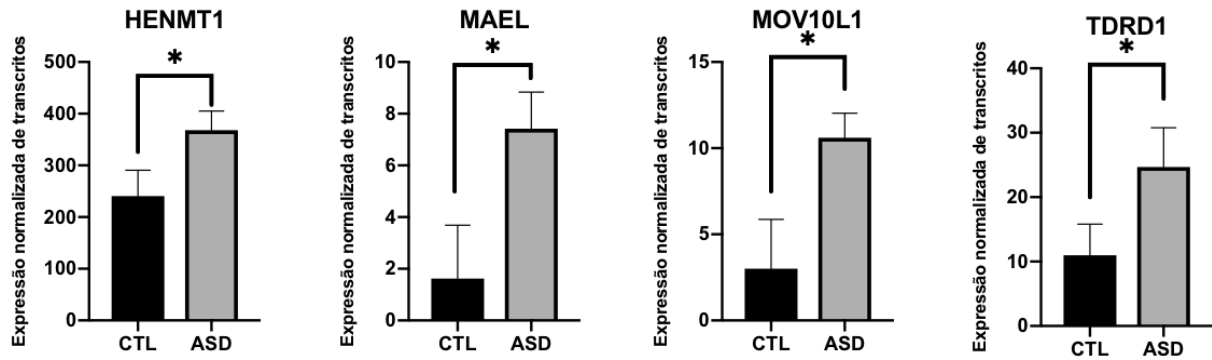


Figura 26: Reguladores L1 em ASD participantes da via da biogênese de piRNAs.

Os gráficos mostram valores de expressão normalizada (RLE) de transcritos para amostras controle e amostras de ASD para os genes *HENMT1*, *MAEL*, *MOV10L1* e *TDRD1*, participantes da via da biogênese de piRNAs. Os gráficos de barras mostram a média ± SD (desvio padrão). *P*-value: '*' $p < 0,05$; '**' $p < 0,01$; '***' $p < 0,001$.

Fonte: o autor.

5.3.3 Expressão de reguladores L1 em NPC de DN

Em NPC, cinco genes são reportados com expressão significativamente alterada comuns a ASD e SCZ (Figura 27 e Tabela 10). Tais genes não foram reportados até o momento apresentando desregulação dentro dos distúrbios abordados para este tipo celular. Um destes genes, *ZC3HAV1* (*Zinc Finger CCCH-Type Containing, Antiviral 1*), possui capacidade de inibir a retrotransposição de L1 através da sua interação com o transcrito de L1, o que ocorre em agregados citoplasmáticos que se co-localizam com proteínas em grânulos de estresse (GOODIER et al., 2015; MOLDOVAN; MORAN, 2015).

Para reguladores com expressão alterada em SCZ, a análise reporta os genes *MRE11A*, *PALB2* e *XRCC2* com expressão significativamente aumentada nesta DN (Figura 28). A análise de enriquecimento de vias expõe estes genes atuando na via de reparo de DNA por recombinação homóloga (Figura 13).

Também são observados os genes *BECN1*, *CALCOCO2* e *TP53* apresentando expressão significativamente reduzida em ASD (Figura 29). Na análise realizada, estes genes participam da via de mitofagia (Figura 16).

Outros dois genes, *PIWIL4* e *PLD6*, são reportados com expressão significativamente alterada em ASD (Figura 30) e, para esta análise, estão associados com a via de biogênese de piRNAs (Figura 17). Estes genes apresentam efeito inibitório

à atividade de L1 em diferentes tipos celulares (GOODIER et al., 2015; KOJIMA-KITA et al., 2016; REUTER et al., 2011; WATANABE et al., 2011).

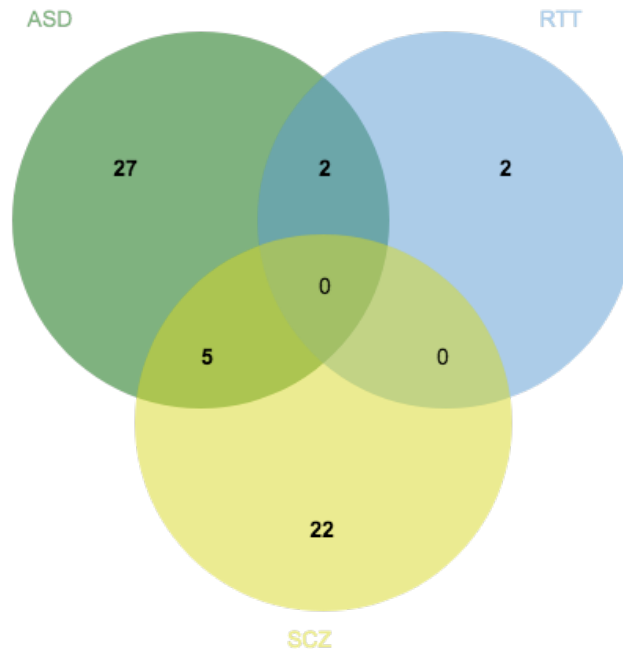


Figura 27: Padrão de expressão de moléculas reguladores de L1 em DN para NPC.

O Diagrama de Venn representa o número de reguladores L1 com um mesmo perfil de expressão (aumentada/reduzida) em amostras de ASD, SCZ e RTT em NPC. Variação de expressão (*fold-change*) entre amostras controle e de DNs considerando $p < 0,05$.

Fonte: o autor.

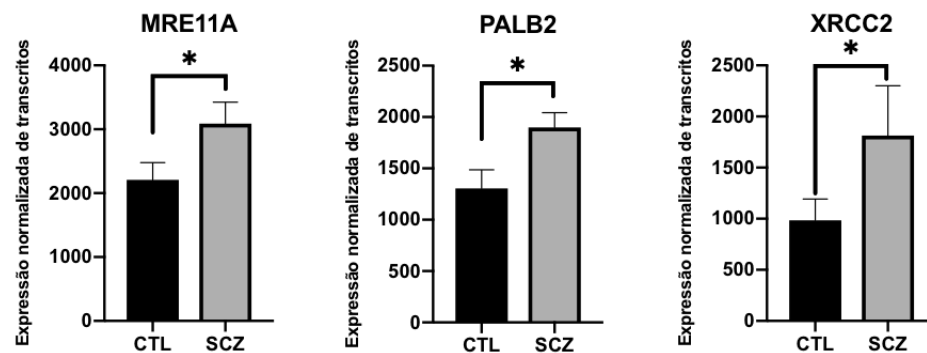


Figura 28: Reguladores L1 em SCZ participantes da via de reparo de DNA por recombinação homóloga.

Os gráficos mostram valores de expressão normalizada (RLE) de transcritos para amostras controle e amostras de SCZ para os genes *MRE11A*, *PALB2* e *XRCC2* participantes da via de reparo de DNA por recombinação homóloga. Os gráficos de barras mostram a média $\pm$ SD (desvio padrão). *P-value*: '*' $p < 0,05$; '**' $p < 0,01$; '***' $p < 0,001$.

Fonte: o autor.

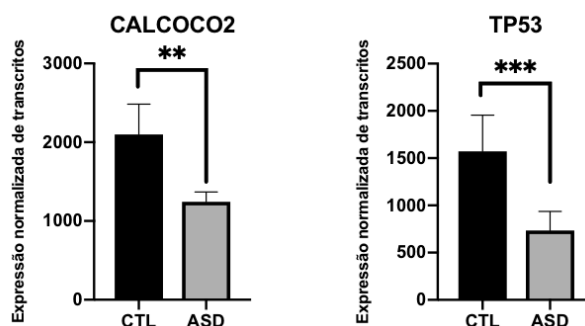


Figura 29: Reguladores L1 em ASD participantes da via de mitofagia.

Os gráficos mostram valores de expressão normalizada (RLE) de transcritos para amostras controle e amostras de ASD para os genes *CALCOCO2* e *TP53* participantes da via de mitofagia. Os gráficos de barras mostram a média ± SD (desvio padrão). *P-value*: *' $p < 0,05$; ***' $p < 0,01$; ****' $p < 0,001$.

Fonte: o autor.

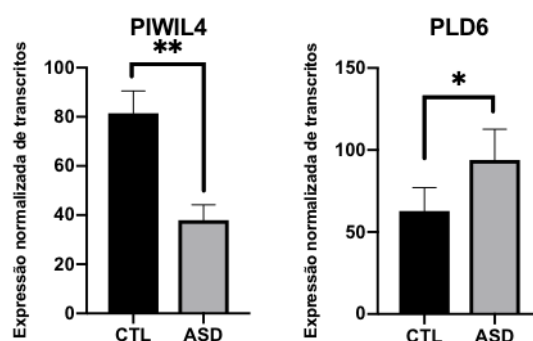


Figura 30: Reguladores L1 em ASD participantes da via da biogênese de piRNAs.

Os gráficos mostram valores de expressão normalizada (RLE) de transcritos para amostras controle e amostras de ASD para os genes *PIWIL4* e *PLD6* participantes da via de biogênese de piRNAs. Os gráficos de barras mostram a média ± SD (desvio padrão). *P-value*: *' $p < 0,05$; ***' $p < 0,01$; ****' $p < 0,001$.

Fonte: o autor.

5.3.4 Expressão de reguladores L1 em neurônios de DN

Em neurônios provenientes de cultura *in vitro* de amostras de ASD, RTT, BPD e SCZ não foram observados reguladores de L1 contendo um mesmo perfil de expressão entre todas as amostras (expressão aumentada ou reduzida em amostras para um determinado DN quando comparado com os respectivos controles) (Figura 31 e Apêndice C - Tabela 9).

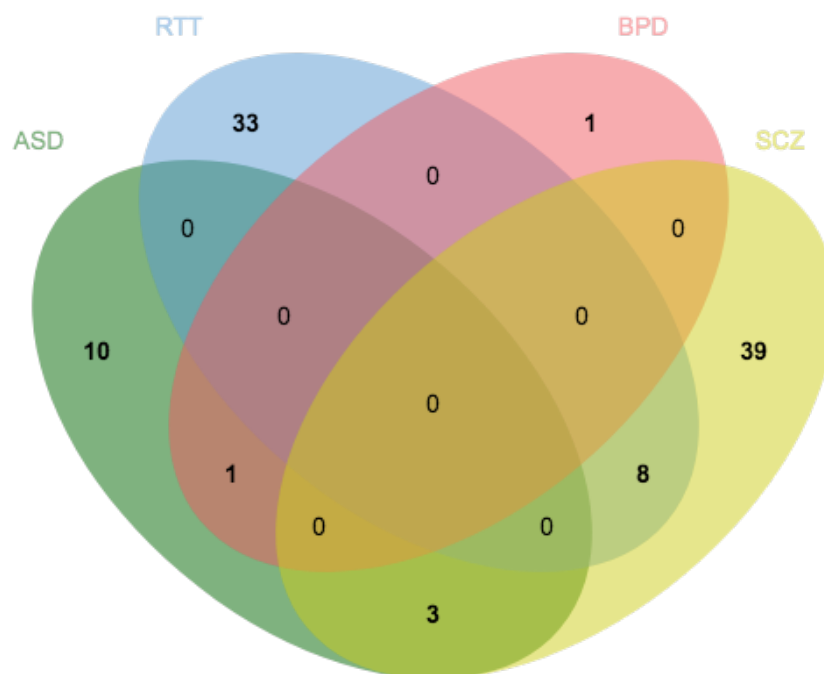


Figura 31: Padrão de expressão de moléculas reguladores de L1 em DN para neurônios provenientes de cultura *in vitro*.

O Diagrama de Venn representa o número de reguladores L1 com um mesmo perfil de expressão (aumentada/reduzida) em amostras de ASD, SCZ, BPD e RTT em neurônios provenientes de cultura *in vitro*. Variação de expressão (*fold-change*) entre amostras controle e afetadas considerando $p < 0,05$.

Fonte: o autor.

Entretanto, foram encontrados 8 reguladores de L1 com o mesmo perfil de expressão entre RTT e SCZ: *APOBEC3C* (*Apolipoprotein B mRNA Editing Enzyme Catalytic Subunit 3C*), *DCLRE1C*, *DGCR8* (*DGCR8 Microprocessor Complex Subunit*), *MAEL*, *RAD50*, *RNASEL*, *SAMHD1* e *UQCRH* (*Ubiquinol-Cytochrome C Reductase Hinge Protein*). Através da análise de enriquecimento de vias, os reguladores de L1 *DCLRE1C* e *RAD50*, com expressão significativamente aumentada nestas DN (Figura 32), participam da via união de extremidade não-homóloga (Figura 14). Um estudo observou, em linhagens celulares DT40 (*Gallus gallus*) com *knockout* para os genes pertencentes a esta via, *DCLRE1C*, *XRCC4* e *XRCC6*, uma redução na atividade de retroelementos L1 transfectados de humanos, associando esta via a regulação de L1 (SUZUKI et al., 2009). Outro estudo expôs que a proteína RAD50 atua regulando positivamente a atividade de L1 (LIU et al., 2018; MITA et al., 2020). O mecanismo de regulação de L1 por estas moléculas ainda não foi determinado.

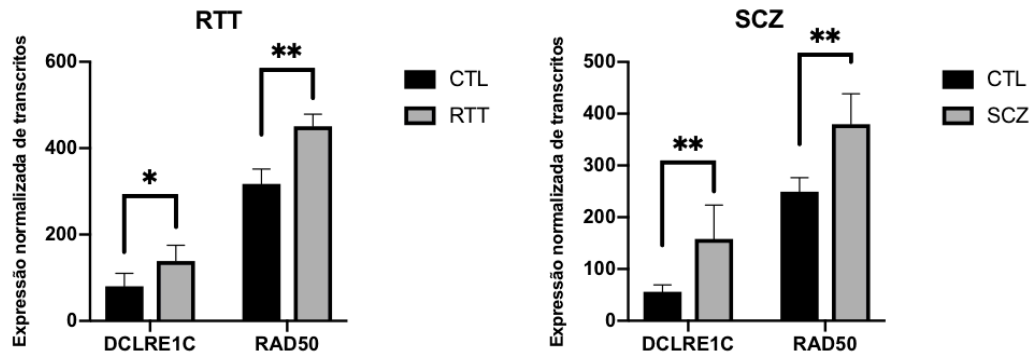


Figura 32: Reguladores L1 em RTT e SCZ participantes da via de reparo da união de extremidade não-homóloga.

Os gráficos mostram valores de expressão normalizada (RLE) de transcritos para amostras controle e amostras de RTT e SCZ para os genes *DCLRE1C* e *RAD50* participantes da via de reparo da união de extremidade não-homóloga. Os gráficos de barras mostram a média $\pm$ SD (desvio padrão). *P-value*: '*' $p < 0,05$; '**' $p < 0,01$; '***' $p < 0,001$.

Fonte: o autor.

Em RTT e SCZ ainda são observados reguladores de L1 com expressão significativamente aumentada nestas DN (Figura 33), e participando da via de sinalização de interferon alfa/beta (Figura 18). Um dos reguladores, a endoribonuclease RNASEL atua restringindo a atividade de L1 em células humanas quando cultivadas em laboratório (ZHANG et al., 2014). Outro regulador, SAMHD1 atua na inibição de retrovírus, vírus de DNA e também da atividade de L1 (HU et al., 2015).

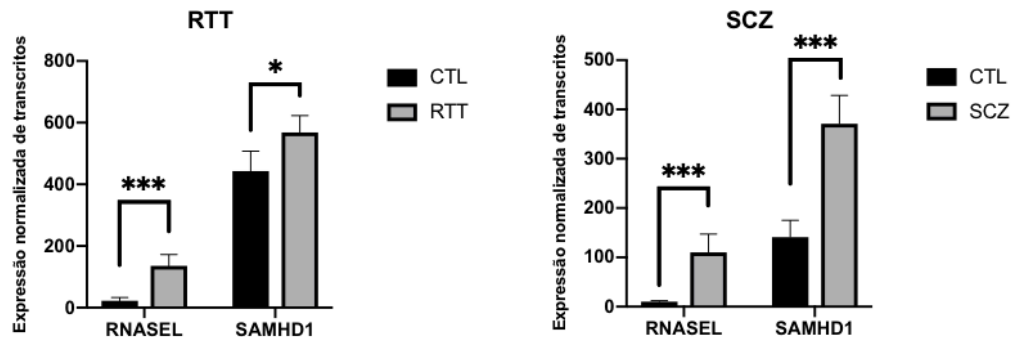


Figura 33: Reguladores L1 em RTT e SCZ participantes da via de sinalização de interferon alfa/beta.

Os gráficos mostram valores de expressão normalizada (RLE) de transcritos para amostras controle e amostras de RTT e SCZ para os genes *RNASEL* e *SAMHD1* participantes da via de sinalização de interferon alfa/beta. Os gráficos de barras mostram a média $\pm$ SD (desvio padrão). *P-value*: '*' $p < 0,05$; '**' $p < 0,01$; '***' $p < 0,001$.

Fonte: o autor.

Em ASD e SCZ foram observados os reguladores de L1, *E2F4* e *TP53* com expressão significativamente aumentada (Figura 34). Os fatores de transcrição pertencentes a família E2F, incluindo o fator de transcrição *E2F4*, são capazes de ligar-

se à região promotora de L1s em células humanas e de camundongos, reduzindo sua expressão (MONTROYA-DURANGO et al., 2009). O gene *TP53* é um fator de transcrição que possui sítios de ligação na região 5' UTR de L1 e está associado a restrição da atividade de L1 (WYLIE et al., 2016). Embora pesquisas não tenham associado a atividade destes genes na regulação de L1 em neurônios, novos estudos são necessários para avaliação da regulação de L1 neste tipo celular. Estes genes são reportados em nossa análise, participando da via de senescência celular (Figura 15).

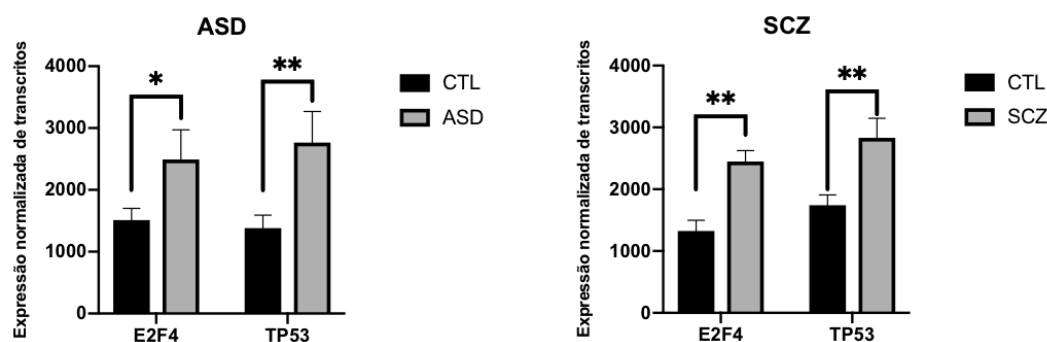


Figura 34: Reguladores L1 em ASD e SCZ participantes da via de senescência celular.

Os gráficos mostram valores de expressão normalizada (RLE) de transcritos para amostras controle e amostras de ASD e SCZ para os genes *E2F4* e *TP53* participantes da via de senescência celular. Os gráficos de barras mostram a média $\pm$ SD (desvio padrão). *P*-value: * $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$. Fonte: o autor.

Foram encontrados também, 39 reguladores de L1 com expressão alterada em SCZ (Figura 31). Após análise, identificamos a via de mitofagia (Figura 16) alterada, com os genes *CALCOCO2*, *BECN1* e *SQSTM1*, apresentando perfil de expressão (Figura 31) reduzido nesta DN (Figura 35). Estes genes foram reportados atuando na repressão da atividade e expressão de L1 em linhagens celulares de Hela e HEK293T (GUO et al., 2014).

Para reguladores L1 em RTT, 33 genes apresentam expressão alterada (Figura 31). Entre estes, os genes *BRCA1*, *FANCM* apresentam expressão reduzida em RTT, enquanto os genes *PALB2* e *SLX4*, expressão aumentada (Figura 36). Através de nossa análise, estes genes são reportados participando da via de anemia da Fanconi (Figura 11). Em relação a regulação de L1, todos estes genes atuam reprimindo a atividade de L1 (BRÉGNARD et al., 2016; LIU et al., 2018; MITA et al., 2020). A literatura reporta também que a proteína BRCA1 liga-se ao transcrito de L1 no citoplasma levando à

inibição da tradução de ORF2 e assim restringindo o processo de retrotransposição de L1 (MITA et al., 2020).

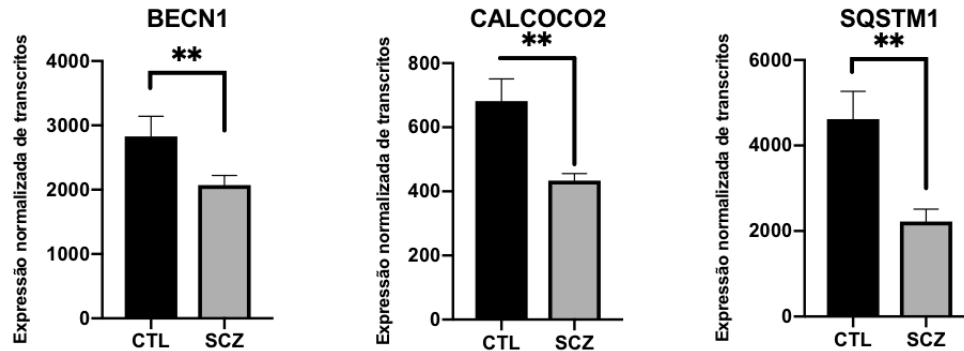


Figura 35: Reguladores L1 em SCZ participantes da via de mitofagia.

Os gráficos mostram valores de expressão normalizada (RLE) de transcritos para amostras controle e amostras de SCZ para os genes *BECN1*, *CALCOCO2* e *SQSTM1* participantes da via de mitofagia. Os gráficos de barras mostram a média $\pm$ SD (desvio padrão). *P*-value: '*' $p < 0,05$; '**' $p < 0,01$; '***' $p < 0,001$. Fonte: o autor.

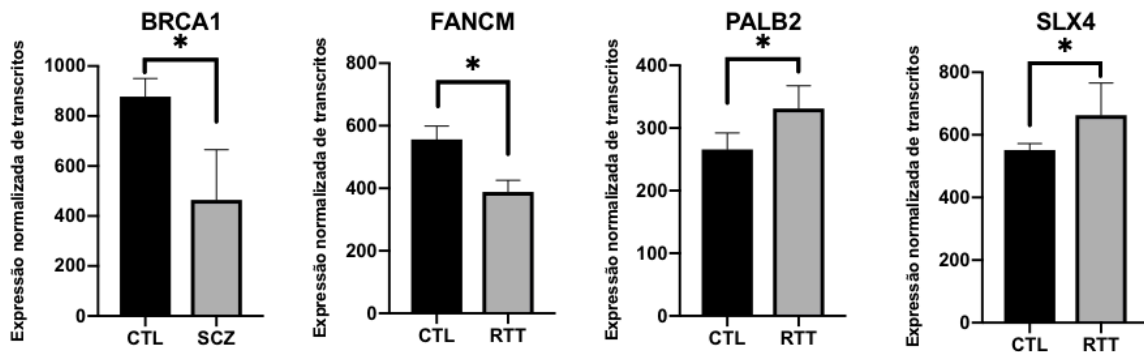


Figura 36: Reguladores L1 em RTT participantes da via da anemia da Fanconi.

Os gráficos mostram valores de expressão normalizada (RLE) de transcritos para amostras controle e amostras de RTT para os genes *BRCA1*, *FANCM*, *PALB2* e *SLX4* participantes da via da anemia de Fanconi. Os gráficos de barras mostram a média $\pm$ SD (desvio padrão). *P*-value: '*' $p < 0,05$; '**' $p < 0,01$; '***' $p < 0,001$.

Fonte: o autor.

Nesta mesma análise, foi identificado também um aumento de expressão em RTT dos genes reguladores de L1, *CALCOCO2* e *SP1* (Figura 37), ambos atuando na via de mitofagia (Figura 16). A proteína SP1 modula positivamente a expressão de elementos L1 (GOODIER, 2016), através da sua ligação a região promotora de L1 (FEDOROV; LUKYANOV; PODGORNAYA, 2006).

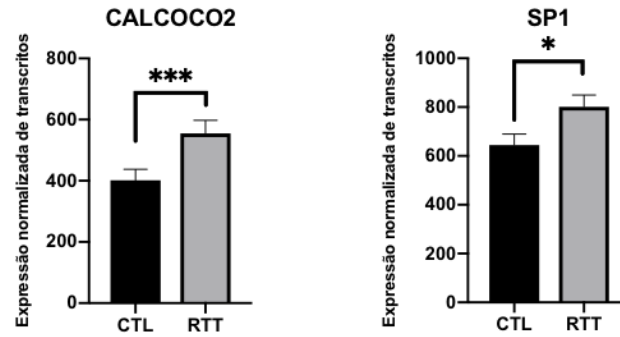


Figura 37: Reguladores L1 em RTT participantes da via de mitofagia.

Os gráficos mostram valores de expressão normalizada (RLE) de transcritos para amostras controle e amostras de RTT para os genes *CALCOCO2* e *SP1* participantes da via de mitofagia. Os gráficos de barras mostram a média $\pm$ SD (desvio padrão). *P*-value: '*' $p < 0,05$; '**' $p < 0,01$; '***' $p < 0,001$.

Fonte: o autor.

Além destes reguladores, foram observados os genes *MRE11A*, *XRCC6* e *PRKDC* com expressão significativamente aumentada em RTT (Figura 38), os quais participam da via de união de extremidade não-homóloga (Figura 14). As proteínas PRKDC e XRCC6 atuam inibindo a expressão/atividade de L1 (LIU et al., 2018; MITA et al., 2020; MORRISH et al., 2007; SUZUKI et al., 2009), enquanto que proteína MRE11A atua regulando positivamente a atividade de L1, favorecendo assim sua atividade (LIU et al., 2018; MITA et al., 2020).

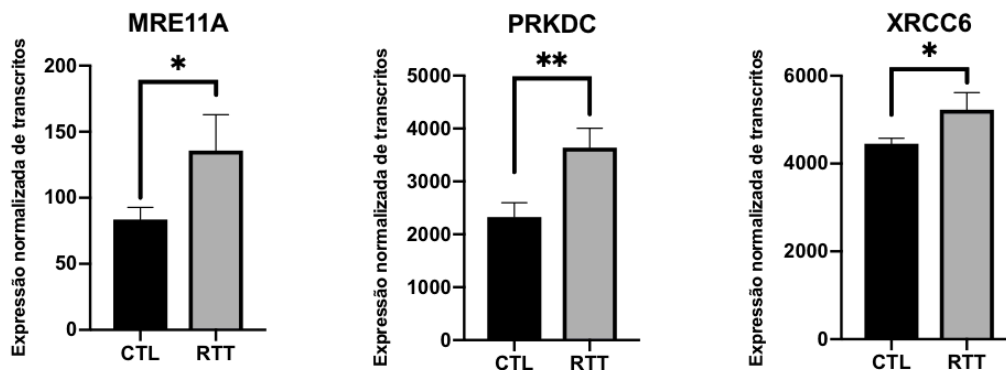


Figura 38: Reguladores L1 em RTT participantes da via da união de extremidade não-homóloga.

Os gráficos mostram valores de expressão normalizada (RLE) de transcritos para amostras controle e amostras de RTT para os genes *MRE11A*, *PRKDC* e *XRCC6* participantes da via da união de extremidade não-homóloga. Os gráficos de barras mostram a média $\pm$ SD (desvio padrão). *P*-value: '*' $p < 0,05$; '**' $p < 0,01$; '***' $p < 0,001$.

Fonte: o autor.

5.3.5 Associação entre genes reguladores de L1 e DN

Dentre os diversos genes reguladores de L1 explorados nos tópicos anteriores, vários compartilham processos biológicos específicos e potencialmente ligados ao desenvolvimento de DN.

Alterações em vias de reparo de DNA são relacionados a etiologia de DN. Em probandos de ASD, foi reportado que a perda de função do gene *XRCC6* devido a inserção de um *stop* códon prematuro no íntron 9 (NC_000022.10:g.42053038T>G), está associada ao desenvolvimento do fenótipo de ASD, provavelmente devido a desregulação deste gene e consequentemente da via da união de extremidade não homóloga (SJAARDA et al., 2020). Ainda, polimorfismos genéticos presentes em outros genes envolvidos em vias de reparo de DNA, como o regulador de L1 *XRCC4*, foram associados como fatores de risco ao desenvolvimento de SCZ (ODEMIS et al., 2015; SAADAT; PAKYARI; FARRASHBANDI, 2008). Outro estudo associou o polimorfismo de nucleotídeo único rs420259 do gene *PALB2*, participante da via de reparo DNA por recombinação homóloga, como um fator de risco para o desenvolvimento de BPD (TESLI et al., 2010). Por fim, foram relacionadas variantes genéticas em genes envolvidos na biogênese de piRNAs, incluindo o regulador de L1 *PIWIL2*, com o desenvolvimento de SCZ (ZHANG et al., 2012).

Para o gene *TP53*, participante de vias como a senescência celular, um estudo identificou que o polimorfismo rs1042522 está associado a suscetibilidade para o desenvolvimento de BPD (YANG et al., 2019a).

Nas análises realizadas, foram encontrados genes desregulados que estão associados a via de mitofagia, como *BECN1*, *CALCOCO2*, *SQSTM1* e *SP1*. Em diversos trabalhos, há a associação da disfunção da remoção autofágica seletiva de mitocôndrias danificadas (ou mitofagia) com o potencial desenvolvimento de ASD (GIULIVI et al., 2010), BPD (SCAINI et al., 2019), RTT (SBARDELLA et al., 2017) e SCZ (PRABAKARAN et al., 2004). Neste contexto, um gene participante desta via, o fator de transcrição *SP1*, que tem um papel central na transcrição de genes envolvidos no crescimento e diferenciação do sistema nervoso central, foi associado ao desenvolvimento de ASD (THANSEEM et al., 2012). Este estudo observou alta expressão da proteína SP1 em amostras *post mortem* de pacientes com ASD (THANSEEM et al., 2012). Os dados

gerados por este trabalho vem ao encontro dessa observação e reportam expressão significativamente elevada do gene *SP1* em neurônios de RTT ($p = 0.01$) (Apêndice C - Tabela 9).

Outro gene envolvido em processos autofágicos, *BECN1*, foi associado ao desenvolvimento de SCZ. Através de um estudo que demonstrou em tecidos *post-mortem* do hipocampo de indivíduos com SCZ, foi observada que a expressão do gene *BECN1* apresenta uma diminuição de 40% em comparação com respectivos controles (MERENLENDER-WAGNER et al., 2015). Os dados gerados por este trabalho expõem também uma redução significativa de expressão de *BECN1* em amostras de tecidos *post-mortem* de ASD ($p = 0,001$), BPD ($p < 0,001$) e SCZ ($p = 0,004$), em neurônios de SCZ ($p = 0,001$) e em NPC de ASD ($p = 0,038$) (Apêndice C - Tabela 7, 9 e 10).

Também, em outro gene envolvido em processos autofágicos, *CALCOCO2*, foi identificada que sua perda de função levou ao fenótipo de ASD (DE RUBEIS et al., 2014). Nos dados gerados por esta pesquisa foram encontrados níveis de expressão significativamente reduzidos de *CALCOCO2* em neurônios de SCZ ($p < 0,001$) e em NPC de ASD ($p = 0,002$) (Apêndice C - Tabela 9 e 10). Porém, para tecidos *post-mortem*, nossos dados apresentam expressão significativamente aumentada de *CALCOCO2* em ASD ($p < 0,001$), BPD ($p < 0,001$), RTT ($p = 0,043$) e SCZ ($p = 0,044$) e também em neurônios de RTT ($p < 0,001$) (Apêndice C - Tabela 7 e 9).

Por fim, para o gene participante de processos autofágicos, *SQSTM1*, outro estudo analisou amostras de sangue total de pacientes com BPD, e encontrou níveis de expressão significativamente reduzidos para o gene *SQSTM1* e para a proteína codificada por este gene, comparados a amostras controle (SCAINI et al., 2019). Para a análise realizada, foi observado expressão significativamente reduzida de *SQSTM1* em amostras de NPC de ASD ($p = 0,006$) e SCZ ($p = 0,031$), expressão reduzida em neurônios de SCZ ($p = 0,009$), e expressão aumentada em células provenientes de tecidos *post mortem* de ASD ($p = 0,019$) (Apêndice C - Tabela 7, 9 e 10).

Os dados da literatura, bem como os dados gerados por esta pesquisa, sugerem que diversos genes participantes de processos biológicos específicos e que atuam na regulação de retroelementos L1 também estão associados, ou são potenciais candidatos ao desenvolvimento de DN. Neste contexto, no próximo item, será analisado um processo

celular específico, o processo da autofagia. Através de análises, foi identificado um gene pertencente a esta via que está associado a fenótipos encontrados em DN, e apresentamos este, como um potencial regulador de retroelementos L1.

5.4 Regulação de retroelementos L1 e via de autofagia

A análise dos dados permitiu observar em células cerebrais alta expressão de transcritos L1, bem como, reguladores da atividade de L1 e processos biológicos alterados em amostras de DN. Dentre os reguladores da atividade de L1 alterados, os genes *BECN1*, *CALCOCO2* e *SQSTM1* atuam diretamente em processos autofágicos.

O processo autofágico (ou macroautofagia), envolve a formação de uma vesícula de membrana dupla chamada autofagossomo. A iniciação deste processo inclui a molécula mTORC1 (*mammalian Target of Rapamycin Complex 1*), a qual, em condições ricas em nutrientes, atua inibindo o catabolismo celular através da repressão da maquinaria de autofagia, que compreende quatro complexos macromoleculares principais: complexo ULK (*mammalian Unc-51-Like Kinase*) - que contém a molécula ULK1 (*Unc-51 Like Autophagy Activating Kinase 1*); complexo PI3K (*Phosphoinositide 3-Kinase - class III*) - que inclui moléculas como *BECN1* (*Beclin 1*) e *PIK3C3* (*Phosphatidylinositol 3-Kinase Catalytic Subunit Type 3*); sistema de conjugação “ubiquitin-like” – composta de moléculas como *ATG12* (*Autophagy Related 12*) e *LC3* (*Microtubule-associated protein 1A/1B-light chain 3*); e proteínas de transmembrana, como *ATG9A* (*Autophagy Related 9A*) (ABADA; ELAZAR, 2014; RABANAL-RUIZ; OTTEN; KOROLCHUK, 2017) (Figura 39).

Buscando ampliar a compreensão sobre a remoção de L1 em células cerebrais, foram investigados genes que interagem com algum daqueles já relacionadas na remoção de L1s na via autofágica (*BECN1*, *CALCOCO2* e *SQSTM1*), e que tenham alguma relação com processos envolvidos no neurodesenvolvimento e o DN.

A partir da via de autofagia, foram analisadas a expressão de todos os genes desta via com relação aos reguladores de L1 do trio discutido anteriormente. Foram considerados para esta análise os dados de expressão de transcritos para células cerebrais gerados/obtidos por esta pesquisa (N = 729), separando os conjuntos de

amostras controle (N = 402) e de DN (N = 327). Também foram obtidos dados do projeto GTEx (*Genotype-Tissue Expression*) (LONSDALE et al., 2013), referentes a expressão de transcritos oriundos de amostras de indivíduos saudáveis, e de diferentes tipos celulares como os provenientes de cérebro, tecido adiposo, glândula adrenal, bexiga, sangue, medula óssea, cólon, esôfago, coração, rim, fígado, pulmão, músculo, ovário, pâncreas, pele, baço, estômago, útero e testículo (N = 17382). Para estes dados, foi realizada uma filtragem separando um conjunto gerado a partir de células cerebrais (N = 2642) e outro conjunto com os demais tipos celulares (N = 14740).

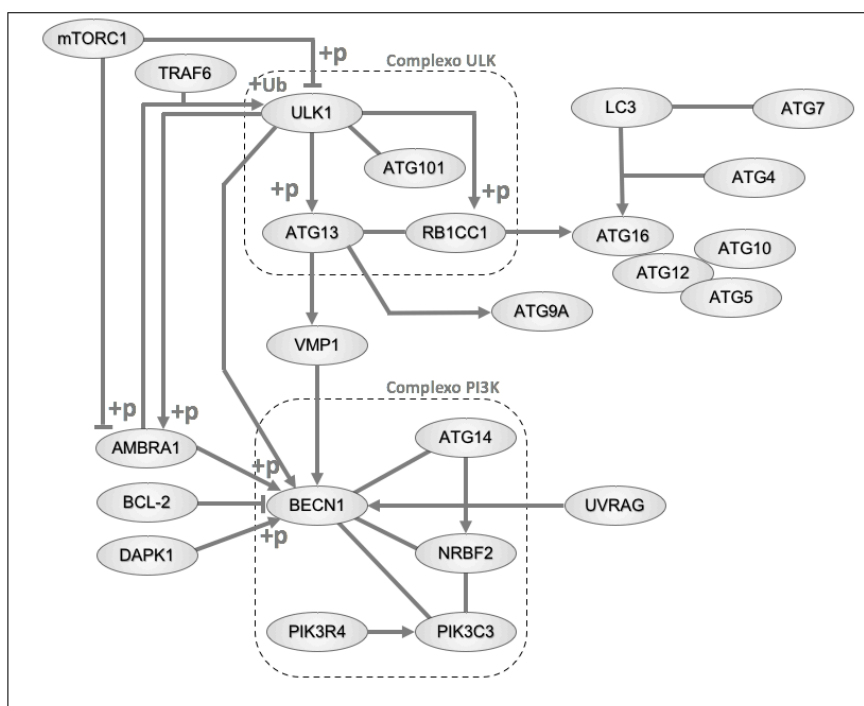


Figura 39: Via de autofagia em células humanas.

O processo canônico da autofagia compreende quatro complexos macromoleculares principais: complexo ULK no qual participam moléculas como a ULK1 (*Unc-51 Like Autophagy Activating Kinase 1*), ATG101 (*Autophagy Related 101*), RB1CC1 (*RB1 Inducible Coiled-Coil 1*) e ATG13 (*Autophagy Related 13*); complexo PI3K, que inclui moléculas como BECN1 (Beclin 1) e PIK3C3 (Phosphatidylinositol 3-Kinase Catalytic Subunit Type 3); sistema de conjugação “ubiquitin-like”, composta de moléculas como ATG12 (*Autophagy Related 12*) e LC3 (*Microtubule-associated protein 1A/1B-light chain 3*); e proteínas transmembrana, como ATG9A (*Autophagy Related 9A*). Após a inibição da proteína mTORC1 (*mammalian Target of Rapamycin Complex 1*), ocorre a iniciação do fluxo autofágico através da regulação dos complexos ULK1 e PI3K.

Fonte: adaptado de (KANEHISA; GOTO, 2000).

Através da análise do coeficiente de correlação de Spearman, foi detectada alta correlação na expressão dos genes *AMBRA1* (*Autophagy And Beclin 1 Regulator 1*) e *BECN1* em células cerebrais saudáveis ($r = 0,80$ e $r = 0,73$) (Figura 40A-40C). Quando

observa-se a correlação entre estes mesmos genes em amostras de DN, é exposta uma menor correlação ($r = 0,63$) (Figura 40D). Já a correlação entre *AMBRA1* e *BECN1*, considerando os demais tipos de tecidos, observa-se o menor valor de correlação quando comparadas com as demais análises ($r = 0,43$) (Figura 40B).

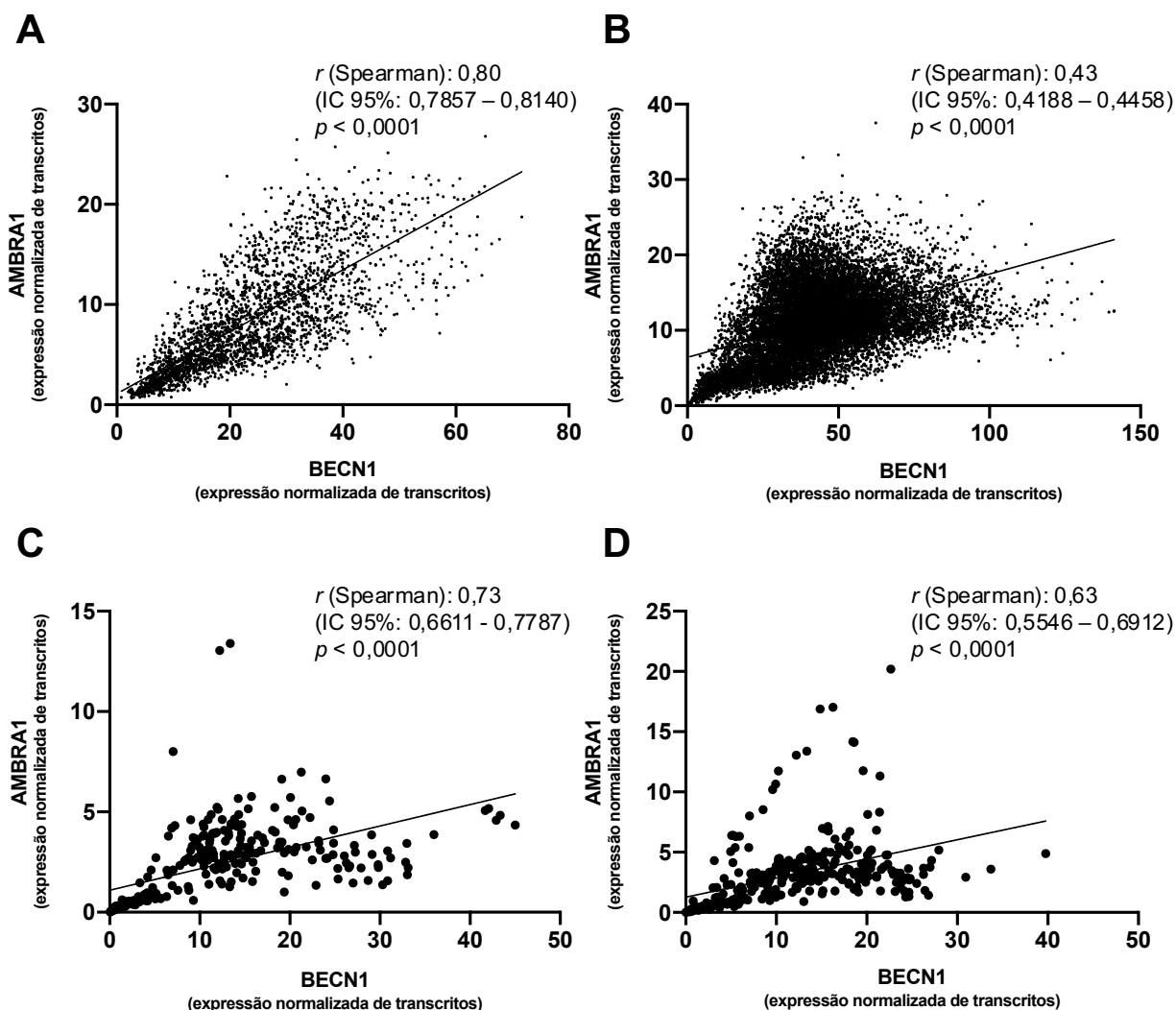


Figura 40: Correlação de expressão dos genes *AMBRA1* e *BECN1* em células cerebrais e não cerebrais.

As análises de correlação para os níveis de expressão de transcritos, consideram dados normalizados (TPM - *Transcripts Per Million*). IC: intervalo de confiança; r : coeficiente de correlação de Spearman. **A.** Análise do coeficiente de correlação de Spearman entre os genes *AMBRA1* e *BECN1* para células cerebrais provenientes do projeto GTex. **B.** Análise do coeficiente de correlação de Spearman entre os genes *AMBRA1* e *BECN1* para células de tecido adiposo, glândula adrenal, bexiga, sangue, vaso sanguíneo, medula óssea, cólon, esôfago, coração, rim, fígado, pulmão, músculo, ovário, pâncreas, pele, baço, estômago, útero e testículo provenientes do projeto GTex. **C.** Análise do coeficiente de correlação de Spearman entre os genes *AMBRA1* e *BECN1* para células cerebrais saudáveis (grupo controle) oriundas de dados gerados/obtidos por esta pesquisa. **D.** Análise do coeficiente de correlação de Spearman entre os genes *AMBRA1* e *BECN1* para células cerebrais de pacientes com DN provenientes de dados gerados/obtidos por esta pesquisa.

Fonte: o autor.

A partir desta observação, foi realizada uma análise da literatura para verificar em detalhes a relação de AMBRA1 com o processo de autofagia, e principalmente se há dados correlacionando este gene com a atividade de L1. A proteína AMBRA1, antes da indução da autofagia, encontra-se ligada a um complexo de dineína. Após a indução da autofagia, a fosforilação mediada por ULK1 libera AMBRA1 deste complexo, o que resulta na sua translocação juntamente com a proteína BECN1 e PIK3C3 para o retículo endoplasmático (do inglês, *Endoplasmic Reticulum* - ER), onde participam da biogênese de autofagossomos (FIMIA et al., 2007). Além disso, em estágios iniciais após a indução da autofagia, AMBRA1 promove a ubiquitinação da proteína ULK1 através de cadeias de ubiquitina ligadas à lisina 63. Esta modificação é essencial para a auto-associação da proteína ULK1, sendo mediada pela proteína TRAF6 (*TNF Receptor Associated Factor 6*), que interage com o complexo AMBRA1-ULK1. Esse mecanismo representa um ciclo no qual AMBRA1 participa no ajuste da resposta autofágica, aprimorando a capacidade de sinalização da proteína ULK1 (Figura 39) (NAZIO et al., 2013).

Também, em um estudo que induziu a superexpressão dos gene *AMBRA1* em células humanas, foi observada a ativação do processo autofágico (FIMIA et al., 2007). Além disso, a diminuição da expressão de *AMBRA1* resulta na redução da resposta autofágica (FIMIA et al., 2007). Outro estudo observou que em fibroblastos humanos, a redução dos níveis da proteína AMBRA1 reduzem também os níveis de BECN1, comprometendo o estágio inicial da formação dos autofagossomos (DI BARTOLOMEO et al., 2010).

Além da atuação de AMBRA1 na via de autofagia, foi levantada a relação entre AMBRA1, neurodesenvolvimento e DN. AMBRA1 desempenha um papel fundamental durante a embriogênese de vertebrados, sendo altamente expresso no sistema nervoso central (SNC) (FIMIA et al., 2007; VÁZQUEZ et al., 2012). A deficiência da proteína Ambra1 em embriões de camundongos causa defeitos graves do tubo neural que estão associados ao comprometimento da autofagia, proliferação celular excessiva nos estágios iniciais da formação do tubo neural. Isso é seguido por um aumento da apoptose e acúmulo de proteínas ubiquitinadas no neuroepitélio, levando a letalidade embrionária (FIMIA et al., 2007). Com base nesta evidência, Ambra1 é uma proteína essencial para

controlar a proliferação celular e a sobrevivência celular durante o desenvolvimento do SNC. Na vida adulta, foi observado em camundongos que a proteína Ambra1 juntamente com Becn1 é vital para a manutenção das células-tronco neuronais dentro da zona subventricular do cérebro (SVZ), controlando o nível de neurônios imaturos e promovendo a sobrevivência de células precursoras neurais (YAZDANKHAH et al., 2014).

Várias condições patológicas, principalmente relacionadas ao sistema nervoso, tem sido associadas a molécula AMBRA1. Rietschel e colegas observaram uma forte associação entre SCZ e uma variante genética (rs11819869) de AMBRA1 (RIETSCHEL et al., 2012). Outro estudo realizou uma análise comportamental abrangente de camundongos heterozigotos para o gene *Ambra1* (*Ambra1^{+/-}*) e revelou um fenótipo semelhante aos encontrados em ASD, incluindo interações sociais comprometidas, uma tendência a exibir estereotipias ou comportamentos repetitivos e flexibilidade cognitiva prejudicada (DERE et al., 2014). Um estudo de associação genética baseado em fenótipo, revelou uma correlação entre a variante genética de AMBRA1 (rs3802890) com características autistas em camundongos. (MITJANS et al., 2017).

Além das evidências sobre a atuação de AMBRA1 na via autofágica, neurodesenvolvimento e relacionamento com DN, buscou-se dados que aprofundem a relação entre AMBRA1 e BECN1. A nível de transcritos, não é reportado na literatura padrões de co-expressão para os genes *AMBRA1* e *BECN1*.

Com base nos dados gerados por esta pesquisa e dados da literatura, foi possível confirmar a forte relação entre AMBRA1 e processos do neurodesenvolvimento, a alta correlação de expressão de transcritos em células cerebrais entre *AMBRA1* e *BECN1*, e a interação fundamental entre estas proteínas para a ocorrência adequada do processo autofágico. Com base nos dados da literatura, o fato de que BECN1 é um regulador comprovado da atividade de L1, e que há alta correlação entre *BECN1* e *AMBRA1*, hipotetiza-se fortemente, que os níveis de regulação de *AMBRA1* na via autofágica vem a interferir diretamente na atividade de L1s em células cerebrais (Figura 41). O processo autofágico, vital para a função celular adequada, é extremamente essencial na manutenção do equilíbrio celular em situações de instabilidade genética. Uma das maiores fontes de instabilidade genética é causada pela atividade de retroelementos L1,

o que pode ocasionar eventos de translocações, deleções, inversões e amplificações no DNA nuclear (ELBARBARY; LUCAS; MAQUAT, 2016). Além disso, esta atividade é considerada como uma das maiores fontes geradoras de DNA citosólico (STETSON et al., 2008). Através da análise realizada nesta pesquisa, observou-se alta expressão de transcritos L1 em células cerebrais humanas. Este acúmulo de DNA pode provocar toxicidade celular, e para a manutenção de níveis de homeostase celular adequados, a via de autofagia deve realizar remoção eficiente de seu acúmulo.

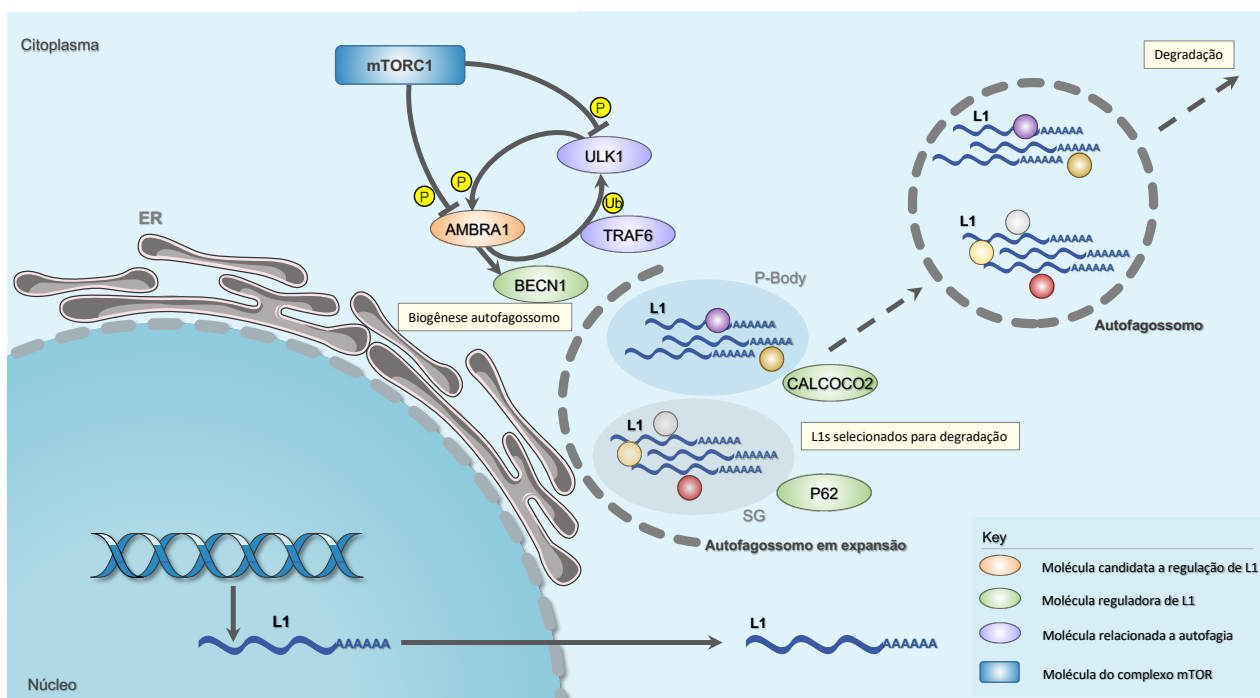


Figura 41: Regulação de L1 via processo autofágico.

A fosforilação promovida por mTORC1 no complexo ULK1-AMBRA1, resulta na inibição da autofagia. Após a indução da autofagia, a fosforilação mediada por ULK1 promove a translocação de AMBRA1 juntamente com a molécula BECN1 para o retículo endoplasmático (ER) onde participam da biogênese de autofagossomos. AMBRA1 promove também a ubiquitinação de ULK1, essencial para a auto-associação da ULK1 mediada pela molécula TRAF6. A regulação de L1 na via de autofagia é realizada por meio das moléculas BECN1 e dos receptores de autofagia CALCOCO2 e P62. Estes receptores direcionam os RNAs provenientes de L1s localizados em corpos de processamento (P-Body) e grânulos de estresse (SG) para degradação autofágica.

Fonte: o autor.

Assim, hipotetiza-se também que, após indução autofágica, células cerebrais humanas com expressão reduzida de *AMBRA1* podem apresentar maior expressão de elementos L1, bem como maior número de inserções (atividade) destes no genoma nuclear. Também, em células *knockout* para *AMBRA1*, as quais podem apresentar alta

concentração de transcritos e proteínas oriundas de L1 no citoplasma, se induzirmos a superexpressão de *AMBRA1* nestas células, é esperado que haja indução autofágica, a interação entre as moléculas *AMBRA1* e *BECN1* seja restabelecida, o fluxo autofágico volte a ocorrer de forma equilibrada e assim, haja a remoção de L1s no citoplasma, reduzindo sua atividade.

6 CONCLUSÕES

Este trabalho explorou a relação entre a regulação de retroelementos L1 e os distúrbios do neurodesenvolvimento, ASD, BPD, RTT e SCZ. Existe uma forte ligação entre o desenvolvimento destes distúrbios e a regulação de L1, dada a observação de diversas moléculas que regulam L1 relacionadas ao desenvolvimento destes distúrbios. Assim, este trabalho buscou contribuir no entendimento de como o processo de expressão e atividade de elementos L1 ocorre a partir da análise de dados de sequenciamento em larga escala de RNA oriundos de células cerebrais de pacientes portadores de DN.

Neste sentido, o primeiro passo foi dado através da criação de uma base de anotações de retroelementos L1, que oferece um catálogo amplo de anotações destes, tanto com sequências completas do elemento L1, quanto com suas regiões, 5'UTR, ORF1, íntron, ORF2 e 3'UTR. O catálogo foi criado para o genoma de referência humano (base de anotações L1RG-REF e L1CP-REF) e para os genomas individuais de um Caucasiano e um Asiático (base de anotações L1RG-IND e L1CP-IND). Tais bases de anotações são um recurso importantíssimo para o estudo individualizado de regiões L1 para experimentos baseados em humanos. Estas bases diferem de outras anotações L1 por exporem elementos que contêm todas as regiões L1 e, assim, oferecem um panorama da localização exata de cada elemento no genoma, além de permitirem inferir, a partir de novas análises *in silico*, quais *loci* apresentam assinatura de elementos L1 potencialmente ativos. Além disso, elementos conflitantes foram eliminados por meio do uso de sequências de consenso e anotações de sequência baseadas em similaridade. Também, foi ampliada a compreensão da distribuição de L1 em genomas individuais, incluindo anotações de regiões L1 e elementos L1 de comprimento total para os genomas individuais Asiático e Caucasiano, permitindo assim, a análise de L1 a nível alélico. Por fim, para a criação desta base de anotações, foi desenvolvido um *software* para a identificação automática de elementos L1 no genoma humano, e que pode ser estendido para genomas de diferentes organismos.

Através desta base, foi possível avaliar o perfil de expressão de L1s em diferentes tipos celulares e DN. Foi avaliada a expressão destes retroelementos considerando todas as subfamílias L1 e também de forma individualizada para a subfamília L1HS, que concentra elementos L1 capazes de realizar novas inserções no genoma humano. Os dados gerados por esta pesquisa demonstraram alta expressão de transcritos L1 em todos os tipos celulares cerebrais e DN. Em nível de transcritos isso expõe um ambiente onde, através da desregulação oportuna de moduladores da atividade de L1s, podem ocorrer novas inserções destes no genoma nuclear, bem como a mobilização de outros retroelementos, o que pode levar ao aumento da instabilidade genômica.

Além destes dados de expressão de L1, foram avaliadas também a expressão de genes reguladores de L1 para diferentes tipos celulares e DN. Para isso, foi realizada uma revisão sistemática da literatura e criado um atlas único de 158 genes que modulam a expressão e atividade de L1s. Como resultado desta revisão, pode-se contextualizar como estes reguladores atuam a nível pré-transcricional e pós-transcricional na regulação de L1s, fornecendo um panorama geral dos processos biológicos envolvidos. A análise realizada mostrou um contexto que engloba inúmeras moléculas reguladoras com perfis de expressão significativamente alterados em DN, o que, juntamente com a alta expressão de L1s observada em células cerebrais, gera um ambiente favorável para o processo de retrotransposição ocorrer. Além disso, para cada tipo celular e DN, também foram listados diversos processos biológicos nos quais estes genes reguladores de L1 atuam. Processos de reparo de DNA genômico como a via de anemia da Fanconi, via de reparo de DNA por excisão de nucleotídeos, via de união de extremidade não-homóloga e a via de reparo de DNA por recombinação homóloga, apresentam genes com expressão significativamente alterada nas amostras analisadas. Foram observados também outros processos biológicos alterados nos dados analisados, como a sinalização de interferon alfa/beta, senescência celular, biogênese de piRNAs e mitofagia.

Observando as diversas moléculas com expressão significativamente alterada participantes da via de mitofagia, foi explorado também seu contexto mais amplo, o processo biológico da autofagia na regulação de L1s. Através da análise realizada, foi

identificado o gene *AMBRA1* como uma forte candidato a regulador de L1s em células cerebrais humanas. Este gene está associado ao desenvolvimento do sistema nervoso central, neurogênese e também está relacionado com condições patológicas cerebrais. Foram analisados mais de dezoito mil amostras e observamos que em cérebro, o gene *AMBRA1* tem um alto grau de correlação com *BECN1*, este último, já relacionado como regulador de L1s. Também, a análise realizada sugere que a remoção de L1s realizada por *BECN1*, ocorre em um nível mais amplo, através da interação com *AMBRA1*, o que nos direciona a hipótese de *AMBRA1* como uma nova molécula reguladora de L1.

Por fim, este trabalho, além dos resultados reportados, gerou um repertório de informações que poderão ser utilizadas como ferramentas em outras pesquisas de forma a expandir o conhecimento sobre a regulação de L1s e também ao desenvolvimento de DN.

7 PERSPECTIVAS

A presente pesquisa apresenta diversas perspectivas que se baseiam na continuidade dos trabalhos para complementar as contribuições realizadas. Desta forma, as principais perspectivas desta pesquisa são listadas a seguir:

- Inserção de L1 associados com genomas individuais: A partir da criação das bases de anotações de L1 em genomas individuais, é possível verificar se há inserções específicas que associem diferenças genômicas entre indivíduos relacionados unicamente com a atividade de L1.
- Gene *AMBRA1* como um novo regulador da atividade de L1: nas análises realizadas, através da avaliação da via de autofagia, foi exposto o gene *AMBRA1*, que desempenha papel fundamental nesta via e também surge como um gene altamente provável em realizar a regulação de L1. Para complementar estes resultados, é interessante a realização de ensaios *in vitro* para explorar o ganho e perda de função de *AMBRA1* em células humanas. Nestes ensaios, são esperados que, nas células avaliadas, ocorra ativação da via de autofagia, através da ausência de nutrientes celulares ou através da indução por meio de drogas como rapamicina, as células onde *AMBRA1* encontra-se com expressão atenuada, sofra um aumento de expressão e atividade de L1. De forma complementar, o aumento de expressão de *AMBRA1* nestas células, deve induzir o processo autofágico *per se* e reduzir a expressão e atividade de L1. Ainda, estes ensaios devem analisar o comportamento de proteínas já relacionadas com regulação de L1, como BECN1, CALCOCO2 e SQSTM1, e assim relacionar o impacto de *AMBRA1* na regulação de L1s, dentro de um contexto mais amplo nesta via.
- Mapeamento *in silico* de retroelementos L1 potencialmente ativos: Esta pesquisa mapeou os *loci* de inserções L1 em todas as amostras analisadas. Através destas informações, podem ser geradas análises *in silico* que identifiquem *loci* com alta expressão de transcritos L1. Com isso, pode-se mapear retroelementos L1 potencialmente ativos dentro de diferentes tecidos e

também de DN, fornecendo meios para o desenvolvimento de novas técnicas que visem a identificação da atividade de L1.

- Criação de um mapa de regulação de L1: Também, com os dados de expressão gerados por esta pesquisa, será possível, além de identificar novos reguladores de L1, criar um mapa de regulação de L1 a nível de múltiplos processos biológicos, como através da interação de vias do sistema imune com sistema de reparo de DNA e processo autofágico. Isto vem na direção da identificação de uma possível assinatura, a nível de múltiplas vias, que revele a desregulação de retroelementos L1 e que possa levar a um potencial desenvolvimento de DN.

8 ATIVIDADES COMPLEMENTARES

Como parte complementar desta pesquisa, foram propostas atividades que envolveram a participação em eventos, publicação de artigos e a realização de colaborações de pesquisa científica com pesquisadores de outras instituições nacionais e internacionais.

8.1 Artigos Publicados

1. ARAUJO, B. H. S. ; KAID, C.; DE SOUZA, J.; GOMES DA SILVA, S.; GOULART, E.; CAIRES, L. C. J.; MUSSO, C. M.; TORRES, L. B.; **FERRASA, A.**; **HERAI, R.**; ZATZ, M.; OKAMOTO, O. K.; CAVALHEIRO, E. A. *Down Syndrome iPSC-Derived Astrocytes Impair Neuronal Synaptogenesis and the mTOR Pathway In Vitro*. Molecular Neurobiology, nov. 2017, DOI: 10.1007/s12035-017-0818-6. (FI: 4.586, Scimago Rank: Q1)
2. FERNANDES, I. R.; CRUZ, A. C. P.; **FERRASA A.**; PHAN, D.; **HERAI, R. H.**; MUOTRI, A. R. *Genetic variations on SETD5 underlying autistic conditions*. Developmental Neurobiology, v. 78, n. 5, p. 500–518, maio 2018, DOI: 10.1002/dneu.22584. (FI: 2.6, Scimago Rank: Q1)
3. CRUZ, A. C. P.; **FERRASA A.**; MUOTRI, A. R. **HERAI, R. H.** *Frequency and association of mitochondrial genetic variants with neurological disorders*. Mitochondrion, 13 set. 2018, DOI: 10.1016/j.mito.2018.09.005. (FI: 3.449, Scimago Rank: Q1)
4. CAVALLARI, T., ARIMA L.Y., **FERRASA A.**, MOYSÉS S.J., MOYSÉS S.T., **HERAI R.H.**, WERNECK R.I.. 2019. *Dental caries: Genetic and protein interactions*. Archives of Oral Biology, 15 ago. 2019, DOI: 10.1016/J.ARCHORALBIO.2019.104522. (FI: 1.663, Scimago Rank: Q1)
5. **ADRIANO FERRASA**; MAYARA M. MURATA; TERESA D. C. G. COFRE; JULIANA S. CAVALLINI; GUSTAVO PERON; JOSÉ BELASQUE JR; HENRIQUE FERREIRA;

MARIA INÊS T. FERRO; RUI P. LEITE JR; HELEN A. PENHA; ALESSANDRO M. VARANI; **ROBERTO H. HERAI**; JESUS A. FERRO. *CitrusKB: A Comprehensive Knowledge Base for Transcriptome and Interactome of Citrus spp. Infected by Xanthomonas citri subsp. citri at Different Infection Stages*. Database: The Journal of Biological Databases and Curation, 2020. (FI: 3.660, Scimago Rank: Q1)

8.2 Artigos submetidos ou em avaliação

1. **ADRIANO FERRASA; ROBERTO H. HERAI**. *L1Farm: an accurate comprehensive annotation of mobile LINE-1 elements in human reference and personal genomes*. FEBS Letters. (FI: 3.370, Scimago Rank: Q1)
2. PINAR MESCI; CHRISTOPHER N. LAROCK; **ADRIANO FERRASA; ROBERTO H. HERAI**; JACOB J. JEZIORSKI; AURIAN SALEH; CLEBER A. TRUJILLO; VICTOR NIZET; ALYSSON R. MUOTRI. *Impact of human microglial cells on neuronal connectivity during development*. Cell. (FI: 38.637, Scimago Rank: Q1)
3. **ADRIANO FERRASA; ROBERTO H. HERAI**. *Regulators of LINE-1 expression and retrotransposition activity in neurological tissues*.
4. **ADRIANO FERRASA; ROBERTO H. HERAI**. *A comprehensive atlas of molecular regulators of expression and activity of LINE-1 retrotransposons and genetic interactions*.
5. ÉRICO M. LINS; NATÁSSIA C. M. OLIVEIRA; OSVALDO REIS; **ADRIANO FERRASA; ROBERTO H. HERAI**; ALYSSON MUOTRI; KATLIN B. MASSIRER; MÁRIO H. BENGTSON. *Genome-wide translation control analysis of developing human neurons*.

6. **ADRIANO FERRASA**; HELEN A. PENHA; JESUS A. FERRO; **ROBERTO H. HERAI**. *SSRHTS: a genomic and genic SSR retrieval and validator for high-throughput sequencing data.*
7. Isabella R Fernandes, Anju Singh, Raul Calvo, Pinar Mesci, Inigo Narvaiza, Osamu Sano, Jordi Serrats, Dylan Phan, **ADRIANO FERRASA**, **ROBERTO H. HERAI**, Priscilla D. Negraes, Cleber Trujillo, Dingyin Tao, Drew Burns, Christopher LeChair, Noel Southall, Marc Ferrer, Juan Marugan, Alysson R Muotri. *Disruption of SETD5 in astrocytes leads to neuroinflammation.*

REFERÊNCIAS

- ABADA, A.; ELAZAR, Z. Getting ready for building: signaling and autophagosome biogenesis. **EMBO reports**, v. 15, n. 8, p. 839–852, 2014.
- AHN, S. M. et al. The first Korean genome sequence and analysis: Full genome sequencing for a socio-ethnic group. **Genome Research**, v. 19, n. 9, p. 1622–1629, 2009.
- AMERICAN PSYCHIATRIC ASSOCIATION. **DSM-5 Development**. 5th. ed. Arlington: Amer Psychiatric Pre, 2013.
- AMIR, R. E. et al. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl- CpG-binding protein 2. **Nature Genetics**, v. 23, p. 185–188, 1999.
- ANCELIN, K. et al. Maternal LSD1/KDM1A is an essential regulator of chromatin and transcription landscapes during zygotic genome activation. **eLife**, v. 5, p. e08851, 2016.
- AOTO, S. et al. Frequent retrotransposition of endogenous genes in ERCC2-deficient cells derived from a patient with xeroderma pigmentosum. **Stem Cell Research and Therapy**, v. 10, n. 1, p. 273, 2019.
- ARAVIN, A. A. et al. Developmentally regulated piRNA clusters implicate MILI in transposon control. **Science**, v. 316, n. 5825, p. 744–747, 2007.
- ARIUMI, Y. et al. DNA repair protein Rad18 restricts LINE-1 mobility. **Scientific Reports**, v. 8, n. 1, p. 15894, 2018.
- ARJAN-OEDRA, S. et al. Endogenous MOV10 inhibits the retrotransposition of endogenous retroelements but not the replication of exogenous retroviruses. **Retrovirology**, v. 9, n. 1, p. 53, 2012.
- ATHANIKAR, J. N.; BADGE, R. M.; MORAN, J. V. A YY1-binding site is required for accurate human LINE-1 transcription initiation. **Nucleic Acids Research**, v. 32, n. 13, p. 3846–3855, 2004.
- ATTIG, J.; ULE, J. Genomic Accumulation of Retrotransposons Was Facilitated by Repressive RNA-Binding Proteins: A Hypothesis. **BioEssays**, v. 41, n. 2, p. 1800132, 2019.
- BABUSHOK, D. V.; KAZAZIAN, H. H. Progress in understanding the biology of the human mutagen LINE-1. **Human Mutation**, v. 28, n. 6, p. 527–39, 2007.
- BAILLIE, J. K. et al. Somatic retrotransposition alters the genetic landscape of the human brain. **Nature**, v. 479, n. 7374, p. 534–7, 2011.
- BAO, W.; KOJIMA, K. K.; KOHANY, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. **Mobile DNA**, v. 6, n. 1, p. 11, 2015.
- BARAU, J. et al. The DNA methyltransferase DNMT3C protects male germ cells from

- transposon activity. **Science**, v. 354, n. 6314, p. 909–912, 2016.
- BARRETT, T. et al. BioProject and BioSample databases at NCBI: Facilitating capture and organization of metadata. **Nucleic Acids Research**, v. 40, p. D57–63, 2012.
- BASSETT, A. S. et al. The schizophrenia phenotype in 22q11 deletion syndrome. **American Journal of Psychiatry**, v. 160, n. 9, p. 1580–6, 2003.
- BECK, C. R. et al. LINE-1 retrotransposition activity in human genomes. **Cell**, v. 141, n. 7, p. 1159–1170, 2010.
- BECKER, K. G. et al. Binding of the ubiquitous nuclear transcription factor YY1 to a cis regulatory sequence in the human LINE-1 transposable element. **Human Molecular Genetics**, v. 2, n. 10, p. 1697–1702, 1993.
- BELANCIO, V. P.; HEDGES, D. J.; DEININGER, P. Mammalian non-LTR retrotransposons: For better or worse, in sickness and in health. **Genome Research**, v. 18, p. 343–358, 2008.
- BEN-PORATH, I.; WEINBERG, R. A. The signals and pathways activating cellular senescence. **International Journal of Biochemistry and Cell Biology**, v. 37, n. 5, p. 961–976, 2005.
- BENITEZ-GUIJARRO, M. et al. RNase H2, mutated in Aicardi-Goutières syndrome, promotes LINE-1 retrotransposition. **The EMBO Journal**, v. 37, n. 15, p. e98506, 2018.
- BENJAMINI, Y.; HOCHBERG, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. **Journal of the Royal Statistical Society**, v. 57, n. 1, p. 289–300, 1995.
- BENTO, C. F. et al. Mammalian Autophagy: How Does It Work? **Annual Review of Biochemistry**, v. 85, p. 685–713, 2016.
- BÉTERMIER, M.; BERTRAND, P.; LOPEZ, B. S. Is Non-Homologous End-Joining Really an Inherently Error-Prone Process? **PLoS Genetics**, v. 10, n. 1, p. e1004086, 2014.
- BHARDWAJ, A. et al. Characterizing TDP-43 interaction with its RNA targets. **Nucleic Acids Research**, v. 41, n. 9, p. 5062–5074, May 2013.
- BOGERD, H. P. et al. APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells. **Nucleic acids research**, v. 34, n. 1, p. 89–95, 2006.
- BOISSINOT, S.; CHEVRET, P.; FURANO, A. V. L1 (LINE-1) retrotransposon evolution and amplification in recent human history. **Molecular Biology and Evolution**, v. 17, n. 6, p. 915–928, 2000.
- BOSTICK, M. et al. UHRF1 plays a role in maintaining DNA methylation in mammalian cells. **Science**, v. 317, n. 5845, p. 1760–1764, 2007.

- BRÉGNARD, C. et al. Upregulated LINE-1 Activity in the Fanconi Anemia Cancer Susceptibility Syndrome Leads to Spontaneous Pro-inflammatory Cytokine Production. **EBioMedicine**, v. 8, p. 184–194, 2016.
- BUCHAN, J. R. et al. Eukaryotic stress granules are cleared by autophagy and Cdc48/VCP function. **Cell**, v. 153, n. 7, p. 1461–1474, 2013.
- BUCHON, N.; VAURY, C. RNAi: a defensive RNA-silencing against viruses and transposable elements. **Heredity**, v. 96, n. 2, p. 195–202, 2006.
- BUISSON, R.; MASSON, J. Y. PALB2 self-interaction controls homologous recombination. **Nucleic Acids Research**, v. 40, n. 20, p. 10312–10323, 2012.
- BULUT-KARSLIOGLU, A. et al. Suv39h-Dependent H3K9me3 Marks Intact Retrotransposons and Silences LINE Elements in Mouse Embryonic Stem Cells. **Molecular Cell**, v. 55, n. 2, p. 277–290, 2014.
- BUNDO, M. et al. Increased L1 retrotransposition in the neuronal genome in schizophrenia. **Neuron**, v. 81, n. 2, p. 306–313, 2014.
- CAPPELLI, E. et al. Homologous recombination proteins are associated with centrosomes and are required for mitotic stability. **Experimental Cell Research**, v. 317, n. 8, p. 1203–1213, 2011.
- CASTELLI, J. A. C. et al. The role of 2'-5' oligoadenylate-activated ribonuclease L in apoptosis. **Cell Death and Differentiation**, v. 5, n. 4, p. 313–20, 1998.
- CASTRO-DIAZ, N. et al. Evolutionally dynamic L1 regulation in embryonic stem cells. **Genes and Development**, v. 28, n. 13, p. 1397–1409, 2014.
- CASTRO-DIAZ, N.; FRIEDLI, M.; TRONO, D. Drawing a fine line on endogenous retroelement activity. **Mobile genetic elements**, v. 5, n. 1, p. 1–6, 2015.
- CHEN, H. et al. APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons. **Current biology : CB**, v. 16, n. 5, p. 480–5, 2006.
- CHEN, J. M. et al. A systematic analysis of LINE-1 endonuclease-dependent retrotranspositional events causing human genetic disease. **Human Genetics**, v. 117, n. 5, p. 411–27, 2005.
- CHEN, L. et al. Naturally occurring endo-siRNA silences LINE-1 retrotransposons in human cells through DNA methylation. **Epigenetics**, v. 7, n. 7, p. 758–771, 2012.
- CHENG, Y. S. et al. MAEL promoter hypermethylation is associated with de-repression of LINE-1 in human hypospermatogenesis. **Human Reproduction**, v. 32, n. 12, p. 2373–2381, 2017.
- CHIU, Y.-L. et al. High-molecular-mass APOBEC3G complexes restrict Alu retrotransposition. **Proceedings of the National Academy of Sciences of the United States of America**, v. 103, n. 42, p. 15588–93, 2006.

- CLÉMOT, M. et al. The replicative histone chaperone CAF1 is essential for the maintenance of identity and genome integrity in adult stem cells. **Development (Cambridge, England)**, v. 145, n. 17, p. dev161190, 2018.
- COUFAL, N. G. L1 retrotransposition in human neural progenitor cells. **Nature**, v. 460, n. 7259, p. 1127–31, 2009.
- COUFAL, N. G. et al. Ataxia telangiectasia mutated (ATM) modulates long interspersed element-1 (L1) retrotransposition in human neural stem cells. **Proceedings of the National Academy of Sciences of the United States of America**, v. 158, n. 51, p. 20382–7, 2011.
- CRICHTON, J. H. et al. Defending the genome from the enemy within: Mechanisms of retrotransposon suppression in the mouse germline. **Cellular and Molecular Life Sciences**, v. 71, n. 9, p. 1581–605, 2014.
- CUELLAR, T. L. et al. Silencing of retrotransposons by SETDB1 inhibits the interferon response in acute myeloid leukemia. **The Journal of cell biology**, v. 216, n. 11, p. 3535–3549, 2017.
- DAI, L. et al. Poly(A) Binding Protein C1 Is Essential for Efficient L1 Retrotransposition and Affects L1 RNP Formation. **Molecular and Cellular Biology**, v. 32, n. 21, p. 4323–4336, 2012.
- DE KONING, A. P. J. et al. Repetitive elements may comprise over Two-Thirds of the human genome. **PLoS Genetics**, v. 7, n. 12, p. e1002384, 2011.
- DE RUBEIS, S. et al. Synaptic, transcriptional and chromatin genes disrupted in autism. **Nature**, v. 515, n. 7526, p. 209–215, 2014.
- DEHARO, D. et al. Regulation of L1 expression and retrotransposition by melatonin and its receptor: Implications for cancer risk associated with light exposure at night. **Nucleic Acids Research**, v. 42, n. 12, p. 7694–7707, 2014.
- DENLI, A. M. et al. Primate-Specific ORF0 Contributes to Retrotransposon-Mediated Diversity. **Cell**, v. 163, n. 3, p. 583–593, 2015.
- DERE, E. et al. Heterozygous Ambra1 Deficiency in Mice: A Genetic Trait with Autism-Like Behavior Restricted to the Female Gender. **Frontiers in Behavioral Neuroscience**, v. 8, p. 181, 2014.
- DEWANNIEUX, M.; ESNAULT, C.; HEIDMANN, T. LINE-mediated retrotransposition of marked Alu sequences. **Nature Genetics**, v. 35, n. 1, p. 41–48, 2003.
- DI BARTOLOMEO, S. et al. The dynamic interaction of AMBRA1 with the dynein motor complex regulates mammalian autophagy. **Journal of Cell Biology**, v. 191, n. 1, p. 155–168, 2010.
- DI GIACOMO, M. et al. Multiple Epigenetic Mechanisms and the piRNA Pathway Enforce LINE1 Silencing during Adult Spermatogenesis. **Molecular Cell**, v. 50, n. 4, p. 601–

608, 2013.

DI GIACOMO, M. et al. G9a co-suppresses LINE1 elements in spermatogonia. **Epigenetics & chromatin**, v. 7, n. 1, p. 24, 2014.

DOBIN, A. et al. STAR: ultrafast universal RNA-seq aligner. **Bioinformatics (Oxford, England)**, v. 29, n. 1, p. 15–21, 2013.

DOMBROSKI, B. A et al. Isolation of an active human transposable element. **Science (New York, N.Y.)**, v. 254, n. 5039, p. 1805–1808, 1991.

DONG, J. et al. UHRF1 suppresses retrotransposons and cooperates with PRMT5 and PIWI proteins in male germ cells. **Nature Communications**, v. 10, n. 1, p. 1–14, 2019.

DOUCET, A. J. et al. Characterization of LINE-1 ribonucleoprotein particles. **PLoS Genetics**, v. 6, n. 10, p. 1–19, 2010.

DU, J. et al. Nucleocytoplasmic shuttling of SAMHD1 is important for LINE-1 suppression. **Biochemical and Biophysical Research Communications**, v. 510, n. 4, p. 551–557, 2019.

ELBARBARY, R. A.; LUCAS, B. A.; MAQUAT, L. E. Retrotransposons as regulators of gene expression. **Science**, v. 351, n. 6274, p. aac7247, 2016.

ERWIN, J. A. et al. L1-associated genomic regions are deleted in somatic cells of the healthy human brain. **Nature Neuroscience**, v. 19, n. 12, p. 1583–1591, 2016.

ERWIN, J. A.; MARCHETTO, M. C.; GAGE, F. H. Mobile DNA elements in the generation of diversity and complexity in the brain. **Nature Reviews Neuroscience**, v. 15, n. 8, p. 497–506, 2014.

ESNAULT, C.; MAESTRE, J.; HEIDMANN, T. Human LINE retrotransposons generate processed pseudogenes. **Nature Genetics**, v. 24, n. 4, p. 363–367, Apr. 2000.

ESTEBAN-JURADO, C. et al. The Fanconi anemia DNA damage repair pathway in the spotlight for germline predisposition to colorectal cancer. **European Journal of Human Genetics**, v. 24, n. 10, p. 1501–1505, 2016.

EVRONY, G. D. et al. Single-neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain. **Cell**, v. 151, n. 3, p. 483–496, 2012.

FABREGAT, A. et al. Reactome diagram viewer: data structures and strategies to boost performance. **Bioinformatics (Oxford, England)**, v. 34, n. 7, p. 1208–1214, 2018.

FEDOROV, A. V.; LUKYANOV, D. V.; PODGORNAYA, O. I. Identification of the proteins specifically binding to the rat LINE1 promoter. **Biochemical and Biophysical Research Communications**, v. 340, n. 2, p. 553–559, 2006.

FIMIA, G. M. et al. Ambra1 regulates autophagy and development of the nervous system. **Nature**, v. 447, n. 7148, p. 1121–1125, 2007.

- FROST, R. J. A. et al. MOV10L1 is necessary for protection of spermatocytes against retrotransposons by Piwi-interacting RNAs. **Proceedings of the National Academy of Sciences of the United States of America**, v. 107, n. 26, p. 11847–52, 2010.
- FUCHS, G. et al. Structural and biochemical basis for misfolded RNA recognition by the Ro autoantigen. **Nature Structural & Molecular Biology**, v. 13, n. 11, p. 1002–1009, 2006.
- FUKUDA, K. et al. A CRISPR knockout screen identifies SETDB1-target retroelement silencing factors in embryonic stem cells. **Genome Research**, v. 28, n. 6, p. 846–858, 2018.
- FUNG, L. et al. MiR-128 restriction of LINE-1 (L1) retrotransposition is dependent on targeting hnRNPA1 mRNA. **International Journal of Molecular Sciences**, v. 20, n. 8, 2019.
- GARCIA-PEREZ, J. L.; WIDMANN, T. J.; ADAMS, I. R. The impact of transposable elements on mammalian development. **Development**, v. 143, n. 22, p. 4101–4114, 2016.
- GASIOR, S. L. et al. The human LINE-1 retrotransposon creates DNA double-strand breaks. **Journal of Molecular Biology**, v. 357, n. 5, p. 1383–1393, 2006.
- GASIOR, S. L.; ROY-ENGEL, A. M.; DEININGER, P. L. ERCC1/XPF limits L1 retrotransposition. **DNA Repair**, v. 7, n. 6, p. 983–989, 1 Jun. 2008.
- GHILDIYAL, M.; ZAMORE, P. D. Small silencing RNAs: An expanding universe. **Nature Reviews Genetics**, v. 10, p. 94–108, 2009.
- GILBERT, N. et al. Multiple Fates of L1 Retrotransposition Intermediates in Cultured Human Cells. **Molecular and Cellular Biology**, v. 25, n. 17, p. 7780–7795, 2005.
- GILBERT, N.; MORAN, J. V. Mammalian LINE-1 Retrotransposons and Related Elements. In: **Mobile DNA II**. Washington, DC: Craig N, Craigie R, Gellert M, Lambowitz A, 2002. p. 836–869.
- GIULIVI, C. et al. Mitochondrial dysfunction in autism. **JAMA - Journal of the American Medical Association**, v. 304, n. 21, p. 2389–2396, 2010.
- GOODIER, J. J. L. et al. MOV10 RNA Helicase Is a Potent Inhibitor of Retrotransposition in Cells. **PLoS Genetics**, v. 8, n. 10, p. e1002941, 2012.
- GOODIER, J. L. et al. LINE-1 ORF1 Protein Localizes in Stress Granules with Other RNA-Binding Proteins, Including Components of RNA Interference RNA-Induced Silencing Complex. **Molecular and Cellular Biology**, v. 27, n. 18, p. 6469–6483, 2007.
- GOODIER, J. L. et al. The Broad-Spectrum Antiviral Protein ZAP Restricts Human Retrotransposition. **PLoS Genetics**, v. 11, n. 5, p. e1005252, 2015.
- GOODIER, J. L. Restricting retrotransposons: A review. **Mobile DNA**, v. 7, n. 1, p. 16,

2016.

- GOODIER, J. L.; CHEUNG, L. E.; KAZAZIAN, H. H. Mapping the LINE1 ORF1 protein interactome reveals associated inhibitors of human retrotransposition. **Nucleic Acids Research**, v. 41, n. 15, p. 7401–7419, 2013.
- GOODIER, J. L.; KAZAZIAN, H. H. Retrotransposons Revisited: The Restraint and Rehabilitation of Parasites. **Cell**, v. 135, n. 1, p. 23–25, 2008.
- GUO, H. et al. Autophagy supports genomic stability by degrading retrotransposon RNA. **Nature Communications**, v. 5, p. 5276, 4 Nov. 2014.
- HAGAN, C. R.; RUDIN, C. M. DNA cleavage and Trp53 differentially affect SINE transcription. **Genes Chromosomes and Cancer**, v. 46, n. 3, p. 248–260, 2007.
- HAMDORF, M. et al. miR-128 represses L1 retrotransposition by binding directly to L1 RNA. **Nature Publishing Group**, v. 22, n. 10, p. 824–831, 2015.
- HARRIS, C. R. et al. P53 Responsive Elements in Human Retrotransposons. **Oncogene**, v. 28, n. 44, p. 3857–65, 2009.
- HASHIMOTO, H. et al. UHRF1, a modular multi-domain protein, regulates replication-coupled crosstalk between DNA methylation and histone modifications. **Epigenetics**, v. 4, n. 1, p. 8–14, 2009.
- HATANAKA, Y. et al. Histone chaperone CAF-1 mediates repressive histone modifications to protect preimplantation mouse embryos from endogenous retrotransposons. **Proceedings of the National Academy of Sciences of the United States of America**, v. 112, n. 47, p. 14641–6, 2015.
- HAVECKER, E. R.; GAO, X.; VOYTAS, D. F. The diversity of LTR retrotransposons. **Genome Biology**, v. 5, n. 6, p. 225, 2004.
- HAZEN, J. L. et al. The Complete Genome Sequences, Unique Mutational Spectra, and Developmental Potency of Adult Neurons Revealed by Cloning. **Neuron**, v. 89, n. 6, p. 1223–1236, 2016.
- HEMPFLING, A. L. et al. Expression patterns of HENMT1 and PIWIL1 in human testis: implications for transposon expression. **Reproduction**, v. 154, n. 4, p. 363–374, 2017.
- HERAS, S. R. et al. The Microprocessor controls the activity of mammalian retrotransposons. **Nat Struct Mol Biol**, v. 20, n. 10, p. 1173–1181, 2013.
- HERAS, S. R. et al. Control of mammalian retrotransposons by cellular RNA processing activities. **Mobile genetic elements**, v. 4, n. 2, p. e28439, 6 Mar. 2014.
- HOWLETT, N. G. et al. The Fanconi anemia pathway is required for the DNA replication stress response and for the regulation of common fragile site stability. **Human Molecular Genetics**, v. 14, n. 5, p. 693–701, 2005.

- HU, S. et al. SAMHD1 Inhibits LINE-1 Retrotransposition by Promoting Stress Granule Formation. **PLoS Genetics**, v. 11, n. 7, p. e1005367, 2015.
- ICHIYANAGI, T. et al. HSP90 α plays an important role in piRNA biogenesis and retrotransposon repression in mouse. **Nucleic acids research**, v. 42, n. 19, p. 11903–11911, 2014.
- IKEDA, T. et al. Intrinsic restriction activity by apolipoprotein B mRNA editing enzyme APOBEC1 against the mobility of autonomous retrotransposons. **Nucleic Acids Research**, v. 39, n. 13, p. 5538–5554, 2011.
- IVASHKIV, L. B.; DONLIN, L. T. Regulation of type I interferon responses. **Nature Reviews Immunology**, v. 14, n. 1, p. 36–49, 2014.
- IYENGAR, S.; FARNHAM, P. J. KAP1 Protein: An Enigmatic Master Regulator of the Genome. **The Journal of Biological Chemistry**, v. 286, n. 30, p. 26267, 2011.
- JACOB-HIRSCH, J. et al. Whole-genome sequencing reveals principles of brain retrotransposition in neurodevelopmental disorders. **Cell Research**, v. 28, n. 2, p. 187–203, 2018.
- JACOBS, F. M. J. et al. An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. **Nature**, v. 516, n. 7530, p. 242–5, 2014.
- JAMES, S. J. et al. Complex epigenetic regulation of Engrailed-2 (EN-2) homeobox gene in the autism cerebellum. **Translational Psychiatry**, v. 3, p. e232, 2013.
- JOHNSON, R. D.; LIU, N.; JASIN, M. Mammalian XRCC2 promotes the repair of DNA double-strand breaks by homologous recombination. **Nature**, v. 401, n. 6751, p. 397–9, 1999.
- JÖNSSON, M. E. et al. Activation of neuronal genes via LINE-1 elements upon global DNA demethylation in human neural progenitors. **Nature Communications**, v. 10, n. 1, p. 1–11, 2019.
- JURKA, J. Sequence patterns indicate an enzymatic involvement in integration of mammalian retrotransposons. **Proceedings of the National Academy of Sciences of the United States of America**, v. 94, n. 4, p. 1872–1877, 1997.
- KAER, K. et al. Intronic L1 retrotransposons and nested genes cause transcriptional interference by inducing intron retention, exonization and cryptic polyadenylation. **PLoS ONE**, v. 6, n. 10, p. e26099, 2011.
- KANEHISA, M. et al. New approach for understanding genome variations in KEGG. **Nucleic Acids Research**, v. 47, n. D1, p. D590–D595, 2019.
- KANEHISA, M.; GOTO, S. KEGG: kyoto encyclopedia of genes and genomes. **Nucleic acids research**, v. 28, n. 1, p. 27–30, 2000.
- KARIMI, M. M. et al. DNA methylation and SETDB1/H3K9me3 regulate predominantly

- distinct sets of genes, retroelements, and chimeric transcripts in mescs. **Cell Stem Cell**, v. 8, n. 6, p. 676–687, 2011.
- KATO, Y. et al. Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. **Human Molecular Genetics**, v. 16, n. 19, p. 2272–80, 2007.
- KAWANO, K. et al. HIV-1 Vpr and p21 restrict LINE-1 mobility. **Nucleic Acids Research**, v. 46, n. 16, p. 8454–8470, 2018.
- KAZAZIAN, H. H. L1 retrotransposons shape the mammalian genome. **Science**, v. 289, n. 5482, p. 1152–1153, 2000.
- KĘDZIERSKA, H.; PIEKIEŁKO-WITKOWSKA, A. Splicing factors of SR and hnRNP families as regulators of apoptosis in cancer. **Cancer Letters**, v. 396, p. 53–65, 2017.
- KEMP, J. R.; LONGWORTH, M. S. Crossing the LINE Toward Genomic Instability: LINE-1 Retrotransposition in Cancer. **Frontiers in chemistry**, v. 3, p. 68, 2015.
- KHAN, H.; SMIT, A.; BOISSINOT, S. Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. **Genome Research**, v. 16, n. 1, p. 78–87, 2005.
- KIM, S. et al. PRMT5 Protects Genomic Integrity during Global DNA Demethylation in Primordial Germ Cells and Preimplantation Embryos. **Molecular Cell**, v. 56, n. 4, p. 564–579, Nov. 2014.
- KINOMOTO, M. et al. All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition. **Nucleic Acids Research**, v. 35, n. 9, p. 2955–2964, 2007.
- KIRINO, Y.; MOURELATOS, Z. Mouse Piwi-interacting RNAs are 2'-O-methylated at their 3' termini. **Nature Structural and Molecular Biology**, v. 14, n. 4, p. 347–348, 2007.
- KITCHENHAM, B. AND CHARTERS, S. Guidelines for performing systematic literature reviews in software engineering. **Technical report, Ver. 2.3 EBSE Technical Report. EBSE**, 2007.
- KOJIMA-KITA, K. et al. MIWI2 as an Effector of DNA Methylation and Gene Silencing in Embryonic Male Germ Cells. **Cell Reports**, v. 16, n. 11, p. 2819–2828, 2016.
- KUBO, S. et al. L1 retrotransposition in nondividing and primary human somatic cells. **Proceedings of the National Academy of Sciences of the United States of America**, v. 103, n. 21, p. 8036–8041, 2006.
- KUNTZ, S. et al. Tex19, a Mammalian-Specific Protein with a Restricted Expression in Pluripotent Stem Cells and Germ Line. **Stem Cells**, v. 26, n. 3, p. 734–744, 2008.
- KUSHIMA, I. et al. Comparative Analyses of Copy-Number Variation in Autism Spectrum Disorder and Schizophrenia Reveal Etiological Overlap and Biological Insights. **Cell Reports**, v. 24, n. 11, p. 2838–2856, 2018.

- KUWABARA, T. et al. Wnt-mediated activation of NeuroD1 and retro-elements during adult neurogenesis. **Nature neuroscience**, v. 12, n. 9, p. 1097–1105, 2009.
- LABAER, J. et al. New functional activities for the p21 family of CDK inhibitors. **Genes and Development**, v. 11, n. 7, p. 847–862, 1997.
- LANDER, E. S. et al. Initial sequencing and analysis of the human genome. **Nature**, v. 409, n. 6822, p. 860–921, 2001.
- LANGMEAD, B.; SALZBERG, S. L. Fast gapped-read alignment with Bowtie 2. **Nat Methods**, v. 9, n. 4, p. 357–359, 2012.
- LEE, D.; SHIN, C. Emerging roles of DROSHA beyond primary microRNA processing. **RNA Biology**, v. 15, n. 2, p. 186–193, 2018.
- LEE, S. H. et al. The impact of CpG island on defining transcriptional activation of the mouse L1 retrotransposable elements. **PLoS ONE**, v. 5, n. 6, 2010.
- LEHALLE, D. et al. STAG1 mutations cause a novel cohesinopathy characterised by unspecific syndromic intellectual disability. **Journal of Medical Genetics**, v. 54, n. 7, p. 479–488, 2017.
- LEIBOLD, D. M. et al. Translation of LINE-1 DNA elements in vitro and in human cells. **Proceedings of the National Academy of Sciences of the United States of America**, v. 87, n. 18, p. 6990–4, 1990.
- LEVIN, H. L.; MORAN, J. V. Dynamic interactions between transposable elements and their hosts. **Nature Reviews Genetics**, v. 12, n. 9, p. 615–627, 18 Aug. 2011.
- LEVY, S. et al. The diploid genome sequence of an individual human. **PLoS Biology**, v. 5, n. 10, p. e254, 2007.
- LI, M. Y. et al. Long interspersed nucleotide acid element-1 ORF-1 protein promotes proliferation and invasion of human colorectal cancer LoVo cells through enhancing ETS-1 activity. **Genetics and Molecular Research**, v. 13, n. 3, p. 6981–6994, 2014.
- LI, S. et al. Hypomethylation of LINE-1 elements in schizophrenia and bipolar disorder. **Journal of Psychiatric Research**, v. 107, n. 1, p. 68–72, 2018.
- LI, X. et al. The MOV10 helicase inhibits LINE-1 mobility. **The Journal of biological chemistry**, v. 288, n. 29, p. 21148–60, 2013.
- LI, Z. et al. Distinct roles of DNMT1-dependent and DNMT1-independent methylation patterns in the genome of mouse embryonic stem cells. **Genome Biology**, v. 16, n. 1, p. 115, 2015.
- LIANG, G. et al. Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements. **Mol Cell Biol**, v. 22, n. 2, p. 480–491, 2002.
- LIANG, W. et al. APOBEC3DE Inhibits LINE-1 Retrotransposition by Interacting with ORF1p and Influencing LINE Reverse Transcriptase Activity. **PloS one**, v. 11, n. 7,

- p. e0157220, 2016.
- LIAO, Y.; SMYTH, G. K.; SHI, W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. **Bioinformatics**, v. 30, n. 7, p. 923–930, 2014.
- LIM, A. K. et al. The nuage mediates retrotransposon silencing in mouse primordial ovarian follicles. **Development (Cambridge)**, v. 140, n. 18, p. 3819–3825, 2013.
- LIM, S. L. et al. HENMT1 and piRNA Stability Are Required for Adult Male Germ Cell Transposon Repression and to Define the Spermatogenic Program in the Mouse. **PLoS Genetics**, v. 11, n. 10, 2015.
- LINDIČ, N. et al. Differential inhibition of LINE1 and LINE2 retrotransposition by vertebrate AID/APOBEC proteins. **Retrovirology**, v. 10, p. 156, 2013.
- LIU, E. Y. et al. Loss of Nuclear TDP-43 Is Associated with Decondensation of LINE Retrotransposons. **Cell Reports**, v. 27, n. 5, p. 1409–1421.e6, 2019.
- LIU, N. et al. Selective silencing of euchromatic L1s revealed by genome-wide screens for L1 regulators. **Nature**, v. 553, n. 7687, p. 228–232, 2018.
- LONSDALE, J. et al. The Genotype-Tissue Expression (GTEx) project. **Nature Genetics**, v. 45, p. 580–585, 2013.
- LOVE, M. I.; HUBER, W.; ANDERS, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. **Genome Biology**, v. 15, n. 12, p. 1–34, 2014.
- LUTZ, S. M. et al. Allelic Heterogeneity in LINE-1 Retrotransposition Activity. **American Journal of Human Genetics**, v. 73, n. 6, p. 1431–1437, 2003.
- MA, L. et al. GASZ is essential for male meiosis and suppression of retrotransposon expression in the male germline. **PLoS Genetics**, v. 5, n. 9, p. e1000635, 2009.
- MACDUFF, D. A.; DEMOREST, Z. L.; HARRIS, R. S. AID can restrict L1 retrotransposition suggesting a dual role in innate and adaptive immunity. **Nucleic Acids Research**, v. 37, n. 6, p. 1854–1867, 2009.
- MACIA, A. et al. Engineered LINE-1 retrotransposition in nondividing human neurons. **Genome Research**, v. 27, n. 3, p. 335–348, 2017.
- MACIAS, S. et al. DGCR8 HITS-CLIP reveals novel functions for the Microprocessor. **Nature structural & molecular biology**, v. 19, n. 8, p. 760–766, 2012.
- MACLENNAN, M. et al. Mobilization of LINE-1 retrotransposons is restricted by Tex19.1 in mouse embryonic stem cells. **eLife**, v. 6, p. e26152, 2017.
- MACMICKING, J. D. Interferon-inducible effector mechanisms in cell-autonomous immunity. **Nature Reviews Immunology**, v. 12, n. 5, p. 367–382, 2012.

- MALONE, C. D.; HANNON, G. J. Small RNAs as Guardians of the Genome. **Cell**, v. 236, n. 4, p. 656–68, 2009.
- MARANGONI, C.; DE CHIARA, L.; FAEDDA, G. L. Bipolar Disorder and ADHD: Comorbidity and Diagnostic Distinctions. **Current Psychiatry Reports**, v. 17, n. 8, p. 604, 2015.
- MARCHETTO, M. C. N. et al. Differential L1 regulation in pluripotent stem cells of humans and apes. **Nature**, v. 503, p. 525–529, 2013.
- MARCHETTO, M. C. N. M. et al. A model for neural development and treatment of Rett syndrome using human induced pluripotent stem cells. **Cell**, v. 143, n. 4, p. 527–39, 2010.
- MARTENS, J. H. A. et al. The profile of repeat-associated histone lysine methylation states in the mouse epigenome. **The EMBO journal**, v. 24, n. 4, p. 800–12, 2005.
- MARTIN, S. L. The ORF1 protein encoded by LINE-1: Structure and function during L1 retrotransposition. **Journal of Biomedicine and Biotechnology**, v. 2006, n. 1, p. 45621, 2006.
- MATHIAS, S. L. et al. Reverse transcriptase encoded by a human transposable element. **Science**, v. 254, n. 5039, p. 1808–1810, 1991.
- MCCONNELL, M. J. et al. Mosaic copy number variation in human neurons. **Science**, v. 342, n. 6158, p. 632–637, 2013.
- MEETEI, A. R. et al. A human ortholog of archaeal DNA repair protein Hef is defective in Fanconi anemia complementation group M. **Nature Genetics**, v. 37, n. 9, p. 958–963, 2005.
- MERENLENDER-WAGNER, A. et al. Autophagy has a key role in the pathophysiology of schizophrenia. **Molecular Psychiatry**, v. 20, n. 1, p. 126–132, 2015.
- METZNER, M.; JÄCK, H. M.; WABL, M. LINE-1 Retroelements Complexed and Inhibited by Activation Induced Cytidine Deaminase. **PLoS ONE**, v. 7, n. 11, 2012.
- MIKL, M. C. et al. Mice Deficient in APOBEC2 and APOBEC3. **Molecular and Cellular Biology**, v. 25, n. 16, p. 7270, 2005.
- MIRET, N. et al. AhR ligands reactivate LINE-1 retrotransposon in triple-negative breast cancer cells MDA-MB-231 and non-tumorigenic mammary epithelial cells NMuMG. **Biochemical Pharmacology**, v. 175, p. 113904, 2020.
- MISIAK, B. et al. Lower LINE-1 methylation in first-episode schizophrenia patients with the history of childhood trauma. **Epigenomics**, v. 7, n. 8, p. 1275–1285, 2015.
- MITA, P. et al. BRCA1 and S phase DNA repair pathways restrict LINE-1 retrotransposition in human cells. **Nature Structural and Molecular Biology**, v. 27, n. 2, p. 179–191, 2020.

- MITJANS, M. et al. Sexual dimorphism of AMBRA1-related autistic features in human and mouse. **Translational psychiatry**, v. 7, p. e1247, 2017.
- MIYOSHI, T.; MAKINO, T.; MORAN, J. V. Poly(ADP-Ribose) Polymerase 2 Recruits Replication Protein A to Sites of LINE-1 Integration to Facilitate Retrotransposition. **Molecular Cell**, v. 75, n. 6, p. 1286- 1298.e12, 2019.
- MOLDOVAN, G. L.; D'ANDREA, A. D. How the fanconi anemia pathway guards the genome. **Annual Review of Genetics**, v. 43, p. 223–249, 2009.
- MOLDOVAN, J. B.; MORAN, J. V. The Zinc-Finger Antiviral Protein ZAP Inhibits LINE and Alu Retrotransposition. **PLoS Genetics**, v. 11, n. 5, p. e1005121, 2015.
- MONTOYA-DURANGO, D. E. et al. Epigenetic control of mammalian LINE-1 retrotransposon by retinoblastoma proteins. **Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis**, v. 665, n. 1–2, p. 20–28, 2009.
- MONTOYA-DURANGO, D. E. et al. LINE-1 silencing by retinoblastoma proteins is effected through the nucleosomal and remodeling deacetylase multiprotein complex. **BMC cancer**, v. 16, p. 38, 2016.
- MORRISH, T. A. et al. Endonuclease-independent LINE-1 retrotransposition at mammalian telomeres. **Nature**, v. 446, n. 7132, p. 208–212, 2007.
- MUCKENFUSS, H. et al. APOBEC3 proteins inhibit human LINE-1 retrotransposition. **Journal of Biological Chemistry**, v. 281, n. 31, p. 22161–22172, 2006.
- MULERO-NAVARRO, S.; FERNANDEZ-SALGUERO, P. M. New Trends in Aryl Hydrocarbon Receptor Biology. **Frontiers in cell and developmental biology**, v. 4, p. 45, 2016.
- MULLIN, A. P. et al. Neurodevelopmental disorders: Mechanisms and boundary definitions from genomes, interactomes and proteomes. **Translational Psychiatry**, v. 3, n. 12, p. e329, 2013.
- MUÑOZ-LÓPEZ, M.; GARCÍA-PÉREZ, J. L. DNA transposons: nature and applications in genomics. **Current genomics**, v. 11, n. 2, p. 115–28, 2010.
- MUOTRI, A. R. et al. Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. **Nature**, v. 435, n. 7044, p. 903–910, 2005.
- MUOTRI, A. R. et al. L1 retrotransposition in neurons is modulated by MeCP2. **Nature**, v. 468, n. 7322, p. 443–446, 2010.
- MURAMATSU, M. et al. Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. **The Journal of biological chemistry**, v. 274, n. 26, p. 18470–6, 1999.
- MYERS, J. S. et al. A comprehensive analysis of recently integrated human Ta L1 elements. **American Journal of Human Genetics**, v. 71, n. 2, p. 312–326, 2002.

- NAIM, V.; ROSSELLI, F. The FANC pathway and BLM collaborate during mitosis to prevent micro-nucleation and chromosome abnormalities. **Nature Cell Biology**, v. 11, n. 6, p. 761–768, 2009.
- NAKANISHI, K. et al. Human Fanconi anemia monoubiquitination pathway promotes homologous DNA repair. **Proceedings of the National Academy of Sciences of the United States of America**, v. 102, n. 4, p. 1110–1115, 2005.
- NAZIO, F. et al. MTOR inhibits autophagy by controlling ULK1 ubiquitylation, self-association and function through AMBRA1 and TRAF6. **Nature Cell Biology**, v. 15, n. 4, p. 406–416, 2013.
- NEUL, J. L.; ZOGHBI, H. Y. Rett Syndrome: A Prototypical Neurodevelopmental Disorder. **Neuroscientist**, v. 10, n. 2, p. 118–128, 2004.
- NIE, Y. et al. FANCD2 is required for the repression of germline transposable elements. **Reproduction**, v. 159, n. 6, p. 659–668, 2020.
- NISHIMURA, T. et al. Human MARF1 is an endoribonuclease that interacts with the DCP1:2 decapping complex and degrades target mRNAs. **Nucleic Acids Research**, v. 46, n. 22, p. 12008–12021, 2018.
- ODEMIS, S. et al. Association Between Polymorphisms of DNA Repair Genes and Risk of Schizophrenia. **Genetic Testing and Molecular Biomarkers**, v. 20, n. 1, p. 11–17, 2015.
- OLLINGER, R. et al. Deletion of the pluripotency-associated Tex19.1 gene causes activation of endogenous retroviruses and defective spermatogenesis in mice. **PLoS Genetics**, v. 4, n. 9, p. e1000199, 2008.
- ONO, T. et al. Differential contributions of condensin I and condensin II to mitotic chromosome architecture in vertebrate cells. **Cell**, v. 115, n. 1, p. 109–121, 2003.
- ORECCHINI, E. et al. ADAR1 restricts LINE-1 retrotransposition. **Nucleic acids research**, v. 45, n. 1, p. 155–168, 2017.
- ORECCHINI, E.; FRASSINELLI, L.; MICHIEZI, A. Restricting retrotransposons: ADAR1 is another guardian of the human genome. **RNA Biology**, v. 14, n. 11, p. 1485–1491, 2017.
- ORQUEDA, A. J. et al. SOX -11 regulates LINE -1 retrotransposon activity during neuronal differentiation. **FEBS Letters**, v. 592, n. 22, p. 3708–3719, 2018.
- OSTERTAG, E. M.; KAZAZIAN, H. H. Biology of Mammalian L1 Retrotransposons. **Annual Review of Genetics**, v. 35, p. 501–538, 2001a.
- OSTERTAG, E. M.; KAZAZIAN, H. H. Twin priming: A proposed mechanism for the creation of inversions in L1 retrotransposition. **Genome Research**, v. 11, n. 12, p. 2059–2065, 2001b.

- PANDEY, R. R. et al. Tudor domain containing 12 (TDRD12) is essential for secondary PIWI interacting RNA biogenesis in mice. **Proceedings of the National Academy of Sciences of the United States of America**, v. 110, n. 41, p. 16492–7, 2013.
- PASTOR, W. A. et al. MORC1 represses transposable elements in the mouse male germline. **Nature communications**, v. 5, p. 5795, 2014.
- PATEL, R. K.; JAIN, M. NGS QC toolkit: A toolkit for quality control of next generation sequencing data. **PLoS ONE**, v. 7, n. 2, p. e30619, 2012.
- PEDDIGARI, S. et al. HnRNPL and nucleolin bind LINE-1 RNA and function as host factors to modulate retrotransposition. **Nucleic Acids Research**, v. 41, n. 1, p. 575–585, 2013.
- PETRUSEVA, I. O.; EVDOKIMOV, A. N.; LAVRIK, O. I. Molecular mechanism of global genome nucleotide excision repair. **Acta Naturae**, v. 6, n. 1, p. 23–34, 2014.
- PIZARRO, J. G.; CRISTOFARI, G. Post-transcriptional control of LINE-1 retrotransposition by cellular host factors in somatic cells. **Frontiers in Cell and Developmental Biology**, v. 4, p. 14, 2016.
- PRABAKARAN, S. et al. Mitochondrial dysfunction in schizophrenia: Evidence for compromised brain metabolism and oxidative stress. **Molecular Psychiatry**, v. 9, p. 684–697, 2004.
- PUSZYK, W. et al. The epigenetic regulator PLZF represses L1 retrotransposition in germ and progenitor cells. **The EMBO journal**, v. 32, n. 13, p. 1941–1952, 2013.
- RABANAL-RUIZ, Y.; OTTEN, E. G.; KOROLCHUK, V. I. mTORC1 as the main gateway to autophagy. **Essays In Biochemistry**, v. 61, n. 6, p. 565–584, 2017.
- RAIZ, J. et al. The non-autonomous retrotransposon SVA is trans -mobilized by the human LINE-1 protein machinery. **Nucleic Acids Research**, v. 40, n. 4, p. 1666–1683, 2012.
- RAJAN, K. S.; RAMASAMY, S. Retrotransposons and piRNA: The missing link in central nervous system. **Neurochemistry International**, v. 77, p. 94–102, 2014.
- RAPOPORT, J. L.; GIEDD, J. N.; GOGTAY, N. Neurodevelopmental model of schizophrenia: Update 2012. **Molecular Psychiatry**, v. 17, n. 12, p. 1228–1238, 2012.
- RAY, D. A. et al. Multiple waves of recent DNA transposon activity in the bat, *Myotis lucifugus*. **Genome research**, v. 18, n. 5, p. 717–28, 2008.
- REICHMANN, J. et al. The genome-defence gene Tex19.1 suppresses LINE-1 retrotransposons in the placenta and prevents intra-uterine growth retardation in mice. **Human molecular genetics**, v. 22, n. 9, p. 1791–806, 2013.
- RENNER, T. M. et al. Characterization of molecular attributes that influence LINE-1

- restriction by all seven human APOBEC3 proteins. **Virology**, v. 520, p. 127–136, 2018.
- REUTER, M. et al. Loss of the Mili-interacting Tudor domain-containing protein-1 activates transposons and alters the Mili-associated small RNA profile. **Nature structural & molecular biology**, v. 16, n. 6, p. 639–46, 2009.
- REUTER, M. et al. Miwi catalysis is required for piRNA amplification-independent LINE1 transposon silencing. **Nature**, v. 480, n. 7376, p. 264–7, 2011.
- RICHARDSON, S. R. et al. APOBEC3A deaminates transiently exposed single-strand DNA during LINE-1 retrotransposition. **eLife**, v. 3, p. e02008, 2014.
- RICHARDSON, S. R. et al. The Influence of LINE-1 and SINE Retrotransposons on Mammalian Genomes. In: **Mobile DNA III**. [s.l.] American Society of Microbiology, 2015. v. 3p. 1165–1208.
- RICHARDSON, S. R. et al. Heritable L1 retrotransposition in the mouse primordial germline and early embryo. **Genome Research**, v. 27, n. 8, p. 1395–1405, 2017.
- RIESS, M. et al. Interferons Induce Expression of SAMHD1 in Monocytes through Down-regulation of miR-181a and miR-30a. **The Journal of biological chemistry**, v. 292, n. 1, p. 264–277, 2017.
- RIETSCHEL, M. et al. Association between genetic variation in a region on chromosome 11 and schizophrenia in large samples from Europe. **Molecular Psychiatry**, v. 17, n. 9, p. 906–917, 2012.
- ROBERTS, J. T.; CARDIN, S. E.; BORCHERT, G. M. Burgeoning evidence indicates that microRNAs were initially formed from transposable element sequences. **Mobile Genetic Elements**, v. 4, p. e29255, 2014.
- RODRÍGUEZ-MARTÍN, C. et al. Familial retinoblastoma due to intronic LINE-1 insertion causes aberrant and noncanonical mRNA splicing of the RB1 gene. **Journal of Human Genetics**, v. 61, n. 5, p. 463–466, 2016.
- ROMAN, A. C. et al. Genome-wide B1 retrotransposon binds the transcription factors dioxin receptor and Slug and regulates gene expression in vivo. **Proceedings of the National Academy of Sciences of the United States of America**, v. 105, n. 5, p. 1632–7, 2008.
- ROUGET, C. et al. Maternal mRNA deadenylation and decay by the piRNA pathway in the early Drosophila embryo. **Nature**, v. 467, n. 7319, p. 1128–1132, 2010.
- ROWE, H. M. et al. KAP1 controls endogenous retroviruses in embryonic stem cells. **Nature**, v. 463, n. 7278, p. 237–240, 2010.
- SAADAT, M.; PAKYARI, N.; FARRASHBANDI, H. Genetic polymorphism in the DNA repair gene XRCC1 and susceptibility to schizophrenia. **Psychiatry Research**, v. 157, n. 1, p. 241–245, 2008.

- SAITO, K. et al. Pimet, the Drosophila homolog of HEN1, mediates 2'-O-methylation of Piwi-interacting RNAs at their 3' ends. **Genes and Development**, v. 21, n. 13, p. 1603–1608, 2007.
- SALTER, J. D.; BENNETT, R. P.; SMITH, H. C. The APOBEC Protein Family: United by Structure, Divergent in Function. **Trends in Biochemical Sciences**, v. 41, n. 7, p. 578–594, 2016.
- SASSAMAN, D. M. et al. Many human L1 elements are capable of retrotransposition. **Nature Genetics**, v. 16, n. 1, p. 37–43, 1997.
- SAXE, J. P. et al. Tdrkh is essential for spermatogenesis and participates in primary piRNA biogenesis in the germline. **The EMBO journal**, v. 32, n. 13, p. 1869–85, 2013.
- SBARDELLA, D. et al. Retention of Mitochondria in Mature Human Red Blood Cells as the Result of Autophagy Impairment in Rett Syndrome. **Scientific Reports**, v. 7, n. 1, p. 12297, 2017.
- SCAINI, G. et al. TSPO upregulation in bipolar disorder and concomitant downregulation of mitophagic proteins and NLRP3 inflammasome activation. **Neuropsychopharmacology**, v. 44, n. 7, p. 1291–1299, 2019.
- SCHÄRER, O. D. **Nucleotide excision repair in Eukaryotes** Cold Spring Harbor **Perspectives in Biology** Cold Spring Harbor Laboratory Press, , 2013.
- SCHNEIDER, V. A. et al. Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. **Genome Research**, v. 27, n. 5, p. 849–864, 2017.
- SERVANT, G. et al. The nucleotide excision repair pathway limits L1 retrotransposition. **Genetics**, v. 205, n. 1, p. 139–153, 2017.
- SHABAN, N. M. et al. The structure of the mammalian RNase H2 complex provides insight into RNA-DNA hybrid processing to prevent immune dysfunction. **Journal of Biological Chemistry**, v. 285, n. 6, p. 3617–3624, 2010.
- SHAIK, S. et al. Ubiquitination and selective autophagy. **Cell Death and Differentiation**, v. 20, n. 1, p. 21–30, 2013.
- SHILOH, Y. ATM (ataxia telangiectasia mutated): expanding roles in the DNA damage response and cellular homeostasis. **Biochemical Society Transactions**, v. 29, n. 6, p. 661, 2001.
- SHIROMOTO, Y. et al. GPAT2 is required for piRNA biogenesis, transposon silencing, and maintenance of spermatogonia in mice. **Biology of Reproduction**, v. 101, n. 1, p. 248–256, 2019.
- SHOJI, M. et al. The TDRD9-MIWI2 complex is essential for piRNA-mediated retrotransposon silencing in the mouse male germline. **Developmental cell**, v. 17, n. 6, p. 775–87, 2009.

- SHPYLEVA, S. et al. Overexpression of LINE-1 Retrotransposons in Autism Brain. **Molecular Neurobiology**, v. 55, n. 2, p. 1740–1749, 2018.
- SILVERMAN, R. H. et al. rRNA cleavage as an index of ppp(A2'p)(n)A activity in interferon-treated encephalomyocarditis virus-infected cells. **Journal of Virology**, v. 46, n. 3, p. 1051–1055, 1983.
- SILVERMAN, R. H. Viral encounters with 2',5'-oligoadenylate synthetase and RNase L during the interferon antiviral response. **Journal of virology**, v. 81, n. 23, p. 12720–9, 2007.
- SINGER, T. et al. LINE-1 retrotransposons: Mediators of somatic variation in neuronal genomes? **Trends in Neurosciences**, v. 33, n. 8, p. 345–354, 2010.
- SJAARDA, C. P. et al. Exome sequencing identifies de novo splicing variant in XRCC6 in sporadic case of autism. **Journal of Human Genetics**, v. 65, n. 3, p. 287–296, 2020.
- SKARIAH, G. et al. Mov10 suppresses retroelements and regulates neuronal development and function in the developing brain. **BMC Biology**, v. 15, n. 1, p. 54, 2017.
- SKENE, P. J. et al. Neuronal MeCP2 Is Expressed at Near Histone-Octamer Levels and Globally Alters the Chromatin State. **Molecular Cell**, v. 37, n. 4, p. 457–468, 2010.
- SMIT, A.; HUBLEY, R.; GREEN, P. **RepeatMasker Open-4.0.6 2013-2015**. <http://www.repeatmasker.org>, 2018.
- SOOKDEO, A.; HEPP, C. M.; BOISSINOT, S. Contrasted patterns of evolution of the LINE-1 retrotransposon in perissodactyls: The history of a LINE-1 extinction. **Mobile DNA**, v. 9, n. 12, 2018.
- SOPER, S. F. C. et al. Mouse maelstrom, a component of nuage, is essential for spermatogenesis and transposon repression in meiosis. **Developmental cell**, v. 15, n. 2, p. 285–97, 2008.
- SPEEK, M. Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. **Molecular and cellular biology**, v. 21, n. 6, p. 1973–85, 2001.
- STENGLEIN, M. D.; HARRIS, R. S. APOBEC3B and APOBEC3F inhibit L1 retrotransposition by a DNA deamination-independent mechanism. **Journal of Biological Chemistry**, v. 281, n. 25, p. 16837–16841, 2006.
- STETSON, D. B. et al. Trex1 prevents cell-intrinsic initiation of autoimmunity. **Cell**, v. 134, n. 4, p. 587–98, 2008.
- SU, Y. Q. et al. MARF1 regulates essential oogenic processes in mice. **Science**, v. 335, n. 6075, p. 1496–1499, 2012a.
- SU, Y. Q. et al. Meiosis arrest female 1 (MARF1) has nuage-like function in mammalian

- oocytes. **Proceedings of the National Academy of Sciences**, v. 109, n. 46, p. 18653–18660, 2012b.
- SUNG, P.; KLEIN, H. Mechanism of homologous recombination: Mediators and helicases take on regulatory functions. **Nature Reviews Molecular Cell Biology**, v. 7, n. 10, p. 739–750, 2006.
- SUZUKI, J. et al. Genetic evidence that the non-homologous end-joining repair pathway is involved in LINE retrotransposition. **PLoS genetics**, v. 5, n. 4, p. e1000461, 2009.
- SWERGOLD, G. D. Identification, characterization, and cell specificity of a human LINE-1 promoter. **Molecular and cellular biology**, v. 10, n. 12, p. 6718–29, 1990.
- SZKLARCZYK, D. et al. STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. **Nucleic Acids Research**, v. 47, n. D1, p. D607–D613, 2019.
- TANG, Z. et al. Human transposon insertion profiling: Analysis, visualization and identification of somatic LINE-1 insertions in ovarian cancer. **Proceedings of the National Academy of Sciences of the United States of America**, v. 114, n. 5, p. E733–E740, 2017.
- TAYLOR, M. S. et al. Affinity Proteomics Reveals Human Host Factors Implicated in Discrete Stages of LINE-1 Retrotransposition. **Cell**, v. 155, n. 5, p. 1034–1048, 2013.
- TCHASOVNIKAROVA, I. A. et al. Epigenetic silencing by the HUSH complex mediates position-effect variegation in human cells. **Science**, v. 348, n. 6242, p. 1481–1485, 2015.
- TCHÉNIO, T.; CASELLA, J. F.; HEIDMANN, T. Members of the SRY family regulate the human LINE retrotransposons. **Nucleic acids research**, v. 28, n. 2, p. 411–5, 2000.
- TEAM, R. C. **R: A Language and Environment for Statistical Computing**. Vienna: R Foundation for Statistical Computing, 2016.
- TENENG, I.; STRIBINSKIS, V.; RAMOS, K. S. Context-specific regulation of LINE-1. **Genes to Cells**, v. 12, n. 10, p. 1101–1110, 2007.
- TESLI, M. et al. Association analysis of PALB2 and BRCA2 in bipolar disorder and schizophrenia in a scandinavian case-control sample. **American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics**, v. 153B, n. 7, p. 1276–1282, 2010.
- THANSEEM, I. et al. Elevated Transcription Factor Specificity Protein 1 in Autistic Brains Alters the Expression of Autism Candidate Genes. **Biological Psychiatry**, v. 71, n. 5, p. 410–418, 2012.
- THOMAS, C. A. et al. Modeling of TREX1-Dependent Autoimmune Disease using Human Stem Cells Highlights L1 Accumulation as a Source of Neuroinflammation. **Cell Stem Cell**, v. 21, n. 3, p. 319–331, 2017.

- TUMBER, A. et al. Potent and Selective KDM5 Inhibitor Stops Cellular Demethylation of H3K4me3 at Transcription Start Sites and Proliferation of MM1S Myeloma Cells. **Cell Chemical Biology**, v. 24, n. 3, p. 371–380, 2017.
- TURELLI, P.; VIANIN, S.; TRONO, D. The innate antiretroviral factor APOBEC3G does not affect human LINE-1 retrotransposition in a cell culture assay. **Journal of Biological Chemistry**, v. 279, n. 42, p. 43371–43373, 2004.
- UPTON, K. R. et al. Ubiquitous L1 mosaicism in hippocampal neurons. **Cell**, v. 161, n. 2, p. 228–239, 2015.
- VAN METER, M. et al. SIRT6 represses LINE1 retrotransposons by ribosylating KAP1 but this repression fails with stress and age. **Nature Communications**, v. 5, p. 5011, 2014.
- VAN TWEST, S. et al. Mechanism of Ubiquitination and Deubiquitination in the Fanconi Anemia Pathway. **Molecular Cell**, v. 65, n. 2, p. 247–259, 2017.
- VARSHNEY, D. et al. SINE transcription by RNA polymerase III is suppressed by histone methylation but not by DNA methylation. **Nature communications**, v. 6, p. 6569, 2015.
- VAZQUEZ, B. N. et al. SIRT7 mediates L1 elements transcriptional repression and their association with the nuclear lamina. **Nucleic Acids Research**, v. 47, n. 15, p. 7870–7885, 2019.
- VÁZQUEZ, P. et al. Atg5 and Ambra1 differentially modulate neurogenesis in neural stem cells. **Autophagy**, v. 8, n. 2, p. 187–199, 2012.
- VOLKMANN, B. et al. Human TRIM5 α senses and restricts LINE-1 elements. **Proceedings of the National Academy of Sciences of the United States of America**, v. 117, n. 30, p. 17965–17976, 2020.
- VOLKMAR, F. R.; STATE, M.; KLIN, A. Autism and autism spectrum disorders: Diagnostic issues for the coming decade. **Journal of Child Psychology and Psychiatry and Allied Disciplines**, v. 50, n. 1–2, p. 108–115, 2009.
- VYAS, S. et al. Family-wide analysis of poly(ADP-ribose) polymerase activity. **Nature Communications**, v. 5, p. 4426, 2014.
- WAGSTAFF, B. J.; BARNERSSOI, M.; ROY-ENGEL, A. M. Evolutionary conservation of the functional modularity of primate and murine LINE-1 elements. **PLoS ONE**, v. 6, n. 5, p. e19672, 2011.
- WALD, A. Sequential tests of statistical hypotheses. **The Annals of Mathematical Statistics**, v. 16, n. 2, p. 117–186, 1945.
- WALLACE, N. A. et al. HPV 5 and 8 E6 Abrogate ATR activity resulting in increased persistence of UVB induced DNA damage. **PLoS Pathogens**, v. 8, n. 7, p. 41, 2012.

- WALLACE, N. A. et al. HPV 5 and 8 E6 expression reduces ATM protein levels and attenuates LINE-1 retrotransposition. **Virology**, v. 443, n. 1, p. 69–79, 2013.
- WANG, Y. et al. Mechanism of alternative splicing and its regulation. **Biomedical reports**, v. 3, n. 2, p. 152–158, 2015.
- WARD, J. R. et al. Condensin II and GAIT complexes cooperate to restrict LINE-1 retrotransposition in epithelial cells. **PLoS Genetics**, v. 13, n. 10, p. e1007051, 2017.
- WARKOCKI, Z. et al. Uridylation by TUT4/7 Restricts Retrotransposition of Human LINE-1s Article Uridylation by TUT4/7 Restricts Retrotransposition of Human LINE-1s. **Cell**, v. 174, n. 6, p. 1537–1548, 2018.
- WATANABE, T. et al. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. **Nature**, v. 453, n. 7194, p. 539–543, 2008.
- WATANABE, T. et al. MITOPLD is a mitochondrial protein essential for nuage formation and piRNA biogenesis in the mouse germline. **Developmental cell**, v. 20, n. 3, p. 364–75, 2011.
- WEICK, E. M.; MISKA, E. A. piRNAs: From biogenesis to function. **Development (Cambridge)**, v. 141, n. 18, p. 3458–3471, 2014.
- WHEELER, T. J. et al. Dfam: A database of repetitive DNA based on profile hidden Markov models. **Nucleic Acids Research**, v. 41, n. D1, p. D70–D82, 2013.
- WHITE, T. E. et al. Effects of human SAMHD1 polymorphisms on HIV-1 susceptibility. **Virology**, v. 460–461, n. 1, p. 34–44, 2014.
- WILLIAMS, E. et al. Transposable elements occur more frequently in autism-risk genes: Implications for the role of genomic instability in autism. **Translational Neuroscience**, v. 4, n. 2, p. 172–202, 2013.
- WISSING, S. et al. Reprogramming somatic cells into ips cells activates line-1 retroelement mobility. **Human Molecular Genetics**, v. 21, n. 1, p. 208–218, 2012.
- WU, T. P. et al. DNA methylation on N6-adenine in mammalian embryonic stem cells. **Nature**, v. 532, n. 7599, p. 1–18, 2016.
- WYLIE, A. et al. p53 genes function to restrain mobile elements. **Genes & development**, v. 30, n. 1, p. 64–77, 2016.
- XIANG, Y. et al. Histone demethylase KDM4B promotes DNA damage by activating long interspersed nuclear element-1. **Cancer Research**, v. 79, n. 1, p. 86–98, 2019.
- XIOL, J. et al. A Role for Fkbp6 and the Chaperone Machinery in piRNA Amplification and Transposon Silencing. **Molecular Cell**, v. 47, n. 6, p. 970–979, 2012.
- XU, H. et al. Cohesin Rad21 Mediates Loss of Heterozygosity and Is Upregulated via Wnt Promoting Transcriptional Dysregulation in Gastrointestinal Tumors. **Cell Reports**, v. 9, n. 5, p. 1781–1797, 2014.

- YABUTA, Y. et al. TDRD5 is required for retrotransposon silencing, chromatoid body assembly, and spermiogenesis in mice. **The Journal of cell biology**, v. 192, n. 5, p. 781–95, 2011.
- YAN, N. et al. The cytosolic exonuclease TREX1 inhibits the innate immune response to human immunodeficiency virus type 1. **Nature immunology**, v. 11, n. 11, p. 1005–13, 2010.
- YANG, J. et al. TP53 Polymorphism Contributes to the Susceptibility to Bipolar Disorder but Not to Schizophrenia in the Chinese Han Population. **Journal of Molecular Neuroscience**, v. 68, n. 4, p. 679–687, 2019a.
- YANG, N. et al. An important role for RUNX3 in human L1 transcription and retrotransposition. **Nucleic Acids Research**, v. 31, n. 16, p. 4929–4940, 2003.
- YANG, N.; KAZAZIAN, H. H. L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. **Nat Struct Mol Biol**, v. 13, n. 9, p. 763–771, 2006.
- YANG, W. R. et al. SQuIRE reveals locus-specific regulation of interspersed repeat expression. **Nucleic acids research**, v. 47, n. 5, p. e27, 2019b.
- YANG, Z. et al. Apolipoprotein(a) gene enhancer resides within a LINE element. **Journal of Biological Chemistry**, v. 273, n. 2, p. 891–897, 1998.
- YANG, Z. et al. PIWI Slicing and EXD1 Drive Biogenesis of Nuclear piRNAs from Cytosolic Targets of the Mouse piRNA Pathway. **Molecular cell**, v. 61, n. 1, p. 138–52, 2016.
- YAZDANKHAH, M. et al. The autophagy regulators Ambra1 and Beclin 1 are required for adult neurogenesis in the brain subventricular zone. **Cell Death and Disease**, v. 5, n. 9, p. e1403, 2014.
- YOSHIMURA, T. et al. Gtsf1/Cue110, a gene encoding a protein with two copies of a CHHC Zn-finger motif, is involved in spermatogenesis and retrotransposon suppression in murine testes. **Developmental Biology**, v. 335, n. 1, p. 216–227, 2009.
- YU, F. Methyl-CpG-binding protein 2 represses LINE-1 expression and retrotransposition but not Alu transcription. **Nucleic Acids Research**, v. 29, n. 21, p. 4493–4501, 2001.
- YU, Q. et al. Type I interferon controls propagation of long interspersed element-1. **Journal of Biological Chemistry**, v. 290, n. 16, p. 10191–10199, 2015.
- ZAFFAGNINI, G.; MARTENS, S. Mechanisms of Selective Autophagy. **Journal of Molecular Biology**, v. 428, n. 9, p. 1714–1724, 2016.
- ZHANG, A. et al. RNase L restricts the mobility of engineered retrotransposons in cultured human cells. **Nucleic Acids Research**, v. 42, n. 6, p. 3803–3820, 2014.

- ZHANG, F. et al. Systematic association analysis of microRNA machinery genes with schizophrenia informs further study. **Neuroscience Letters**, v. 520, n. 1, p. 47–50, 2012.
- ZHANG, Y. et al. SEPA-1 Mediates the Specific Recognition and Degradation of P Granule Components by Autophagy in *C. elegans*. **Cell**, v. 136, n. 2, p. 308–321, 2009.
- ZHAO, B. et al. Somatic LINE-1 retrotransposition in cortical neurons and non-brain tissues of Rett patients and healthy individuals. **PLoS Genetics**, v. 15, n. 4, p. e1008043, 2019.
- ZHAO, K. et al. Modulation of LINE-1 and Alu/SVA Retrotransposition by Aicardi-Goutières Syndrome-Related SAMHD1. **Cell Reports**, v. 4, n. 6, p. 1108–1115, 2013.
- ZHAO, Y.; TIAN, E.; ZHANG, H. Selective autophagic degradation of maternally-loaded germline P granule components in somatic cells during *C. elegans* embryogenesis. **Autophagy**, v. 5, n. 5, p. 717–719, 2009.
- ZHU, C. et al. Structural insight into dGTP-dependent activation of tetrameric SAMHD1 deoxynucleoside triphosphate triphosphohydrolase. **Nature Communications**, v. 4, p. 2722, 2013.

APÊNDICE A - REVISÃO SISTEMÁTICA DA LITERATURA PARA LEVANTAMENTO DE GENES PARTICIPANTES DO PROCESSO DE REGULAÇÃO DE RETROELEMENTOS L1

Uma revisão sistemática da literatura (RSL), baseada no protocolo desenvolvido por Kitchenham e Charters (2007), foi empregada para identificar o conhecimento atual a respeito de reguladores de retroelementos L1. Buscas para identificar artigos com esta informação foram realizadas nas bases de dados PubMed e Scopus até 15 de agosto de 2020.

A fim de incluir todos os estudos existentes a respeito do tema considerado, os seguintes termos e seus respectivos sinônimos foram empregados: combinação de palavras-chave para L1, combinação de palavras-chave para denotar controle/restrrição/ativação e palavras-chave para vias/mecanismos envolvidos na regulação de L1, resultando em uma combinação de palavras-chaves para a construção da busca (Tabela 4). Os resultados obtidos na busca eletrônica foram extraídos diretamente para uma planilha do Excel e tabuladas com as seguintes informações: Autor, título, ano, revista, volume, edição, DOI, Link e base de dados. Em seguida foi realizada uma triagem para remover artigos duplicados para então realizar a seleção dos artigos relevantes para a revisão sistemática de literatura.

Após a remoção dos trabalhos duplicados, cada título e resumo foram revisados por dois autores. Estes trabalhos pré-selecionados foram então analisados pelo seu conteúdo completo a fim de identificar somente os estudos que atendem os critérios de inclusão.

Foram definidos como critérios de inclusão: i) fatores moleculares que regulem a expressão/atividade de retroelementos pertencentes a família de retroelementos L1; ii) fatores moleculares avaliados em humanos e demais espécies animais; iii) estudos publicados até 15 de agosto de 2020.

Com base nos critérios definidos para esta revisão, foram extraídos 7130 artigos a partir da base de busca PubMed e 5670 artigos da base Scopus. Aplicando-se os critérios de inclusão e exclusão, 122 trabalhos foram utilizados para coleta das informações (Figura 42).

Tabela 4: Parâmetros de busca utilizados para levantamento de moléculas reguladoras de retroelementos L1.

Base de Dados	Parâmetros de Busca
PubMed	("L1* element"[TIAB] OR "L1* retrotransposition"[TIAB] OR "L1* transposon"[TIAB] OR "LINE 1"[TIAB] OR "LINE-1"[TIAB] OR "LINE1"[TIAB] OR "Long interspersed nuclear element"[TIAB] OR "endogenous retroelement"[TIAB] OR retrotransposition[TIAB] OR retrotransposon*[TIAB]) AND (accumulate*[TIAB] OR activate[TIAB] OR activation[TIAB] OR activity[TIAB] OR clearing[TIAB] OR control[TIAB] OR damage*[TIAB] OR decrease*[TIAB] OR defense[TIAB] OR degrade[TIAB] OR degrading[TIAB] OR derepression[TIAB] OR facilitate[TIAB] OR frustrate[TIAB] OR increase*[TIAB] OR inhibit[TIAB] OR interact*[TIAB] OR limit[TIAB] OR misregulation[TIAB] OR modulate*[TIAB] OR modulation[TIAB] OR regulation[TIAB] OR regulatory[TIAB] OR relax*[TIAB] OR replication[TIAB] OR repress*[TIAB] OR restrict[TIAB] OR restriction[TIAB] OR retards[TIAB] OR silence[TIAB] OR silencing[TIAB] OR start[TIAB] OR stimulate*[TIAB] OR trigger[TIAB])
Scopus	TITLE-ABS-KEY ("L1 element" OR "L1 retrotransposition" OR "L1 transposon" OR "LINE 1" OR "LINE-1" OR "LINE1" OR "Long interspersed nuclear element" OR "Endogenous retroelement" OR retrotransposition OR retrotransposon) AND TITLE-ABS-KEY (accumulate OR activate OR activation OR activity OR clearing OR control OR damage OR decrease OR defense OR degrade OR degrading OR derepression OR facilitate OR frustrate OR increase OR inhibit OR interact OR limit OR misregulation OR modulate OR modulation OR regulation OR regulatory OR relax OR replication OR repress OR restrict OR restriction OR retards OR silence OR silencing OR start OR stimulate OR trigger) AND (LIMIT-TO (SUBJAREA , "BIOC") OR LIMIT-TO (SUBJAREA , "MEDI") OR LIMIT-TO (SUBJAREA , "IMMU") OR LIMIT-TO (SUBJAREA , "NEUR") OR LIMIT-TO (SUBJAREA , "PHAR") OR LIMIT-TO (SUBJAREA , "CHEM"))

As seguintes informações foram coletadas dos artigos selecionados: nome do gene/proteína, espécie observada, tipo celular, tipo de atividade, família L1 e domínios da proteína.

Foram encontrados 158 genes reguladores de L1 atuando a nível pré-transcricional e pós-transcricional. Estes encontram-se dispostos na Tabela 5 e são explorados a seguir.

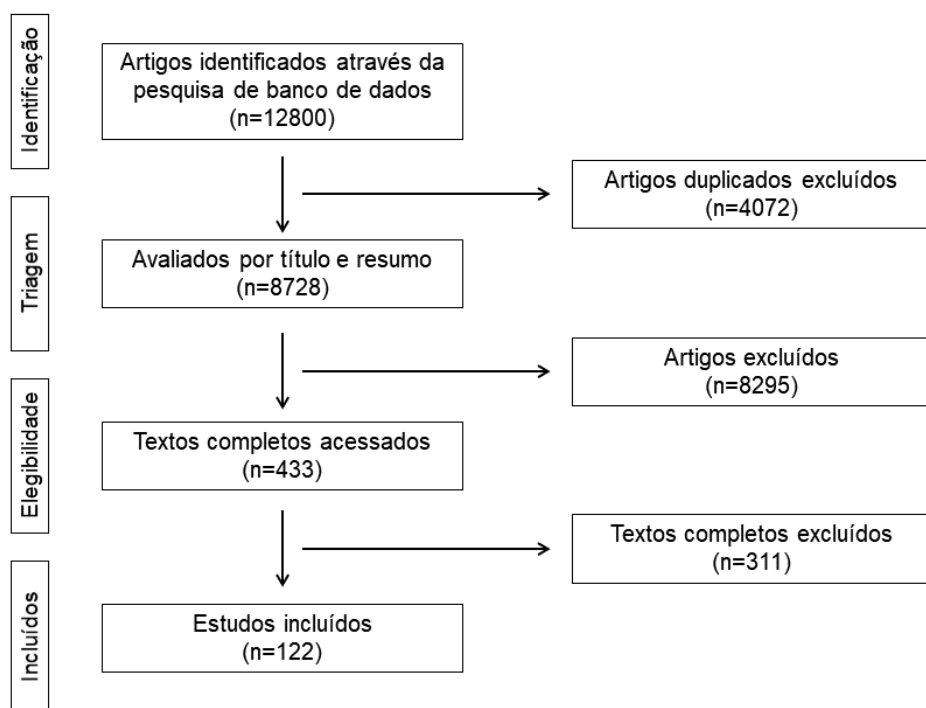


Figura 42: Fluxograma de identificação e seleção de artigos para a RSL.

Reguladores Pré-transcricionais de L1s

Regulação de L1 mediada por metilação do DNA

A metilação do DNA é uma modificação epigenética importante que regula vários fenômenos biológicos, incluindo impressão genômica e silenciamento de retroelementos (HATANAKA et al., 2015; KATO et al., 2007). O processo de metilação do DNA é mediado por DNA metiltransferases (DNMTs) (GOODIER, 2016).

Estudos mostram que os elementos L1 são metilados pelas proteínas DNMT1 (*DNA Methyltransferase 1*), DNMT3A (*DNA Methyltransferase 3 Alpha*), DNMT3B (*DNA Methyltransferase 3 Beta*) e DNMT3L (*DNA Methyltransferase 3 Like*), as quais atuam cooperativamente na supressão de L1 (KATO et al., 2007; LI et al., 2015; LIANG et al., 2002). Dnmt3c (*DNA methyltransferase 3C*) encontrada nas células germinativas masculinas de camundongos, é responsável pela metilação dos promotores de retroelementos e assim promovendo seu silenciamento (BARAU et al., 2016).

Além das DNMTs, MORC1 (*MORC Family CW-Type Zinc Finger 1*) também pode facilitar a metilação do DNA e repressão de L1 na linha germinativa masculina de camundongos (GOODIER, 2016; PASTOR et al., 2014).

A proteína SIRT6 (*Sirtuin 6*) também reprime L1 ligando-se a região 5'UTR de L1 e promovendo a heterocromatização através da ribosilação de TRIM28 (*Tripartite Motif Containing 28*) (VAN METER et al., 2014). TRIM28 é um co-repressor transcricional essencial para o desenvolvimento normal e diferenciação celular (GOODIER, 2016) que atua ligando-se a proteínas de ligação ao DNA (IYENGAR; FARNHAM, 2011). Foi observado que ele atua controlando retroelementos durante o desenvolvimento embrionário inicial (ROWE et al., 2010).

O gene *ALKBH1* (*AlkB Homolog 1, Histone H2A Dioxygenase*) codifica uma desmetilase para N6-metiladenina (WU et al., 2016). A ausência de *Alkbh1* leva ao aumento da deposição de N6-mA nas extremidades 5' de retroelementos L1, acarretando no silenciamento epigenético de tais retroelementos em células-tronco embrionárias de camundongos. (WU et al., 2016).

MECP2 (*Methyl-CpG Binding Protein 2*) é uma proteína nuclear envolvida na metilação global do DNA e doenças do neurodesenvolvimento humano (MUOTRI et al., 2010; SKENE et al., 2010). Estudos demonstram que a deficiência da proteína MECP2, resulta em transcrição elevada de elementos repetitivos (MUOTRI et al., 2010; SKENE et al., 2010; YU, 2001). Por outro lado, quando presente, MECP2 liga-se à região 5'UTR de L1 limitando sua expressão e retrotransposição (YU, 2001).

ZBTB16 (*Zinc Finger And BTB Domain Containing 16*) é um regulador do crescimento e diferenciação celular que se liga ao DNA de L1 (GOODIER, 2016). Esta ligação altera a acetilação e metilação da cromatina local e acaba reprimindo a expressão de L1 em células germinativas e progenitoras (PUSZYK et al., 2013).

Outra proteína, TARDBP (*TAR DNA Binding Protein*) é uma proteína de ligação a RNA (BHARDWAJ et al., 2013). A perda de TARDBP nuclear está associada à descondensação da cromatina em torno de retroelementos e aumento a expressão de L1 (LIU et al., 2019). A perda de TARDBP também leva ao aumento da retrotransposição (LIU et al., 2019).

A proteína SIRT7 (*Sirtuin 7*), participa da repressão transcricional epigenética de todo o genoma L1 em células de camundongo e humanas (VAZQUEZ et al., 2019). Essa repressão a nível pré-transcricional do retroelemento L1 ocorre via desacetilação H3K18Ac mediada por SIRT7 e promove a associação de L1 com a lâmina nuclear.

A histona desmetilase KDM4B é frequentemente superexpressa em vários tipos de câncer e possui capacidade de desmetilar H3K9me3 (TUMBER et al., 2017). Um estudo mostrou que em células cancerosas, uma proporção significativa de H3K9me3 dependente de KDM4B, foi localizada em retroelementos L1 de forma a reprimir a atividade de retrotransposição (XIANG et al., 2019). Além disso observou-se que a superexpressão de KDM4B aumentou a eficácia de retrotransposição de L1. Ainda, a superexpressão de KDM4B aumentou a eficácia de retrotransposição de L1 e danos ao DNA (XIANG et al., 2019).

As proteínas MRE11A (*MRE11 Homolog, Double Strand Break Repair Nuclease*) e RAD50 (*RAD50 Double Strand Break Repair Protein*), componentes do complexo MRN (o qual desempenha papel importante no reparo de quebra de fita dupla) atuam regulando positivamente a atividade de L1 (LIU et al., 2018; MITA et al., 2020).

Ainda, a proteína UHRF1 (*Ubiquitin Like With PHD And Ring Finger Domains 1*) é um regulador chave entre a manutenção da metilação do DNA e nas modificações das histonas (BOSTICK et al., 2007; HASHIMOTO et al., 2009). Esta proteína de múltiplos domínios, é responsável pela repressão de L1s participando através da metilação do DNA, modificação de histona e via piRNA. UHRF1 interage com PRMT5 (*Protein Arginine Methyltransferase 5*) para regular as modificações repressivas da histona (H4R3me2s e H3R2me2s) e coopera com a via PIWI durante a espermatogênese (DONG et al., 2019).

Regulação de L1 mediada por modificações de histonas

Proteínas associadas a cromatina também são responsáveis pela regulação de retroelementos. CHAF1 (*Chromatin assembly factor 1*) é uma chaperona capaz de reunir histonas H3 e H4 em DNA recém-sintetizado, durante a replicação e reparo do DNA (CLÉMOT et al., 2018). Estudos mostram que embriões de camundongos são protegidos de retroelementos por modificações repressivas de histonas mediadas por CHAF1

(HATANAKA et al., 2015). Tais modificações incluem a substituição de H3.3 por H3.1/3.2 e também a deposição de marcas repressivas de histona (como H4K20me3 e H3K9me3), em regiões contendo retroelementos L1 (HATANAKA et al., 2015). Ainda, a marca epigenética H3K9me2 mediado por EHMT2 (*Euchromatic Histone Lysine Methyltransferase 2*) também é um importante mecanismo para o silenciamento de L1 (DI GIACOMO et al., 2014). Ainda, foi observado que o acúmulo de H3K9me3 mediado pelas metiltransferases SUV39H1 e SUV39H2 em células-tronco embrionárias de camundongos leva à repressão transcricional de L1s (BULUT-KARSLIOGLU et al., 2014).

Também, a ausência de KDM1A (*Lysine Demethylase 1A*) resulta em níveis elevados das marcas H3K4me3 e H3K9me3 e também no aumento da atividade do L1 e dos níveis de proteína L1 ORF1 (ANCELIN et al., 2016).

Outra proteína observada na regulação de L1, é a SETDB1 (*SET Domain Bifurcated 1*). A ausência desta proteína em células humanas de leucemia mielóide aguda (AML) levou a uma redução de H3K9me3 e limitou a expressão de retroelementos L1 (CUELLAR et al., 2017).

Nas células humanas, existe um complexo multiproteico, chamado de complexo HUSH, que atua na repressão epigenética (TCHASOVNIKAROVA et al., 2015). A proteína MORC2 (*MORC Family CW-Type Zinc Finger 2*) junto com as proteínas MPP8 (*M-Phase Phosphoprotein 8*) e TASOR (*Transcription Activation Suppressor*), que são subunidades do complexo HUSH, atuam restringindo a atividade de L1z (LIU et al., 2018). Estudos indicam que o complexo HUSH e MORC2 ligam-se a L1s e promovem a deposição de marcas epigenéticas repressivas, como H3K9me3 (LIU et al., 2018). Tais eventos de silenciamento geralmente ocorrem dentro de íntrons de genes transcricionalmente ativos e levam à regulação negativa da expressão do gene hospedeiro de maneira dependente de HUSH, MORC2 e L1 (LIU et al., 2018).

Regulação de L1 mediada por proteínas de reparo de DNA

Algumas proteínas de reparo de DNA também desempenham papel regulatório em retroelementos L1. Estudos tem demonstrado resultados diferentes para o papel da

proteína ATM (*ATM Serine/Threonine Kinase*) na retrotransposição de L1 (COUFAL et al., 2011; GASIOR et al., 2006; WALLACE et al., 2012).

Em células humanas com deficiência de ATM (ESC, NPC derivados de ESC, NSC e células HCT116), Coufal e colaboradores observaram um aumento na eficiência da retrotransposição de L1s (COUFAL et al., 2011). O mesmo resultado foi observado em camundongos transgênicos *knockout* para ATM. Neste mesmo trabalho é relatado um aumento no número de cópias de DNA L1 em amostras de cérebro de pacientes com Ataxia-telangiectasia (AT) (COUFAL et al., 2011), uma doença causada por mutações autossômicas recessivas que inativam ATM, fazendo com que os pacientes com AT apresentem neurodegeneração progressiva (SHILOH, 2001). Por outro lado, estudos com linhagens celulares com níveis reduzidos de proteína ATM, levaram a diminuição da atividade de L1 (GASIOR et al., 2006; WALLACE et al., 2012).

A via de reparo por excisão de nucleotídeos (NER) é a principal via para remover lesões volumosas de DNA (SCHÄRER, 2013). Foi observado que três proteínas essenciais da via NER limitam a retrotransposição de L1: XPA (*XPA, DNA Damage Recognition And Repair Factor*), ERCC2 (*ERCC Excision Repair 2, TFIIH Core Complex Helicase Subunit*) e XPC (*XPC Complex Subunit, DNA Damage Recognition And Repair Factor*) (AOTO et al., 2019; SERVANT et al., 2017).

A via de reparo de danos ao DNA da anemia de Fanconi (FA) possui proteínas que colaboram em processos envolvendo a restauração da integridade do genoma, reparando o DNA danificado (ESTEBAN-JURADO et al., 2016), atuando na replicação do DNA e controle do ciclo celular (HOWLETT et al., 2005; NAIM; ROSSELLI, 2009; NAKANISHI et al., 2005). FANCD2 (*FA Complementation Group D2*) é um membro chave da via FA (NIE et al., 2020). FANCD2 interage com componentes importantes das vias Prmt5/piRNA específicas para células germinativas que atuam na repressão de retroelementos (NIE et al., 2020). PMRT5 (*Protein Arginine Methyltransferase 5*), uma arginina metiltransferase, é necessária para a modificação repressiva da cromatina H2A/H4R3me2 para retroelementos L1, levando a seu silenciamento durante a hipometilação do DNA (KIM et al., 2014). Estudos mostram que Fancd2 e Prmt5 co-ocupam o promotor de L1 (NIE et al., 2020), sugerindo que a via FA também participa da repressão de retroelementos L1, provavelmente por meio de um mecanismo envolvendo

marcas repressivas H2A/H4R3me2 catalisadas por Fancd2 e Prmt5 (NIE et al., 2020). Além disso, FANCD2 também é importante para ativação de SLX4 (*SLX4 Structure-Specific Endonuclease Subunit*) (BRÉGNARD et al., 2016), e ambos atuam também reprimindo diretamente a retrotransposição de L1 e evitando o acúmulo de ácidos nucleicos citoplasmáticos. Outro componente da via FA (MEETEI et al., 2005) e que também participa da repressão de L1 é FANCM (*FA Complementation Group M*) (LIU et al., 2018; MITA et al., 2020).

FANCB (*FA Complementation Group B*), que atua na ubiquitinação de FANCD2 (VAN TWEST et al., 2017), também reprime L1 (LIU et al., 2018; MITA et al., 2020). Por outro lado, FANCA (*FA Complementation Group A*) leva a uma regulação positiva de retroelementos L1 (LIU et al., 2018; MITA et al., 2020).

Outras duas proteínas, PALB2 (*Partner And Localizer Of BRCA2*) e XRCC2 (*X-Ray Repair Cross Complementing 2*) desempenham papel crítico na via de reparo DNA por recombinação homóloga (HRR) (BUISSON; MASSON, 2012; CAPPELLI et al., 2011) e são capazes de regular de forma distinta retroelementos L1. Enquanto PALB2 regula negativamente a atividade de L1, XRCC2 ativa a retrotransposição de L1 (LIU et al., 2018; MITA et al., 2020). Ainda, XRCC6 (*X-Ray Repair Cross Complementing 6*), desempenha papel importante na translocação de cromossomos e regula negativamente a atividade de L1s (LIU et al., 2018; MITA et al., 2020; SUZUKI et al., 2009).

As proteínas NCAPH2 (*Non-SMC Condensin II Complex Subunit H2*) e NCAPD3 (*Non-SMC Condensin II Complex Subunit D3*) são subunidades reguladoras do complexo Condensina II, um complexo que estabelece a arquitetura cromossômica mitótica e está envolvida na rigidez física do eixo da cromátide (ONO et al., 2003). NCAPH2 e NCAPD3 interagem com membros do complexo Inibidor de Tradução Ativado por Interferon Gama (GAIT), e se associam ao RNA de L1 o que resulta na repressão da atividade de L1 (WARD et al., 2017).

PARP1 (*Poly(ADP-Ribose) Polymerase 1*) e PARP2 (*Poly(ADP-Ribose) Polymerase 2*) promovem a poli-ADP-ribosilação de proteínas e também são importantes no reparo do DNA (VYAS et al., 2014). Ambas atuam regulando de forma positiva retroelementos L1 (MIYOSHI; MAKINO; MORAN, 2019; TAYLOR et al., 2013).

Regulação de L1 mediada por fatores de transcrição

A região 5'UTR do L1 abriga locais de ligação para alguns fatores de transcrição: RUNX3 (*Runt Related Transcription Factor 3*), SOX11 (*SRY-Box 11*), SOX2 (*SRY-Box 2*), SP1 (*Sp1 Transcription Factor*) (RICHARDSON et al., 2015) e YY1 (*YY1 Transcription Factor*) (BECKER et al., 1993).

RUNX3 tem papel importante para aumentar a transcrição e retrotransposição de L1. Mutações no primeiro sítio de ligação de RUNX na região 5'UTR de L1HS, reduziram significativamente a interação, transcrição e retrotransposição de L1 (YANG et al., 2003). SOX11 regula positivamente a atividade do retroelemento L1 durante a diferenciação neuronal ligando-se a região promotora de L1 (ORQUEDA et al., 2018; TCHÉNIO; CASELLA; HEIDMANN, 2000). A redução na expressão de SOX2 durante os estágios iniciais da diferenciação neuronal leva a aumentos na transcrição e retrotransposição de L1 (MUOTRI et al., 2005). O sítio de ligação a YY1 funciona como um componente essencial para direcionar o início preciso da transcrição do promotor de L1s (ATHANIKAR; BADGE; MORAN, 2004; LEE et al., 2010).

SP1 modula a expressão de elementos transponíveis (GOODIER, 2016). Um estudo *in vitro*, mostrou que a proteína SP1 se liga especificamente ao promotor L1 de camundongos (FEDOROV; LUKYANOV; PODGORNAYA, 2006).

Outros fatores de transcrição que também regulam L1 incluem: TP53 (*Tumor protein p53*) (HARRIS et al., 2009), ZNF91 (*Zinc Finger Protein 91*), ZNF93 (*Zinc Finger Protein 93*) (CASTRO-DIAZ; FRIEDLI; TRONO, 2015; JACOBS et al., 2014), E2F1 (*E2F Transcription Factor 1*) e E2F4 (*E2F Transcription Factor 4*) (MONTROYA-DURANGO et al., 2009, 2016) que atuam restringindo a expressão L1, além dos fatores ETS1 (*ETS Proto-Oncogene 1, Transcription Factor*) (WANG et al., 2015) e RAD21 (*RAD21 Cohesin Complex Component*) (XU et al., 2014) que atuam de forma a aumentar a expressão de L1s e por fim temos AHR (*Aryl Hydrocarbon Receptor*) que pode regular tanto de forma negativa como positiva a atividade de L1 (MIRET et al., 2020; MULERO-NAVARRO; FERNANDEZ-SALGUERO, 2016; ROMAN et al., 2008; TENENG; STRIBINSKIS; RAMOS, 2007).

Regulação de L1 mediada por outras proteínas celulares

O complexo de ribonuclease HII (RNase HII) é composto de três subunidades RNASEH2A (*Ribonuclease H2 Subunit A*), RNASEH2B (*Ribonuclease H2 Subunit B*) e RNASEH2C (*Ribonuclease H2 Subunit C*) e possui função crítica no metabolismo de ácido nucleicos (SHABAN et al., 2010). Estudos mostram que RNase HII é necessário para uma retrotransposição eficiente de L1 (BENITEZ-GUIJARRO et al., 2018). Isso ocorre através da degradação do RNA de L1 após a transcrição reversa, via RNase HII, permitindo que a retrotransposição seja concluída.

CDKN1A (*Cyclin Dependent Kinase Inhibitor 1A*) desempenha papel importante no controle da progressão do ciclo celular (LABAER et al., 1997). Estudos indicam que CDKN1A coimmunoprecipita com ORF2p e suprime a atividade da transcriptase reversa de L1, sugerindo que CDKN1A inibe a transcrição reversa mediada por ORF2p, limitando a mobilidade de L1 (KAWANO et al., 2018).

Reguladores Pós-transcricionais de L1s

Regulação de L1 por proteínas da família APOBEC

A família APOBEC (*Apolipoprotein B mRNA Editing Catalytic Polypeptide-like*) consiste em 11 produtos genéticos primários e variantes alternadas que incluem as proteínas APOBEC1 (*Apolipoprotein B MRNA Editing Enzyme Catalytic Subunit 1*), APOBEC2 (*Apolipoprotein B MRNA Editing Enzyme Catalytic Subunit 2*), APOBEC3A-H (*Apolipoprotein B MRNA Editing Enzyme Catalytic Subunit 3A-H*), APOBEC4 (*Apolipoprotein B MRNA Editing Enzyme Catalytic Polypeptide Like 4*) e AICDA (*Activation Induced Cytidine Deaminase*) (SALTER; BENNETT; SMITH, 2016). Possuem atividade de deaminases, capazes de modificar as bases da citosina em uracilas (C – U) no DNA e/ou RNA (IKEDA et al., 2011).

Entre as proteínas APOBEC, três membros desta família foram caracterizadas como inibidoras de L1: APOBEC1 (IKEDA et al., 2011), todas as proteínas APOBEC3 (A3A-H) (KINOMOTO et al., 2007) e AICDA (MACDUFF; DEMOREST; HARRIS, 2009; METZNER; JÄCK; WABL, 2012).

As proteínas da família APOBEC1, em mamíferos, reduzem a mobilidade dos retroelementos L1 por mecanismo independente da desaminação (IKEDA et al., 2011).

As proteínas da família APOBEC3 são essenciais para o desenvolvimento, sobrevivência ou fertilidade de camundongos (MIKL et al., 2005). Estas proteínas inibem o nível de retrotransposição de L1 (KINOMOTO et al., 2007; LINDIČ et al., 2013; MUCKENFUSS et al., 2006). Uma vez que os membros da família APOBEC3 têm funções distintas, elas parecem ter evoluído para resistir a várias classes de elementos genéticos (CHEN et al., 2006). Entre os membros da família APOBEC3, APOBEC3A é o membro mais ativo na inibição dos elementos L1 (CHEN et al., 2006).

AICDA é uma enzima de edição de DNA, que desempenha papéis importantes na diversificação de genes de anticorpos em células B ativadas (IKEDA et al., 2011; MURAMATSU et al., 1999). Estudos tem demonstrado que AICDA pode inibir a retrotransposição de L1 através de um mecanismo independente da desaminação do DNA, que pode se manifestar no compartimento citoplasmático (MACDUFF; DEMOREST; HARRIS, 2009; METZNER; JÄCK; WABL, 2012). Metzner e colaboradores (2012) mostraram, *in vitro*, que o AICDA forma complexos citoplasmáticos com o mRNA de L1, o que pode contribuir para a restrição destes retroelementos.

Regulação de L1 por autofagia

A autofagia desempenha um papel no metabolismo dos transcritos de L1, impedindo novas inserções de retroelementos no genoma (GUO et al., 2014).

Os RNAs são direcionados para a degradação autofágica através de grânulos de RNA (como os grânulos de estresse) (BUCHAN et al., 2013). Entre os receptores de autofagia, CALCOCO2 (*Calcium Binding And Coiled-Coil Domain 2*) e SQSTM1 (*Sequestosome 1*), são responsáveis pela degradação seletiva de grânulos de RNA, os quais carregam RNA de L1 (GUO et al., 2014).

Em células de camundongos, a ausência de ATG5 (Autophagy Related 5) e BECN1 (*Beclin 1*), leva ao acúmulo de RNA de L1 (GUO et al., 2014; KEMP; LONGWORTH, 2015). ATG5 e BECN1 são genes críticos para a formação de autofagossomos. Estes dados mostram que o impedimento deste processo de

degradação do RNA de L1 por autofagia, resulta no acúmulo de RNA do retroelemento e a novas inserções genômicas (GUO et al., 2014).

Regulação de L1 por vias de resposta a interferons

A resposta imune inata envolve fatores celulares com atividades antivirais, entre as quais a via de resposta ao interferon (IFN) desempenha um papel central. Essas vias levam à ativação de genes estimulados por IFN e alguns destes genes também são responsáveis pela restrição de retroelementos L1 (PIZARRO; CRISTOFARI, 2016).

MOV10 (*Mov10 RISC Complex RNA Helicase*) e RNASEL (*Ribonuclease L*) clivam regiões de fita simples de RNA viral e celular, suprimindo a síntese, replicação e disseminação de proteínas virais (SILVERMAN, 2007). Ambas atuam na repressão de retroelementos L1 (ARJAN-ODEDRA et al., 2012; GOODIER et al., 2012; ZHANG et al., 2014). O *knockdown* de MOV10 resulta no aumento dos níveis de retrotransposição de L1, possivelmente devido ao mecanismo de silenciamento de RNAi (GOODIER et al., 2012). Já o *knockdown* da RNASEL aumenta em ~87% a retrotransposição de L1, indicando que RNASEL interfere negativamente no ciclo de retrotransposição L1 e pode contribuir para a manutenção da estabilidade genômica (ZHANG et al., 2014).

SAMHD1 (*SAM And HD Domain Containing Deoxynucleoside Triphosphate Triphosphohydrolase 1*) e TREX1 (*Three Prime Repair Exonuclease 1*) também atuam em vias de resposta a interferons (RIESS et al., 2017; YAN et al., 2010). SAMHD1 é uma proteína que possui atividade de dNTPase e que inibe retrovírus, vírus de DNA e retroelementos L1 (HU et al., 2015). Estudos preveem que a restrição de L1s por SAMHD1 decorre de um mecanismo diferente da atividade do dNTPase [Revisado em: (GOODIER, 2016)]. A inibição da L1 por SAMHD1 é observada na divisão celular (ZHAO et al., 2013). Ainda, SAMHD1 tem a capacidade de reduzir a transcrição reversa de L1 através da inibição de ORF2p (mas não ORF1p) indicando que L1 ORF2p pode ser um alvo potencial de SAMHD1 (ZHAO et al., 2013). Além disso, SAMHD1 tem a capacidade de induzir a formação de grandes grânulos de estresse e, assim, aumentar o sequestro de L1 RNP dentro desses pontos citoplasmáticos levando ao bloqueio da retrotransposição de L1 (HU et al., 2015).

A superexpressão de TREX1 reduz a retrotransposição dos elementos L1 (STETSON et al., 2008). Além disso, o DNA de fita simples derivado de retroelementos endógenos, como L1s, SINEs e ERVs, se acumula nas células com deficiência de Trex1 de ratos (GOODIER, 2016; STETSON et al., 2008).

A proteína antiviral TRIM5 (*Tripartite Motif Containing 5*), além de ser importante na sinalização imunidade inata, desempenha importante papel na proteção do genoma contra retroelementos através de sua restrição (VOLKMANN et al., 2020). TRIM5 interage com complexos de ribonucleoproteína de L1 no citoplasma, induzindo a sinalização imune inata. Os eventos de sinalização ativam os fatores de transcrição AP-1 e NF- κ B, levando à regulação negativa da atividade do promotor L1 (VOLKMANN et al., 2020).

Outras proteínas antivirais estimuladas por interferon, como BST2 (*Bone Marrow Stromal Cell Antigen 2*), ISG20 (*Interferon Stimulated Exonuclease Gene 20*), MAVS (*Mitochondrial Antiviral Signaling Protein*), MX2 (*MX Dynamin Like GTPase 2*) e ZC3HAV1 (*Zinc Finger CCCH-Type Containing, Antiviral 1*), também apresentam efeitos de inibição de L1 quando superexpressadas (GOODIER et al., 2015). Estudos sugerem que ZC3HAV1 inibe a retrotransposição de L1 através da sua interação com o RNA L1, que ocorre em agregados citoplasmáticos que se co-localizam com proteínas de grânulos de estresse (GOODIER et al., 2015; MOLDOVAN; MORAN, 2015).

Regulação pelas vias de RNA de interferência (RNAi)

A interferência de RNAs (RNAi) é outro mecanismo que inibe os retroelementos L1 (BUCHON; VAURY, 2006; YANG; KAZAZIAN, 2006).

DGCR8 (*DGCR8, Microprocessor Complex Subunit*) e DROSHA (*Drosha ribonuclease type III*) formam um complexo conhecido como Microprocessador, que catalisa a etapa nuclear da biogênese de miRNA (HERAS et al., 2013). O DGCR8 reconhece o substrato do RNA, enquanto o Drosha atua como endonuclease (MACIAS et al., 2012). Esse Microprocessador (Drosha-DGCR8) é capaz de reconhecer e se ligar a RNAs derivados de retroelementos L1 humanos.

DGCR8 é uma proteína de ligação a RNA de fita dupla. Os alvos de DGCR8 incluem centenas de mRNAs, snRNAs (pequenos RNAs nucleares), longos RNAs não codificantes e também elementos repetitivos, principalmente retroelementos L1 e LTR (HERAS et al., 2013; MACIAS et al., 2012). Possivelmente DGCR8 se liga à RNAs derivados de L1 em estruturas em *hairpin* afetando a retrotransposição (GOODIER, 2016).

DROSHA é a subunidade catalítica do complexo Microprocessador, que inicia a maturação miRNA no núcleo, reconhecendo e clivando precursores de *hairpin* embutidos nos transcritos primários (LEE; SHIN, 2018). DROSHA regula negativamente a retrotransposição de L1 e Alu, clivando estruturas em *hairpin* pri-miRNA incorporadas em seus transcritos cognatos (LEE; SHIN, 2018).

O promotor humano de L1 5'-UTR pode levar à produção de dsRNAs, os quais são processados por DICER1 (*Dicer 1, Ribonuclease III*) para produzir uma série de siRNAs específicos de L1 (SPEEK, 2001; YANG; KAZAZIAN, 2006). O *knockdown* do gene *Dicer1* (*Dicer 1, Ribonuclease III*) em células-tronco embrionárias de camundongos aumenta a taxa de atividade de retrotransposição de L1 (CHEN et al., 2012; YANG; KAZAZIAN, 2006).

Os miRNAs também parecem ser capazes de mediar a estabilidade genômica, suprimindo a mobilidade dos retroelementos (HAMDORF et al., 2015). Um estudo feito com o miR128-1 (MicroRNA 128-1) (HAMDORF et al., 2015), mostra que este é capaz de reprimir novos eventos de retrotransposição em células cancerígenas humanas e iPSCs, através da sua ligação direta ao RNA de L1.

Os RNAs que interagem com PIWI (piRNAs), são pequenos RNAs não codificantes que restringem a mobilidade de vários retroelementos, e assim mantêm a integridade do genoma (RAJAN; RAMASAMY, 2014). Estudos tem demonstrado a importância de algumas proteínas para o silenciamento de retroelementos mediado por piRNA em células germinativas de camundongos. A ausência das seguintes proteínas levam ao aumento da atividade de retroelementos L1: ASZ1 (*Ankyrin Repeat, SAM And Basic Leucine Zipper Domain Containing 1*) (MA et al., 2009), EXD1 (*Exonuclease 3'-5' Domain Containing 1*) (YANG et al., 2016), FKBP6 (*FK506 Binding Protein 6*) (XIOL et al., 2012), GTSF1 (*Gametocyte Specific Factor 1*) (YOSHIMURA et al., 2009), HENMT1

(*HEN Methyltransferase 1*) (LIM et al., 2015), HSP90AA1 (*Heat Shock Protein 90 Alpha Family Class A Member 1*) (ICHIYANAGI et al., 2014), MAEL (*Maelstrom Spermatogenic Transposon Silencer*) (SOPER et al., 2008), PIWIL1 (*Piwi Like RNA-Mediated Gene Silencing 1*) (REUTER et al., 2011), PIWIL2 (*Piwi Like RNA-Mediated Gene Silencing 2*) (DI GIACOMO et al., 2013), PIWIL4 (*Piwi Like RNA-Mediated Gene Silencing 4*) (REUTER et al., 2011), PLD6 (*Phospholipase D Family Member 6*) (WATANABE et al., 2011), TDRD1 (*Tudor Domain Containing 1*) (REUTER et al., 2009), TDRD5 (*Tudor Domain Containing 5*) (YABUTA et al., 2011), TDRD9 (*Tudor Domain Containing 9*) (SHOJI et al., 2009), TDRD12 (*Tudor Domain Containing 12*) (PANDEY et al., 2013), TDRKH (*Tudor And KH Domain Containing*) (SAXE et al., 2013) e MOV10L1 (*Mov10 RISC Complex RNA Helicase Like 1*) (FROST et al., 2010) GPAT2 (*Glycerol-3-Phosphate Acyltransferase 2, Mitochondrial*) (SHIROMOTO et al., 2019).

Regulação de L1 por proteínas de ligação a poli (A)

Retroelementos funcionais L1 apresentam uma sequência poli (A), localizada na extremidade 3' do RNA, que é um substrato candidato para proteínas de ligação a poli (A) (PABPs) (DAI et al., 2012). As proteínas PABPC1 (*Poly(A) Binding Protein Cytoplasmic 1*), PABPN1 (*Poly(A) Binding Protein Nuclear 1*) e PABPC4 (*Poly(A) Binding Protein Cytoplasmic 4*) parecem ter papel importante na retrotransposição positiva de L1 (DAI et al., 2012). Tais proteínas apresentam associação com o complexo da ribonucleoproteína L1, possivelmente por meio do mRNA de L1 (DAI et al., 2012). Um estudo *knockdown* de PABPC1 e PABPN1 (DAI et al., 2012) mostrou que a ausência destas proteínas leva a redução da atividade de L1 em 70 a 80%, sem afetar significativamente a quantidade de RNA L1 e os níveis de tradução.

PABPC1 é essencial para a formação de complexos de RNA-proteína L1 e pode desempenhar um papel na translocação de RNP L1 na célula hospedeira (DAI et al., 2012). A diminuição dos níveis da proteína PABPC1 leva a um defeito na formação de L1 RNP, enquanto o *knockdown* do inibidor de PABPC1, PAIP2, aumenta a retrotransposição de L1 em até 2 vezes (DAI et al., 2012). Além disso, alterações nos níveis de PABPC1 altera a localização subcelular de ORF1p (GOODIER, 2016).

Já o *knockdown* de PABPN1 não alterou o nível relativo de ORF1p no complexo RNP (DAI et al., 2012), e apesar desta proteína também apresentar associação com o RNA L1, seu papel específico na retrotransposição de L1 ainda não é claro (PIZARRO; CRISTOFARI, 2016) assim como o da proteína PABPC4.

Regulação de L1 por outras proteínas celulares

Outras proteínas que restringem L1 incluem as ribonucleoproteínas nucleares heterogêneas (HNRNPs) atuam regulando o *splicing* alternativo (KĘDZIERSKA; PIEKIEŁKO-WITKOWSKA, 2017). Quatro proteínas HNRNPs parecem interagir especificamente com a ORF2 em sítios internos de entrada do ribossomo (IRESs): R, Q e L e NCL (*Nucleolin*) (PEDDIGARI et al., 2013). Peddigari e colaboradores (2013) mostraram que o *knockdown* de HNRNPL (*Heterogeneous Nuclear Ribonucleoprotein L*) aumenta drasticamente a retrotransposição de L1, bem como o RNA de L1 e a proteína ORF1, indicando que regula negativamente a retrotransposição de L1. Por outro lado a proteína SYNCRIP (*Synaptotagmin Binding Cytoplasmic RNA Interacting Protein*) atua de forma positiva na atividade de retroelementos L1.

MARF1 (*Meiosis Regulator and mRNA Stability Factor 1*) é uma proteína de ligação ao RNA citoplasmático essencial para a progressão da meiose e controle da retrotransposição em oócitos (NISHIMURA et al., 2018; SU et al., 2012a). A ausência de MARF1 em oócitos resultam no acúmulo de mRNA de L1 (NISHIMURA et al., 2018).

TEX19 (*Testis Expressed 19*) é uma proteína específica de mamífero com expressão restrita em células-tronco pluripotentes e linhagem germinativa (KUNTZ et al., 2008). Em camundongos, TEX19 é importante para o desenvolvimento normal da placenta e espermatogênese (REICHMANN et al., 2013).

TUT4 (*Terminal Uridylyl Transferase 4*) e TUT7 (*Terminal Uridylyl Transferase 7*) restringem L1 agindo ao nível do mRNA de L1, Para isso é necessária a atividade da MOV10 seguida da uridilação das extremidades 3' dos mRNAs de L1 por TUT4 e TUT7 (WARKOCKI et al., 2018).

A proteína BRCA1 (*BRCA1 DNA Repair Associated*), liga-se ao mRNA de L1 no citoplasma levando à inibição da tradução de ORF2 e restringindo a retrotransposição de L1 (MITA et al., 2020).

Já a proteína PCNA (*Proliferating Cell Nuclear Antigen*) é responsável pela ativação de L1. PCNA está envolvida nas vias de reparo do DNA e sua ligação com o *motif* PIP-box (região de interação com PCNA, nos resíduos 407-415 de ORF2p, localizados entre o domínio EN e RT) em ORF2p é crítica para retrotransposição (TAYLOR et al., 2013). Outras proteínas que também são responsáveis pela regulação positiva da atividade de L1 são: STK17A (*Serine/Threonine Kinase 17a*), FCGR1B (*Fragment Of IgG Receptor Ib*) e TIMM8B (*Translocase Of Inner Mitochondrial Membrane 8 Homolog B*) (LIU et al., 2018; MITA et al., 2020)

A ausência de UPF1 (*UPF1, RNA Helicase And ATPase*), UPF2 (*UPF2 Regulator Of Nonsense Mediated MRNA Decay*) e UPF3A (*UPF3A Regulator Of Nonsense Mediated MRNA Decay*) reduzem a retrotransposição enquanto aumentam os níveis de mRNA de L1 (TAYLOR et al., 2013).

Existem dois principais receptores de melatonina em mamíferos, MT1 e MT2. MT1 (*Melatonin Receptor 1A*) é amplamente expresso no sistema nervoso central (DEHARO et al., 2014). Deharo e colaboradores (2014) mostraram a partir de um modelo isolado de tecido de câncer de próstata que a expressão de elementos L1 é suprimida pela melatonina que circula no sangue humano, sendo que a inibição de L1 ocorre através da regulação negativa do mRNA de L1 e da proteína ORF1 (DEHARO et al., 2014).

RB1 (*RB Transcriptional Corepressor 1*) junto com E2F, forma o complexo E2F-RB que participa da regulação epigenética dos elementos L1 por meio de modificações de histonas e recrutamento de histona desacetilases (HDACs) HDAC1 e HDAC2 (MONTROYA-DURANGO et al., 2009).

TROVE2 (*TROVE Domain Family Member 2*) possui especificidade apenas com ORF1p (FUCHS et al., 2006; TAYLOR et al., 2013).

ADAR (*Adenosine Deaminase RNA Specific*) restringe a retrotransposição de L1 através da atividade de edição do ADAR. Além disso, ADAR liga-se ao complexo L1 RNP, sustentando seu papel como regulador do ciclo de vida de L1 (ORECCHINI et al., 2017).

Tabela 5: Genes reguladores de retroelementos L1.

Gene: descreve os genes encontrados que de alguma forma regulam a atividade de L1; Descrição: nome do gene; Localização: fornece informações sobre localização celular de genes; Espécie: espécie onde foi observada a ação de reguladores L1; Tipo celular: tipo celular onde foi observado a ação de reguladores L1; Regulação: indica se a molécula em questão atua de forma negativa, restringindo a regulação de L1 (N) ou de forma positiva (P), favorecendo a expressão/atividade de L1; Nível: pré-transcricional (PRÉ) ou pós-transcricional (PÓS); Domínios proteicos: fornece informações anotadas sobre domínios e famílias de proteínas humanas (*Homo sapiens*) de acordo com UniProtKB; Função, fornece informações anotadas sobre a função da molécula; Referência: refere-se a lista individual de todos os trabalhos consultados.

(continua)

Gene	Descrição	Localização	Espécie	Tipo celular	Regulação	Nível	Domínios proteicos	Função	Referência
ADAR	<i>Adenosine Deaminase, RNA Specific</i>	Núcleo e citoplasma	Hs	HEK293T, HeLa	N	PÓS	Possui dois domínios <i>Z-binding</i> e três DRBM. DRBM 3 é importante para reunir a parte N-terminal e a parte C-terminal do sinal de localização nuclear para importação mediada por TNPO1. A ligação do RNA interfere na importação nuclear. O primeiro domínio de ligação Z liga-se ao Z-DNA.	Catalisa a desaminação hidrolítica de adenosina em inosina em RNA de fita dupla (dsRNA). Isso pode afetar a expressão e a função do gene de várias maneiras.	(ORECCHINI et al., 2017; ORECCHINI; FRASSINELLI; MICHIEZI, 2017)
AHR	<i>Aryl Hydrocarbon Receptor</i>	Núcleo e citoplasma	Hs, Mm	HeLa, HMEC, mVSMC, mK4, Hepa 1, MDA-MB-231, NMuMG	N/P	PRÉ	Possui os domínios bHLH, PAS1, PAS2 e PAC. O domínio PAS1 é essencial para a dimerização e também necessário para heterodimerização AHR: ARNT.	Fator de transcrição.	(MIRET et al., 2020; MULERO-NAVARRO; FERNANDEZ-SALGUERO, 2016; ROMAN et al., 2008; TENENG; STRIBINSKIS; RAMOS, 2007)
AICDA	<i>Activation Induced Cytidine Deaminase</i>	Núcleo	Hs, Mm	HeLa, HEK293, WEHI-231	N	PÓS	Possui domínio CMP/dCMP-tipo desaminase.	Citidina deaminase específica de DNA de cadeia simples. Envolvida na hipermutação somática, conversão de genes e recombinação de troca de classe de genes de imunoglobulina.	(MACDUFF; DEMOREST; HARRIS, 2009; METZNER; JÄCK; WABL, 2012)

Hs: *Homo sapiens*; Mm: *Mus musculus*; Gg: *Gallus gallus*; Dr: *Danio rerio*; Cg: *Cricetulus griseus*; Bm: *Bombyx mori*.

N: regulador negativo de L1; P: regulador positivo de L1.

ND: dados não disponíveis.

(continuação)

Gene	Descrição	Localização	Espécie	Tipo celular	Regulação	Nível	Domínios proteicos	Função	Referência
ALKBH1	<i>AlkB Homolog 1, Histone H2A Dioxygenase</i>	Núcleo e mitocôndria	Mm	ES	N	PRÉ	Possui um domínio Fe2OG dioxygenase e regiões de ligação de tRNA, ligação de substrato e de ligação alpha-cetoglutarato.	Dioxygenase que atua como ácidos nucleicos, como DNA e tRNA. Atua principalmente em tRNA e promove sua desmetilação ou oxidação, dependendo do contexto e do compartimento subcelular.	(WU et al., 2016)
APOBEC1	<i>Apolipoprotein B MRNA Editing Enzyme Catalytic Subunit 1</i>	Núcleo	Hs	HeLa, HEK293T	N	PÓS	Consiste em 3 domínios RRM (RRM1, RRM2 e RRM3) para ligação ao mRNA de APOB. Resíduos adicionais nas regiões pré-RRM e C-terminal são necessários para a ligação do RNA e para complementar a atividade de APOBEC1. Contém uma sequência para localização nuclear.	Envolvido na edição pós-transcricional do códon CAA para um códon UAA no mRNA de ApoB. Também funciona no códon CGA para UGA no mRNA da Neurofibromina 1 e e em processos epigenéticos.	(IKEDA et al., 2011; LINDIČ et al., 2013)
APOBEC3A	<i>Apolipoprotein B MRNA Editing Enzyme Catalytic Subunit 3A</i>	Núcleo	Hs	HeLa, HEK293T	N	PÓS	Possui domínio CMP/dCMP-tipo deaminase.	Citidina deaminase com atividade de restrição contra vírus e pode inibir a mobilidade de retroelementos LTR e não-LTR.	(BOGERD et al., 2006; CHIU et al., 2006; KINOMOTO et al., 2007; LIANG et al., 2016; LINDIČ et al., 2013; MUCKENFUS S et al., 2006; RENNER et al., 2018; RICHARDSON et al., 2014; STENGLEIN; HARRIS, 2006; TURELLI; VIANIN; TRONO, 2004)

Hs: *Homo sapiens*; Mm: *Mus musculus*; Gg: *Gallus gallus*; Dr: *Danio rerio*; Cg: *Cricetulus griseus*; Bm: *Bombyx mori*.

N: regulador negativo de L1; P: regulador positivo de L1.

ND: dados não disponíveis.

(continuação)

Gene	Descrição	Localização	Espécie	Tipo celular	Regulação	Nível	Domínios proteicos	Função	Referência
APOBEC3B	<i>Apolipoprotein B MRNA Editing Enzyme Catalytic Subunit 3B</i>	Núcleo	Hs	HeLa, HEK293T	N	PÓS	O domínio CMP/dCMP deaminase 1 atua como medidor na ligação de RNA, enquanto o domínio CMP/dCMP deaminase 2 confere atividade de desoxicitidina deaminase e especificidade de sequência de substrato.	Citidina deaminase que atua como um inibidor da replicação de retrovírus e da mobilidade de retroelementos (LTR e não-LTR) por meio de mecanismos dependentes e independentes da deaminase.	(BOGERD et al., 2006; CHIU et al., 2006; KINOMOTO et al., 2007; LIANG et al., 2016; LINDIČ et al., 2013; MUCKENFUS S et al., 2006; RENNER et al., 2018; RICHARDSON et al., 2014; STENGLEIN; HARRIS, 2006; TURELLI; VIANIN; TRONO, 2004)
APOBEC3C	<i>Apolipoprotein B MRNA Editing Enzyme Catalytic Subunit 3C</i>	Núcleo	Hs	HeLa, HEK293T	N	PÓS	Possui domínio CMP/dCMP-tipo deaminase e uma região para interação com proteína de vírus sincicial humano.		
APOBEC3D	<i>Apolipoprotein B MRNA Editing Enzyme Catalytic Subunit 3D</i>	Núcleo	Hs	HeLa, HEK293T	N	PÓS	O domínio CMP/dCMP deaminase 1 atua como mediador na ligação de RNA, enquanto o domínio CMP/dCMP deaminase 2 confere atividade de desoxicitidina deaminase e especificidade de sequência de substrato.		
APOBEC3F	<i>Apolipoprotein B MRNA Editing Enzyme Catalytic Subunit 3F</i>	Núcleo	Hs	HeLa, HEK293T	N	PÓS			
APOBEC3G	<i>Apolipoprotein B MRNA Editing Enzyme Catalytic Subunit 3G</i>	Núcleo e citoplasma	Hs	HeLa, HEK293T	N	PÓS			
APOBEC3H	<i>Apolipoprotein B MRNA Editing Enzyme Catalytic Subunit 3H</i>	Núcleo	Hs	HeLa, HEK293T	N	PÓS	Possui domínio CMP/dCMP-tipo deaminase.		
ASZ1	<i>Ankyrin Repeat, SAM And Basic Leucine Zipper Domain Containing 1</i>	Núcleo e citoplasma	Mm	ES	N	PÓS	Possui um domínio SAM.	Desempenha um papel central durante a espermatogênese, reprimindo elementos transponíveis e impedindo sua mobilização, essencial para a integridade da linha germinativa.	(MA et al., 2009)

Hs: *Homo sapiens*; Mm: *Mus musculus*; Gg: *Gallus gallus*; Dr: *Danio rerio*; Cg: *Cricetulus griseus*; Bm: *Bombyx mori*.

N: regulador negativo de L1; P: regulador positivo de L1.

ND: dados não disponíveis.

(continuação)

Gene	Descrição	Localização	Espécie	Tipo celular	Regulação	Nível	Domínios proteicos	Função	Referência
ATG5	<i>Autophagy Related 5</i>	Núcleo e citoplasma	Hs	HeLa, HEK293T	N	PÓS	ND	Envolvido na formação de vesícula autofágica.	(GUO et al., 2014)
ATM	<i>ATM Serine/Threonine Kinase</i>	Núcleo	Hs, Mm	HeLa, hESCs, HUES hESC, NPCs	N	PRÉ	Possui os domínios: FAT, PI3K/PI4K e FATC. O domínio FATC é necessário para interação com KAT5. Possui ainda uma região de interação com ABL1.	Ativa a sinalização de quebra de fita dupla, atuando assim como um sensor de dano ao DNA. Também envolvido na transdução de sinal e controle do ciclo celular.	(COUFAL et al., 2011; GASIOR et al., 2006; WALLACE et al., 2012)
BECN1	<i>Beclin 1</i>	Citoplasma, mitocôndria, núcleo, RE e endossomo	Hs	HeLa, HEK293T	N	PÓS	O domínio <i>coiled coil</i> pode mediar a dimerização com os domínios <i>coiled coil</i> de ATG14 ou UVRAG envolvidos na formação de complexos PI3K. O domínio ECD contém domínios de ligação poli-Gln, como o <i>motif</i> poli-Gln ATXN3. Como alguma ligação é observada com BECN1 sem o ECD, outros domínios de BECN1 também podem interagir com ATXN3.	Desempenha papel central na autofagia.	(GUO et al., 2014)
BRCA1	<i>BRCA1 DNA Repair Associated</i>	Núcleo, membrana plasmática e citoesqueleto	Hs	HeLa, HeLa-M2	N	PÓS	Os domínios BRCT reconhecem e se ligam ao <i>motif</i> pSXXF fosforilado nas proteínas. A interação com o <i>motif</i> pSXXF fosforilado de ABRAXAS1, recruta BRCA1 em sítios de dano de DNA. O domínio <i>zinc finger RING-type</i> interage com BAP1.	Atua na formação de cadeias de poliubiquitina ligadas a 'Lys-6' e desempenha um papel central no reparo do DNA, facilitando as respostas celulares aos danos do DNA. Atua como um ativador transcricional.	(LIU et al., 2018; MITA et al., 2020)

Hs: *Homo sapiens*; Mm: *Mus musculus*; Gg: *Gallus gallus*; Dr: *Danio rerio*; Cg: *Cricetulus griseus*; Bm: *Bombyx mori*.

N: regulador negativo de L1; P: regulador positivo de L1.

ND: dados não disponíveis.

(continuação)

Gene	Descrição	Localização	Espécie	Tipo celular	Regulação	Nível	Domínios proteicos	Função	Referência
BST2	<i>Bone Marrow Stromal Cell Antigen 2</i>	Membrana plasmática, endossomo, citoplasma, lisossomo e complexo de golgi	Hs	HEK 293T, 2102Ep, HeLa-HÁ, HeLa-JVM	N	PÓS	O domínio <i>coiled coil</i> forma uma estrutura extensa semelhante a uma haste semiflexível 170 Å, importante para a retenção do vírion na superfície da célula e prevenção da propagação do vírus.	Fator de restrição de hospedeiro antiviral induzido por IFN que inibe a infecção por retrovírus.	(GOODIER et al., 2015)
CALCOCO2	<i>Calcium Binding And Coiled-Coil Domain 2</i>	Citoplasma e núcleo	Hs	HeLa, HEK293T	N	PÓS	O domínio de ligação de MYO6 é necessário para a degradação mediada por autofagia de bactérias infectantes. O <i>motif</i> CLIR é necessário para interação com MAP1LC3C. O <i>motif</i> LIR-like é necessário para interação com MAP1LC3A, MAP1LC3B e GABARAPL2, bem como para maturação de autofagossomo mediada por CALCOCO2. O domínio de ligação do LGALS8 é essencial para recrutamento de bactérias infectantes.	Receptor específico para xenofagia necessário para degradação de bactérias intracelulares mediada por autofagia.	(GUO et al., 2014)
CBX1	<i>Chromobox 1</i>	Núcleo e citoesqueleto	Hs	HeLa	P	PRÉ	Possui os domínios Chromo 1 e Chromo 2.	Componente da heterocromatina. Reconhece e liga as caudas de histonas H3 metiladas em 'Lys-9', levando à repressão epigenética.	(LIU et al., 2018; MITA et al., 2020)
CDKN1A	<i>Cyclin Dependent Kinase Inhibitor 1A</i>	Núcleo e citoplasma	Hs	HEK293T, TET293T	N	PRÉ	O <i>motif</i> PIP-box K+4 atua tanto na interação com PCNA quanto o recrutamento do complexo DCX (DTL). O C-terminal é necessário para a localização nuclear do complexo ciclina D-CDK4.	Pode estar envolvido na inibição da proliferação celular mediada por p53 em resposta a danos no DNA.	(KAWANO et al., 2018)

Hs: *Homo sapiens*; Mm: *Mus musculus*; Gg: *Gallus gallus*; Dr: *Danio rerio*; Cg: *Cricetulus griseus*; Bm: *Bombyx mori*.

N: regulador negativo de L1; P: regulador positivo de L1.

ND: dados não disponíveis.

(continuação)

Gene	Descrição	Localização	Espécie	Tipo celular	Regulação	Nível	Domínios proteicos	Função	Referência
CHAF1A	<i>Chromatin assembly factor 1</i>	Núcleo	Mm	Células germinativas	N	PRÉ	Contém um <i>motif</i> PxVxL, que é necessário para a interação com domínios de sombra cromática.	Componente central do complexo CAF-1, um complexo que se pensa mediar a montagem da cromatina na replicação e reparo do DNA.	(HATANAKA et al., 2015)
DCLRE1C	<i>DNA Cross-Link Repair 1C</i>	Núcleo e complexo de golgi	Gg	DT40	P	PRÉ	ND	Necessário para a recombinação V(D)J, processo pelo qual os éxons que codificam os domínios de ligação ao antígeno de imunoglobulinas e proteínas do receptor de células T são montados a partir de segmentos gênicos individuais V,(D) e J.	(SUZUKI et al., 2009)
DDX21	<i>DExD-Box Helicase 21</i>	Núcleo e citoplasma	Hs	2102Ep, HeLa, HEK293T, PA-1	N	PRÉ	Atividades de helicase e foldase.	RNA helicase que atua como um sensor do status transcricional da RNA polimerase (Pol) I e II: promove o processamento e a transcrição do RNA ribossomal (rRNA) da polimerase II (Pol II).	(GOODIER; CHEUNG; KAZAZIAN, 2013; MOLDOVAN; MORAN, 2015)
DGCR8	<i>DGCR8, Microprocessor Complex Subunit</i>	Núcleo	Hs	HEK293, HeLa	N	PÓS	Os domínios DRBM1 e DRBM2 são necessários para a ligação eficiente a pri-miRNA. Possui ainda um domínio WW além de regiões necessárias para: interação com NCL, localização nuclear e de interação com DROSHA.	Componente do complexo do microprocessador que atua como uma proteína de ligação ao RNA, envolvida na etapa inicial da biogênese de miRNA.	(HERAS et al., 2013; MACIAS et al., 2012)

Hs: *Homo sapiens*; Mm: *Mus musculus*; Gg: *Gallus gallus*; Dr: *Danio rerio*; Cg: *Cricetulus griseus*; Bm: *Bombyx mori*.

N: regulador negativo de L1; P: regulador positivo de L1.

ND: dados não disponíveis.

(continuação)

Gene	Descrição	Localização	Espécie	Tipo celular	Regulação	Nível	Domínios proteicos	Função	Referência
DHX9	<i>DExH-Box Helicase 9</i>	Núcleo, citoesqueleto e citoplasma	Hs	2102Ep, HeLa, HEK293T, PA-1	N	PRÉ	Os domínios DRBM cooperam para a ligação ao ácido nucleico. A região HA2 é essencial para a atividade da helicase de desenrolamento do RNA. A região MTAD atua na interação com RNA polimerase II e estimula a ativação transcricional de uma maneira dependente de CREB. As regiões OB-fold e as regiões repetidas de arginina e glicina-glicina (RGG) são dispensáveis para ambas as atividades de helicase de ligação e desenrolamento de RNA. A região RGG contém o sinal de localização nuclear (NLS) e o sinal de exportação nuclear (NES) e é necessária e suficiente para o transporte nucleocitoplasmático de uma maneira independente do RNA.	Possui funções reguladoras na replicação de DNA, ativação pré-transcricional, regulação de RNA pós-transcricional, tradução de mRNA e silenciamento de gene mediado por RNA. Também é capaz de se ligar e neutralizar estruturas de RNA de fita dupla formadas por retroelementos Alu que são inseridos e transcritos como partes de genes.	(GOODIER; CHEUNG; KAZAZIAN, 2013; MOLDOVANA; MORAN, 2015)
DICER1	<i>Dicer 1, Ribonuclease III</i>	Extracelular, núcleo e citoplasma	Hs	HeLa, HEK293	N	PÓS	Possui os domínios Helicase ATP-binding, Helicase C-terminal, Dicer dsRNA-binding fold, PAZ, RNase III 1, RNase III 2 e DRBM.	Endoribonuclease de dsRNA que desempenha um papel central no silenciamento de genes a nível pós-transcricional mediado por dsRNA.	(CHEN et al., 2012; YANG; KAZAZIAN, 2006)

Hs: *Homo sapiens*; Mm: *Mus musculus*; Gg: *Gallus gallus*; Dr: *Danio rerio*; Cg: *Cricetulus griseus*; Bm: *Bombyx mori*.

N: regulador negativo de L1; P: regulador positivo de L1.

ND: dados não disponíveis.

(continuação)

Gene	Descrição	Localização	Espécie	Tipo celular	Regulação	Nível	Domínios proteicos	Função	Referência
DNMT1	DNA Methyltransferase 1	Núcleo	Hs, Mm	ES, hES	N	PRÉ	A parte N-terminal é necessária para homodimerização. O <i>zinc finger</i> CXXC- <i>type</i> liga-se a dinucleotídeos CpG não metilados, posicionando o ligante autoinibitório entre o DNA e o sítio ativo, proporcionando assim um mecanismo para garantir que apenas os dinucleotídeos CpG hemimetilados sofram metilação.	Responsável por manter os padrões de metilação celular.	(CASTRO-DIAZ et al., 2014; JÖNSSON et al., 2019; LIANG et al., 2002)
DNMT3A	DNA Methyltransferase 3 Alpha	Núcleo	Hs, Mm	ES, hES	N	PRÉ	Possui domínios PWWP, ADD e SAM-dependente MTase C5-tipo. O domínio PWWP é essencial para direcionar a heterocromatina pericêntrica. Possui também regiões de interação com DNMT1, DNMT3A e com o complexo PRC2/EED-EZH2.	Necessário para metilações novas em todo o genoma e é essencial para o estabelecimento de padrões de metilação de DNA durante o desenvolvimento.	(CASTRO-DIAZ et al., 2014; KATO et al., 2007; LI et al., 2015; LIANG et al., 2002)
DNMT3B	DNA Methyltransferase 3 Beta	Núcleo	Hs, Mm	ES, hES	N	PRÉ			
Dnmt3c	DNA methyltransferase 3C	Núcleo	Mm	ESCs	N	PRÉ	Possui domínios ADD e SAM-dependente MTase C5-tipo.	DNA metiltransferase que metila especificamente os promotores de retroelementos evolutivamente jovens na linha germinativa.	(BARAU et al., 2016)

Hs: *Homo sapiens*; Mm: *Mus musculus*; Gg: *Gallus gallus*; Dr: *Danio rerio*; Cg: *Cricetulus griseus*; Bm: *Bombyx mori*.
N: regulador negativo de L1; P: regulador positivo de L1.
ND: dados não disponíveis.

(continuação)

Gene	Descrição	Localização	Espécie	Tipo celular	Regulação	Nível	Domínios proteicos	Função	Referência
DNMT3L	<i>DNA Methyltransferase 3 Like</i>	Núcleo	Mm	células germinativas	N	PRÉ	Possui domínio ADD.	Fator regulador cataliticamente inativo de DNA metiltransferases que podem promover ou inibir a metilação do DNA.	(KATO et al., 2007)
DROSHA	<i>Drosha ribonuclease type III</i>	Núcleo	Hs	HEK293, PA-1, HeLa	N	PÓS	Os domínios RNase III1 e RNase III2 formam um dímero intramolecular em que o 1 corta a cadeia 3's enquanto o 2 cliva a cadeia 5's de pri-miRNAs. Possui também um domínio DRBM e uma região necessária para interação com DGCR8.	Ribonuclease III específica de RNA de fita dupla que está envolvida na etapa inicial da biogênese do miRNA. Componente do complexo de microprocessador necessário para processar os transcritos primários de miRNA (pri-miRNAs) para liberar miRNA precursor (pré-miRNA) no núcleo.	(HERAS et al., 2013)
E2F1	<i>E2F Transcription Factor 1</i>	Núcleo e cito esqueleto	Hs, Mm	MCF7, HeLa, MEFs	N	PRÉ	Possui o <i>motif</i> DEF box.	Ativador de transcrição que liga o DNA de forma cooperativa às proteínas DP através do sítio de reconhecimento E2, 5'-TTTC [CG] CGC-3' encontrado na região promotora de vários genes cujos produtos estão envolvidos na regulação do ciclo celular ou na replicação do DNA.	(MONTROYA-DURANGO et al., 2009, 2016)
E2F4	<i>E2F Transcription Factor 4</i>	Núcleo	Hs, Mm	MCF7, HeLa, MEFs	N	PRÉ	Possui os <i>motifs</i> DEF box e HBM, além de regiões de dimerização, transativação e de interação com RBL1 e RBL2.	ND	(MONTROYA-DURANGO et al., 2009, 2016)

Hs: *Homo sapiens*; Mm: *Mus musculus*; Gg: *Gallus gallus*; Dr: *Danio rerio*; Cg: *Cricetulus griseus*; Bm: *Bombyx mori*.

N: regulador negativo de L1; P: regulador positivo de L1.

ND: dados não disponíveis.

(continuação)

Gene	Descrição	Localização	Espécie	Tipo celular	Regulação	Nível	Domínios proteicos	Função	Referência
EHMT2	<i>Euchromatic Histone Lysine Methyltransferase 2</i>	Núcleo	Mm	SSC	N	PRÉ	O domínio SET faz a interação com o WIZ. No domínio pré-SET, os resíduos Cys ligam-se a 3 íons zinco; alguns desses resíduos Cys contribuem para a ligação de dois íons zinco dentro do cluster. Possui ainda 7 repetições ANK que ligam H3K9me1 e H3K9me2.	Histona metiltransferase que mono e dimetila 'Lys-9' da histona H3 na eucromatina. Também faz a monometilação de 'Lys-56' da histona H3 na fase G1, levando a promover a interação entre a histona H3 e o PCNA e regulando a replicação do DNA. Também metila fracamente 'Lys-27' da histona H3.	(DI GIACOMO et al., 2014)
ERCC1	<i>ERCC Excision Repair 1, Endonuclease Non-Catalytic Subunit</i>	Núcleo e citoplasma	Hs	HeLa	N	PRÉ	Contém um <i>motif</i> de sinal de localização nuclear. E uma região HhH2, dimerização com ERCC4.	Componente não catalítico de uma endonuclease de reparo de DNA específica da estrutura responsável pela incisão em 5' durante o reparo do DNA.	(GASIOR; ROY-ENGEL; DEININGER, 2008)
ERCC2	<i>ERCC Excision Repair 2, TFIIH Core Complex Helicase Subunit</i>	Núcleo, citoesqueleto e citoplasma	Hs	iPSC	N	PRÉ	Possui um domínio Helicase ATP-binding, um <i>motif</i> DEAH box e um <i>motif</i> de sinal de localização nuclear. Também apresenta uma região de interação com MMS19.	Componente do complexo geral de transcrição e do fator de reparo de DNA IIH (TFIIH), que está envolvido no reparo por excisão de nucleotídeo acoplado a transcrição (NER) geral e, quando complexado com CAK, na transcrição de RNA pela RNA polimerase II.	(AOTO et al., 2019; SERVANT et al., 2017)
ERCC4	<i>ERCC Excision Repair 4, Endonuclease Catalytic Subunit</i>	Núcleo	Hs	HeLa	N	PRÉ	Possui um domínio ERCC4 e contém um <i>motif</i> de sinal de localização nuclear.	Juntamente com o ERCC1, o ERCC4 forma o complexo enzimático ERCC1-XPF que participa do reparo e recombinação do DNA.	(GASIOR; ROY-ENGEL; DEININGER, 2008)

Hs: *Homo sapiens*; Mm: *Mus musculus*; Gg: *Gallus gallus*; Dr: *Danio rerio*; Cg: *Cricetulus griseus*; Bm: *Bombyx mori*.

N: regulador negativo de L1; P: regulador positivo de L1.

ND: dados não disponíveis.

(continuação)

Gene	Descrição	Localização	Espécie	Tipo celular	Regulação	Nível	Domínios proteicos	Função	Referência
ETS1	<i>ETS Proto-Oncogene 1, Transcription Factor</i>	Núcleo	Hs	LoVo	P	PRÉ	Possui domínio PNT e domínio de ativação (necessário para a ativação da transcrição).	Controla diretamente a expressão de genes de citocinas e quimiocinas. Pode controlar a diferenciação, sobrevivência e proliferação de células linfóides. Pode regular a angiogênese.	(LI et al., 2014; YANG et al., 1998)
EXD1	<i>Exonuclease 3'-5' Domain Containing 1</i>	Núcleo e citoplasma	Mm, Dm, Bm	Célula germinativa, BmN4	N	PÓS	Possui o domínio 3'-5' exonuclease não possui os resíduos conservados Asp-Glu-Asp-Asp (DEDD) que coordenam íons divalentes essenciais para a atividade de exonuclease.	Componente de ligação a RNA do complexo PET, um complexo multiproteico necessário para o processamento de piRNAs durante a espermatogênese.	(YANG et al., 2016)
FAM120A	<i>Family With Sequence Similarity 120A</i>	Citoplasma e núcleo	Hs	2102Ep, HeLa, HEK293T, PA-1	N	PÓS	Possui uma região de interação com YES1, SRC e FYN e uma região de ligação de RNA.	Pode participar do transporte de mRNA no citoplasma. Componente crítico da sinalização de sobrevivência induzida pelo estresse oxidativo.	(GOODIER; CHEUNG; KAZAZIAN, 2013; MOLDOVAN; MORAN, 2015)
FANCA	<i>FA Complementation Group A</i>	Núcleo	Hs	HeLa	P	PRÉ	Possui um <i>motif</i> sinal de localização nuclear.	Proteína de reparo de DNA que pode operar em um reparo pós-replicação ou em uma função de checkpoint do ciclo celular.	(LIU et al., 2018; MITA et al., 2020)
FANCB	<i>FA Complementation Group B</i>	Núcleo	Hs	HeLa	N	PRÉ	ND	Reparo de DNA, necessário para ubiquitinação de FANCD2.	(LIU et al., 2018; MITA et al., 2020)
FANCD2	<i>FA Complementation Group D2</i>	Núcleo e citoplasma	Mm, Hs	PGCs, PD20, RA3331	N	PRÉ	Possui regiões de interação com FENCE e BRCA2.	Necessário para manutenção da estabilidade cromossômica.	(BRÉGNARD et al., 2016; NIE et al., 2020)

Hs: *Homo sapiens*; Mm: *Mus musculus*; Gg: *Gallus gallus*; Dr: *Danio rerio*; Cg: *Cricetulus griseus*; Bm: *Bombyx mori*.

N: regulador negativo de L1; P: regulador positivo de L1.

ND: dados não disponíveis.

(continuação)

Gene	Descrição	Localização	Espécie	Tipo celular	Regulação	Nível	Domínios proteicos	Função	Referência
FANCM	<i>FA Complementation Group M</i>	Núcleo	Hs	HeLa	N	PRÉ	Possui um domínio Helicase ATP-binding e um domínio Helicase C-terminal. Regiões de interação com CENPS/CENPSX e FAAP24. Um <i>motif</i> DEAH box.	Componente do complexo central da anemia de Fanconi (AF). Necessário para a ativação normal da via AF, levando à monoubiquitinação do complexo FANCI-FANCD2 em resposta a danos no DNA, resistência celular a drogas de reticulação de DNA e prevenção de quebra cromossômica.	(LIU et al., 2018; MITA et al., 2020)
FASTKD2	<i>FAST Kinase Domains 2</i>	Núcleo, mitocôndria e extracelular	Hs	HeLa	P	PRÉ	Possui um domínio RAP.	Tem papel na montagem da grande subunidade ribossomal mitocondrial. Controla a abundância de 16S mt-rRNA e é necessário para a tradução intra-mitocondrial.	(LIU et al., 2018; MITA et al., 2020)
FCGR1B	<i>Fc Fragment Of IgG Receptor 1b</i>	Membrana plasmática e endossomo	Hs	HeLa	P	PÓS	Possui os domínios Ig-like C2-type 1 e Ig-like C2-type 2.	Pode atuar na resposta imune humoral.	(LIU et al., 2018; MITA et al., 2020)
FKBP6	<i>FK506 Binding Protein 6</i>	Núcleo e citoplasma	Hs	HEK293T	N	PÓS	Possui um domínio PPlase FKBP-tipo.	Necessário durante a espermatogênese para reprimir elementos transponíveis e impedir sua mobilização, essencial para a integridade da linha germinativa.	(XIOL et al., 2012)
GPAT2	<i>Glycerol-3-Phosphate Acyltransferase 2, Mitochondrial</i>	Mitocôndria	Mm	Célula germinativa	N	PÓS	O <i>motif</i> HXXXXD é essencial para a atividade da aciltransferase.	Necessário para a etapa de processamento primário durante a biossíntese de piRNA.	(SHIROMOTO et al., 2019)

Hs: *Homo sapiens*; Mm: *Mus musculus*; Gg: *Gallus gallus*; Dr: *Danio rerio*; Cg: *Cricetulus griseus*; Bm: *Bombyx mori*.

N: regulador negativo de L1; P: regulador positivo de L1.

ND: dados não disponíveis.

(continuação)

Gene	Descrição	Localização	Espécie	Tipo celular	Regulação	Nível	Domínios proteicos	Função	Referência
GTSF1	<i>Gametocyte Specific Factor 1</i>	Núcleo	Mm	ES	N	PÓS	Possui <i>zinc finger</i> CHHC U11-48K-tipo 1 e CHHC U11-48K-tipo 2.	Necessário para espermatogênese e está envolvido na supressão da transcrição de retroelementos.	(YOSHIMURA et al., 2009)
HECTD1	<i>HECT Domain E3 Ubiquitin Protein Ligase 1</i>	Citoplasma	Hs	HeLa	P	PÓS	ND	Atua na poliubiquitinação ligada a 'Lys-63' de HSP90AA1 que leva à sua localização intracelular e secreção reduzida.	(LIU et al., 2018; MITA et al., 2020)
HENMT1	<i>HEN Methyltransferase 1</i>	Núcleo	Mm	Células germinativas	N	PÓS	ND	Metiltransferase que adiciona um grupo 2'-O-metil na extremidade 3' de piRNAs, estabilizando estas moléculas.	(HEMPFLING et al., 2017; LIM et al., 2015)
HNRNPA2B1	<i>Heterogeneous Nuclear Ribonucleoprotein A2/B1</i>	Extracelular e núcleo	Hs	2102Ep, HeLa, HEK293T, PA-1	N	PÓS	Possui dois domínios RRM. A região de baixa complexidade (LC) é intrinsecamente desordenada. Uma série de evidências sugere que a formação de estruturas beta cruzadas por regiões LC medeiam a formação de grânulos de RNA.	hnRNP que se associa a pré-mRNAs nascentes, empacotando-os em partículas de hnRNP.	(GOODIER; CHEUNG; KAZAZIAN, 2013; MOLDOVAN; MORAN, 2015)
HNRNPC	<i>Heterogeneous Nuclear Ribonucleoprotein C (C1/C2)</i>	Citoesqueleto, núcleo e citoplasma	Hs	2102Ep, HeLa, HEK293T, PA-1	N	PÓS	Possui domínio RRM e um sinal de localização nuclear.	Liga o pré-mRNA e faz a nucleação da montagem de partículas de 40S hnRNP. Interage com os tratos poli-U na 3'-UTR ou 5'-UTR do mRNA e modula a estabilidade e o nível de tradução das moléculas de mRNA ligadas.	(ATTIG; ULE, 2019; GOODIER; CHEUNG; KAZAZIAN, 2013; MOLDOVAN; MORAN, 2015)

Hs: *Homo sapiens*; Mm: *Mus musculus*; Gg: *Gallus gallus*; Dr: *Danio rerio*; Cg: *Cricetulus griseus*; Bm: *Bombyx mori*.

N: regulador negativo de L1; P: regulador positivo de L1.

ND: dados não disponíveis.

(continuação)

Gene	Descrição	Localização	Espécie	Tipo celular	Regulação	Nível	Domínios proteicos	Função	Referência
HNRNPL	<i>Heterogeneous Nuclear Ribonucleoprotein L</i>	Núcleo, citoplasma, extracelular	Hs	HeLa, HeLa-JVM	P	PÓS	O domínio RRM2 tem afinidade de ligação ao RNA moderada. Os domínios RRM 3 e 4 podem facilitar o loop de RNA quando se ligam a dois locais de ligação separados adequadamente dentro do mesmo pré-mRNA alvo.	Fator de <i>splicing</i> agindo como ativador ou repressor da inclusão de éxons.	(GOODIER; CHEUNG; KAZAZIAN, 2013; MOLDOVAN; MORAN, 2015; PEDDIGARI et al., 2013)
HSP90AA1	<i>Heat Shock Protein 90 Alpha Family Class A Member 1</i>	Núcleo, citoplasma, membrana plasmática, extracelular e lisossomo	Mm	Células germinativas	N	PÓS	Possui regiões de interação com: NRC1, FLCN e FNIP1, FNIP2 e TSC1 e NR1D1. O <i>motif</i> TPR <i>repeat-binding</i> atua como mediador da interação com proteínas contendo repetição de TPR, como a co-chaperona STUB1.	Chaperona que promove a maturação, manutenção estrutural e regulação adequada de proteínas alvo específicas envolvidas, por exemplo, no controle do ciclo celular.	(ICHIYANAGI et al., 2014)
IFNB	<i>Interferon-Beta</i>	Extracelular e núcleo	Mm, Hs	MEFs, HeLa	N	PÓS	ND	Possui atividades antiviral, antibacteriana e anticâncer.	(YU et al., 2015)
ILF3	<i>Interleukin Enhancer Binding Factor 3</i>	Núcleo e mitocôndria	Hs	2102Ep, HeLa, HEK293T, PA-1	N	PRÉ	Possui os domínios: DZF, DRBM1 e DRBM2. Tem uma região de interação com PRMT1 e um sinal de localização nuclear bipartido.	Proteína de ligação a RNA que desempenha um papel essencial na biogênese de circRNAs. Dentro do núcleo, promove o processamento de circRNAs ao estabilizar os elementos reguladores que residem nos íntrons flanqueadores dos éxons circularizados. Desempenha, portanto, um papel no <i>back-splicing</i> de um subconjunto de circRNAs.	(GOODIER; CHEUNG; KAZAZIAN, 2013; MOLDOVAN; MORAN, 2015)

Hs: *Homo sapiens*; Mm: *Mus musculus*; Gg: *Gallus gallus*; Dr: *Danio rerio*; Cg: *Cricetulus griseus*; Bm: *Bombyx mori*.

N: regulador negativo de L1; P: regulador positivo de L1.

ND: dados não disponíveis.

(continuação)

Gene	Descrição	Localização	Espécie	Tipo celular	Regulação	Nível	Domínios proteicos	Função	Referência
ISG20	<i>Interferon Stimulated Exonuclease Gene 20</i>	Núcleo	Hs	HEK 293T, 2102Ep, HeLa	N	PÓS	ND	Exoribonuclease antiviral induzida por interferon que atua no RNA de fita simples e também possui menor atividade em relação ao DNA de fita simples. Pode também desempenhar papéis adicionais na maturação de snRNAs e rRNAs e na biogênese do ribossomo.	(GOODIER et al., 2015)
KDM1A	<i>Lysine Demethylase 1A</i>	Núcleo	Mm	Células germinativas	N	PRÉ	Possui o domínio SWIRM que pode atuar como um local de ancoragem para uma cauda de histona. E possui uma região de atividade de desmetilase.	Histona desmetilase que pode desmetilar 'Lys-4' (H3K4me) e 'Lys-9' (H3K9me) da histona H3, agindo assim como um coativador ou um corepressor.	(ANCELIN et al., 2016)
KDM4B	<i>Lysine Demethylase 4B</i>	Núcleo	Hs	MCF7, T47D, MDA-MB-231, MCF10A, ZR-75-30, HEK293T, HCT116, U2OS	P	PRÉ	Os 2 domínios Tudor reconhecem e ligam histonas metiladas. Ambos domínios Tudor tem uma estrutura interdigitada e a dobra incomum é necessária para sua capacidade de ligar caudas de histonas metiladas. Também possui os domínios JmjN e JmjC.	Histona desmetilase que desmetila 'Lys-9' da histona H3, desempenhando assim um papel no código da histona.	(XIANG et al., 2019)

Hs: *Homo sapiens*; Mm: *Mus musculus*; Gg: *Gallus gallus*; Dr: *Danio rerio*; Cg: *Cricetulus griseus*; Bm: *Bombyx mori*.

N: regulador negativo de L1; P: regulador positivo de L1.

ND: dados não disponíveis.

(continuação)

Gene	Descrição	Localização	Espécie	Tipo celular	Regulação	Nível	Domínios proteicos	Função	Referência
KPTN	<i>Kaptein, Actin Binding Protein</i>	Citoesqueleto e lisossomo.	Hs	HeLa	N	PÓS	ND	Como parte do complexo KICSTOR atua na detecção de aminoácidos da via de sinalização TORC1. Recruta o complexo GATOR1 para as membranas lisossomais e permite sua interação com GATOR2 e as RAG GTPases. É necessário para regular negativamente a sinalização de mTORC1.	(LIU et al., 2018; MITA et al., 2020)
LARP1	<i>La Ribonucleoprotein Domain Family Member 1</i>	Núcleo	Hs	2102Ep, HeLa, HEK293T, PA-1	N	PRÉ	Possui um domínio HTH <i>La-type RNA-binding</i> . A região C-terminal faz a mediação da interação com o mRNA e polissomos. É necessário para a repressão translacional de mRNAs com um <i>motif</i> 5'TOP. A região N-terminal faz a interação com PABPC1.	Proteína de ligação ao RNA que regula a tradução de espécies-alvo específicas de mRNA a jusante do complexo mTORC1, em função de sinais de crescimento e disponibilidade de nutrientes.	(GOODIER; CHEUNG; KAZAZIAN, 2013; MOLDOVAN; MORAN, 2015)
LIG4	<i>DNA Ligase 4</i>	Núcleo	Gg	DT40	N	PRÉ	Possui os domínios BRCT1 e BRCT2.	Reúne quebras de fita simples em um polidesoxinucleotídeo de fita dupla em uma reação dependente de ATP. Envolvido na união de extremidade não homóloga de DNA (NHEJ) necessária para reparo de quebra de fita dupla e recombinação V(D)J. O complexo LIG4-XRCC4 é responsável pela etapa de ligação NHEJ.	(SUZUKI et al., 2009)

Hs: *Homo sapiens*; Mm: *Mus musculus*; Gg: *Gallus gallus*; Dr: *Danio rerio*; Cg: *Cricetulus griseus*; Bm: *Bombyx mori*.

N: regulador negativo de L1; P: regulador positivo de L1.

ND: dados não disponíveis.

(continuação)

Gene	Descrição	Localização	Espécie	Tipo celular	Regulação	Nível	Domínios proteicos	Função	Referência
MAEL	<i>Maelstrom Spermatogenic Transposon Silencer</i>	Núcleo	Mm	ES	N	PÓS	ND	Desempenha um papel central durante a espermatogênese, reprimindo elementos transponíveis e impedindo sua mobilização.	(CHENG et al., 2017; SOPER et al., 2008)
MARF1	<i>Meiosis Regulator and mRNA Stability Factor 1</i>	Peroxissomo e complexo de Golgi	Hs	oócitos	N	PÓS	Possui os domínios: NYN, RRM, e oito domínios HTH OST-type.	Regulador essencial da oogênese necessário para a progressão meiótica feminina, reprimir os elementos transponíveis e prevenir sua mobilização. Também necessário para proteger contra quebras de fita dupla de DNA.	(SU et al., 2012b, 2012a)
MATR3	<i>Matrin 3</i>	Núcleo	Hs	2102Ep, HeLa, HEK293T, PA-1	N	PÓS	Possui domínios RRM (RRM1 e RRM2), um sinal de localização nuclear e um <i>zinc finger matrin-type</i> .	Pode desempenhar papel na transcrição ou interagir com outras proteínas da matriz nuclear para formar a rede fibrogranular interna. Em associação com o heterômero SFPQ-NONO pode desempenhar papel na retenção nuclear de RNAs defeituosos. Tem papel na regulação da resposta imune inata mediada por vírus de DNA, reunindo-se no complexo HDP-RNP. Pode se ligar a hairpins de miRNA específicos.	(ATTIG; ULE, 2019; GOODIER; CHEUNG; KAZAZIAN, 2013; MOLDOVAN; MORAN, 2015)

Hs: *Homo sapiens*; Mm: *Mus musculus*; Gg: *Gallus gallus*; Dr: *Danio rerio*; Cg: *Cricetulus griseus*; Bm: *Bombyx mori*.

N: regulador negativo de L1; P: regulador positivo de L1.

ND: dados não disponíveis.

Gene	Descrição	Localização	Espécie	Tipo celular	Regulação	Nível	Domínios proteicos	Função	Referência
MAVS	<i>Mitochondrial Antiviral Signaling Protein</i>	Mitocôndria e peroxissomo	Hs	HEK 293T, 2102Ep	N	PÓS	O motif pLxIS constitui um <i>motif</i> de ligação a IRF3: após a fosforilação por TBK1, o <i>motif</i> pLxIS fosforilado de MAVS recruta IRF3. IRF3 é então fosforilado e ativado por TBK1 para induzir interferons tipo I e outras citocinas. Os domínios CARD e transmembrana são essenciais para a função antiviral. O domínio CARD é responsável pela interação com DDX58/RIG-I e IFIH1/MDA5. O domínio transmembrana e os resíduos 300-444 são essenciais para sua interação com DHX58/LGP2.	Necessário para a defesa imune inata contra vírus. MAVS peroxissomal e mitocondrial agem sequencialmente para criar um estado celular antiviral. Após a infecção viral, o MAVS peroxissômico induz a rápida expressão independente do interferon de fatores de defesa que fornecem proteção de curto prazo, enquanto o MAVS mitocondrial ativa uma via de sinalização dependente do interferon com cinética retardada, que amplifica e estabiliza a resposta antiviral. Pode ativar as mesmas vias após a detecção de dsRNA extracelular por TLR3. Pode proteger as células da apoptose.	(GOODIER et al., 2015)
MECP2	<i>Methyl-CpG Binding Protein 2</i>	Núcleo e citoplasma	Hs, Mm	HeLa	N	PRÉ	Possui um domínio MBD. Regiões de interação com NCOR2 e com TBLXR1. Além de sequências ricas em histidina e prolina.	Proteína cromossômica que se liga ao DNA metilado. Atua na repressão transcrricional por meio da interação com a histona desacetilase e o corepressor SIN3A. Liga o DNA contendo 5-metilcitosina (5mC) e 5-hidroximetilcitosina (5hmC).	(MUOTRI et al., 2010; SKENE et al., 2010; YU, 2001)

Hs: *Homo sapiens*; Mm: *Mus musculus*; Gg: *Gallus gallus*; Dr: *Danio rerio*; Cg: *Cricetulus griseus*; Bm: *Bombyx mori*.

N: regulador negativo de L1; P: regulador positivo de L1.

ND: dados não disponíveis.

(continuação)

Gene	Descrição	Localização	Espécie	Tipo celular	Regulação	Nível	Domínios proteicos	Função	Referência
MEPCE	<i>Methylphosphate Capping Enzyme</i>	Núcleo	Hs	2102Ep, HeLa, HEK293T	N	PRÉ	Possui domínio Bin3-type SAM, três regiões de ligação S-adenosyl-L-methionine e uma região rica em glicina.	Metiltransferase dependente de S-adenosil-L-metionina que adiciona metilfosfato na extremidade 5' do 7SK snRNA, para estabilizá-lo. Também tem uma função não enzimática como parte do complexo 7SK RNP.	(GOODIER; CHEUNG; KAZAZIAN, 2013; TAYLOR et al., 2013)
miR128-1	<i>MicroRNA 128-1</i>	Núcleo	Hs	HeLa, iPSCs, Tera-1.	N	PÓS	ND	Ligação a mRNAs. Envolvido no silenciamento gênico pós-transcricional.	(FUNG et al., 2019; HAMDORF et al., 2015)
MORC1	<i>MORC Family CW-Type Zinc Finger 1</i>	Núcleo	Mm	Célula germinativa embrionária	N	PRÉ	Possui um zinc finger CW-type.	Necessário para espermatogênese.	(PASTOR et al., 2014)
MORC2	<i>MORC Family CW-Type Zinc Finger 2</i>	Núcleo e citosol	Mm, Hs	mESCs, K562, HeLa	N	PRÉ	ND	Essencial para o silenciamento epigenético pelo complexo HUSH. Reprime genes relacionados a células germinativas e retroelementos L1 em colaboração com SETDB1 e o complexo HUSH, o silenciamento é dependente de modificações epigenéticas repressivas, como H3K9me3.	(FUKUDA et al., 2018; LIU et al., 2018)

Hs: *Homo sapiens*; Mm: *Mus musculus*; Gg: *Gallus gallus*; Dr: *Danio rerio*; Cg: *Cricetulus griseus*; Bm: *Bombyx mori*.

N: regulador negativo de L1; P: regulador positivo de L1.

ND: dados não disponíveis.

(continuação)

Gene	Descrição	Localização	Espécie	Tipo celular	Regulação	Nível	Domínios proteicos	Função	Referência
MOV10	<i>Mov10 RISC Complex RNA Helicase</i>	Núcleo e citoplasma	Hs	HeLa, HEK293, HEK293T, 2102Ep	N	PÓS	Possui região de interação com AGO2 e APOBEC3G e um motif DEAG box.	Necessário para o silenciamento de genes mediado por miRNA pelo complexo RISC. Restringe a retrotransposição de L1 em cooperação com TUT4 e TUT7 neutralizando a atividade de RNA chaperonne de L1RE1. Facilita a uridilação LINE-1 por TUT4 e TUT7. Importante para viabilidade embrionária e desenvolvimento e função normais do SNC. Desempenha dois papéis críticos no desenvolvimento inicial do cérebro: suprime retroelementos no núcleo, inibindo diretamente a síntese de cDNA, enquanto regula mRNAs do citoesqueleto para influenciar o crescimento de neurites no citosol.	(ARJAN-ODEDRA et al., 2012; GOODIER et al., 2012; LI et al., 2013; MOLDOVAN; MORAN, 2015; SKARIAH et al., 2017)
MOV10L1	<i>Mov10 RISC Complex RNA Helicase Like 1</i>	Citoplasma	Mm	Célula germinativa	N	PÓS	Possui um motif DEAG box.	RNA helicase dependente de ATP, necessária durante a espermatogênese para reprimir elementos transponíveis e impedir sua mobilização.	(FROST et al., 2010)

Hs: *Homo sapiens*; Mm: *Mus musculus*; Gg: *Gallus gallus*; Dr: *Danio rerio*; Cg: *Cricetulus griseus*; Bm: *Bombyx mori*.

N: regulador negativo de L1; P: regulador positivo de L1.

ND: dados não disponíveis.

(continuação)

Gene	Descrição	Localização	Espécie	Tipo celular	Regulação	Nível	Domínios proteicos	Função	Referência
MPP8	<i>M-Phase Phosphoprotein 8</i>	Núcleo, membrana plasmática e citosol	Hs	K562, HeLa	N	PRÉ	Possui um domínio cromo que atua na interação com 'Lys-9' metilada da histona H3 (H3K9me). E quatro domínios ANK.	Promove o recrutamento de proteínas que atuam na repressão epigenética. Reprime L1 em colaboração com MORC2 e, provavelmente, SETDB1, o silenciamento é dependente de modificações epigenéticas repressivas, como a marca H3K9me3.	(LIU et al., 2018)
MRE11A	<i>MRE11 Homolog, Double Strand Break Repair Nuclease</i>	Núcleo e citoplasma	Hs	HeLa	P	PRÉ	ND	Componente do complexo MRN, que desempenha papel central no reparo de quebra de fita dupla, recombinação de DNA, manutenção da integridade dos telômeros e meiose.	(LIU et al., 2018; MITA et al., 2020)
MRPL28	<i>Mitochondrial Ribosomal Protein L28</i>	Mitocôndria	Hs	HeLa	P	PÓS	ND	Potencialmente representa um importante reagente terapêutico para pacientes com HLA-A24.	(LIU et al., 2018; MITA et al., 2020)

Hs: *Homo sapiens*; Mm: *Mus musculus*; Gg: *Gallus gallus*; Dr: *Danio rerio*; Cg: *Cricetulus griseus*; Bm: *Bombyx mori*.

N: regulador negativo de L1; P: regulador positivo de L1.

ND: dados não disponíveis.

(continuação)

Gene	Descrição	Localização	Espécie	Tipo celular	Regulação	Nível	Domínios proteicos	Função	Referência
MTNR1A	<i>Melatonin Receptor 1A</i>	Membrana plasmática	Hs	HeLa	N	PÓS	ND	Receptor de alta afinidade para melatonina. Provavelmente atua na mediação das ações reprodutivas e circadianas da melatonina.	(DEHARO et al., 2014)
MX2	<i>MX Dynamin Like GTPase 2</i>	Núcleo, citoplasma e mitocôndria	Hs	HEK 293T, 2102Ep, HeLa	N	PÓS	Possui um domínio Dynamin-type G e um GED.	GTPase do tipo dinamina induzida por interferon com potente atividade antiviral contra o vírus HIV-1.	(GOODIER et al., 2015)
NCAPD3	<i>Non-SMC Condensin II Complex Subunit D3</i>	Núcleo	Hs	HT-29	N	PRÉ	ND	Subunidade reguladora do complexo da condensina-2, o qual estabelece a arquitetura cromossômica mitótica e está envolvida na rigidez física do eixo da cromátide. Especificamente necessário para a decatenação de pontes de DNA ultrafinas centroméricas durante a anáfase. No início da neurogênese, pode desempenhar um papel essencial para garantir a condensação precisa dos cromossomos mitóticos nas células-tronco dos neurônios.	(WARD et al., 2017)

Hs: *Homo sapiens*; Mm: *Mus musculus*; Gg: *Gallus gallus*; Dr: *Danio rerio*; Cg: *Cricetulus griseus*; Bm: *Bombyx mori*.

N: regulador negativo de L1; P: regulador positivo de L1.

ND: dados não disponíveis.

(continuação)

Gene	Descrição	Localização	Espécie	Tipo celular	Regulação	Nível	Domínios proteicos	Função	Referência
NCAPH2	<i>Non-SMC Condensin II Complex Subunit H2</i>	Núcleo e extracelular	Hs	HT-29	N	PRÉ	ND	Subunidade reguladora do complexo da condensina-2, importante no estabelecimento da arquitetura cromossômica mitótica. No início da neurogênese, pode desempenhar um papel essencial para garantir a condensação precisa dos cromossomos mitóticos nas células-tronco dos neurônios.	(WARD et al., 2017)
NCL	<i>Nucleolin</i>	Núcleo	Hs	2102Ep, HeLa, HEK293T, PA-1	N	PÓS	Possui domínios RRM: RRM 1, RRM 2, RRM 3 e RRM 4	Principal proteína nucleolar do crescimento de células eucarióticas. É encontrada associada à cromatina intranucleolar e partículas pré-ribossômicas. Induz descondensação da cromatina pela ligação à histona H1.	(GOODIER; CHEUNG; KAZAZIAN, 2013; MOLDOVAN; MORAN, 2015)
PABPC1	<i>Poly(A) Binding Protein Cytoplasmic 1</i>	Núcleo e citoplasma	Hs	HeLa, HEK293T	P	PÓS	Os domínios de ligação a RNA RRM1 e RRM2 e as regiões do C-terminal (últimos 138 aminoácidos) interagem com o motif-1 de interação PABPC1 (PAM1) e -2 (PAM2) de PAIP1. Enquanto RRM2 e RRM3 e as regiões do C-terminal (últimos 138 aminoácidos) interagem com o motif-1 de interação PABPC1 (PAM1) e -2 (PAM2) de PAIP2.	Liga a cauda poli (A) do mRNA, incluindo a de seu próprio transcrito, e regula os processos do metabolismo do mRNA, como <i>splicing</i> pré-mRNA e estabilidade do mRNA.	(DAI et al., 2012; LIU et al., 2018; MITA et al., 2020)

Hs: *Homo sapiens*; Mm: *Mus musculus*; Gg: *Gallus gallus*; Dr: *Danio rerio*; Cg: *Cricetulus griseus*; Bm: *Bombyx mori*.

N: regulador negativo de L1; P: regulador positivo de L1.

ND: dados não disponíveis.

(continuação)

Gene	Descrição	Localização	Espécie	Tipo celular	Regulação	Nível	Domínios proteicos	Função	Referência
PABPC4	<i>Poly(A) Binding Protein Cytoplasmic 4</i>	Núcleo e citoplasma	Hs	2102Ep, HeLa, HEK293T	P	PÓS	Possui quatro domínios RRM e um PABC. Também apresenta uma região rica em alanina.	Liga a cauda poli(A) do mRNA. Pode estar envolvido nos processos de regulação citoplasmática do metabolismo do mRNA. Provavelmente pode se ligar a sequências de RNA citoplasmáticas diferentes de poli(A) in vivo.	(DAI et al., 2012; GOODIER; CHEUNG; KAZAZIAN, 2013; TAYLOR et al., 2013)
PABPN1	<i>Poly(A) Binding Protein Nuclear 1</i>	Núcleo	Hs	2102Ep, HeLa, HEK293T	P	PÓS	O domínio RRM é essencial para o reconhecimento de bases de adenina específicas na cauda poli (A), mas não é suficiente para a ligação de poli (A). Contém ainda sequências para interação com SKIP, para homooligomerização e para estimular e interagir com PAPOLA.	Envolvido na formação da extremidade 3' de pré-mRNAs pela adição de uma cauda poli (A) de 200-250 nt ao produto de clivagem a montante. Também está presente em vários estágios do metabolismo do mRNA.	(DAI et al., 2012; GOODIER; CHEUNG; KAZAZIAN, 2013)
PALB2	<i>Partner And Localizer Of BRCA2</i>	Núcleo	Hs	HeLa	N	PRÉ	A interação com BRCA2 ocorre através de uma bolsa hidrofóbica entre as repetições WD 4 e 5. O domínio <i>coiled coil</i> atua na autoassociação. O <i>motif</i> ChAM atua na associação com a cromatina.	Tem papel crítico no reparo de DNA por recombinação homóloga (HRR) por meio de sua capacidade de recrutar BRCA2 e RAD51 para quebras de DNA. Atua como arcabouço molecular na formação do complexo BRCA1-PALB2-BRCA2, que é essencial para a recombinação homóloga.	(LIU et al., 2018; MITA et al., 2020)

Hs: *Homo sapiens*; Mm: *Mus musculus*; Gg: *Gallus gallus*; Dr: *Danio rerio*; Cg: *Cricetulus griseus*; Bm: *Bombyx mori*.

N: regulador negativo de L1; P: regulador positivo de L1.

ND: dados não disponíveis.

(continuação)

Gene	Descrição	Localização	Espécie	Tipo celular	Regulação	Nível	Domínios proteicos	Função	Referência
PARP1	<i>Poly(ADP-Ribose) Polymerase 1</i>	Núcleo e mitocôndria	Hs	HEK293T	P	PRÉ	Possui os domínios BRCT, PARP alpha-helical e PARP catalytic. Possui também dois sinais de localização nuclear e <i>motifs</i> PARP-type.	Atua na poli-ADP-ribosilação de proteínas e desempenha um papel fundamental no reparo do DNA.	(MIYOSHI; MAKINO; MORAN, 2019; TAYLOR et al., 2013)
PARP2	<i>Poly(ADP-Ribose) Polymerase 2</i>	Núcleo	Hs	HEK293T	P	PRÉ	Possui os domínios PARP alpha-helical e PARP catalytic. Possui também dois sinais de localização nuclear.	Desempenha um papel fundamental no reparo de DNA.	(MIYOSHI; MAKINO; MORAN, 2019)
PCNA	<i>Proliferating Cell Nuclear Antigen</i>	Citoesqueleto e núcleo	Hs	HeLa, HEK293T	P	PÓS	Possui uma região de interação com NUDT15.	Envolvida no controle da replicação do DNA, aumentando a processabilidade da polimerase durante o alongamento da fita principal. Desempenha papel fundamental na resposta ao dano ao DNA.	(TAYLOR et al., 2013)
PIWIL1	<i>Piwi Like RNA-Mediated Gene Silencing 1</i>	Núcleo	Mm	ES	N	PÓS	O domínio PAZ se liga à extremidade 3'-2'-O-metilada de piRNAs. A região MID é necessária para o reconhecimento da uridina na primeira posição dos piRNAs. A D-box (caixa de destruição) atua como um sinal de reconhecimento para associação com o complexo APC/C, ubiquitinação e degradação. Possui ainda um domínio Piwi.	Endoribonuclease que desempenha um papel central em células germinativas pós natais, reprimindo elementos transponíveis e impedindo sua mobilização. . Atua por meio do processo metabólico piRNA, reprimindo os elementos transponíveis durante a meiose, formando complexos compostos por piRNAs e proteínas Piwi e governa a metilação e a subsequente repressão dos retroelementos.	(REUTER et al., 2011)

Hs: *Homo sapiens*; Mm: *Mus musculus*; Gg: *Gallus gallus*; Dr: *Danio rerio*; Cg: *Cricetulus griseus*; Bm: *Bombyx mori*.

N: regulador negativo de L1; P: regulador positivo de L1.

ND: dados não disponíveis.

(continuação)

Gene	Descrição	Localização	Espécie	Tipo celular	Regulação	Nível	Domínios proteicos	Função	Referência
PIWIL2	<i>Piwi Like RNA-Mediated Gene Silencing 2</i>	Núcleo	Mm	ESC	N	PÓS	Possui domínio PAZ e domínio Piwi.	Desempenha papel central durante a espermatogênese, reprimindo elementos transponíveis e evitando sua mobilização. Atua por meio do processo metabólico piRNA, reprimindo os elementos transponíveis durante a meiose, formando complexos compostos por piRNAs e proteínas Piwi e governa a metilação e a subsequente repressão dos retroelementos.	(DI GIACOMO et al., 2013)
PIWIL4	<i>Piwi Like RNA-Mediated Gene Silencing 4</i>	Núcleo	Mm	Células germinativas, ES	N	PÓS	Possui domínio PAZ e domínio Piwi.	ND	(KOJIMA-KITA et al., 2016; REUTER et al., 2011)
PLD6	<i>Phospholipase D Family Member 6</i>	Extracelular e mitocôndria.	Mm	ES	N	PÓS	Contém um domínio de fosfodiesterase PLD, sugerindo que tem um único meio-catalítico e requer homodimerização para formar um sítio ativo completo. Possui também uma região necessária para localização mitocondrial.	Endonuclease que desempenha um papel crítico na biogênese do RNA que interage com PIWI (piRNA) durante a espermatogênese.	(WATANABE et al., 2011)
PRKDC	<i>Protein Kinase, DNA-Activated, Catalytic Polypeptide</i>	Núcleo e citoplasma	Cg	CHO	N	PRÉ	Possui os seguintes domínios: HEAT1, HEAT2, TPR1, FAT, TPR2, TPR3, PI3K/PI4K e FATC.	Atua como um sensor molecular para danos ao DNA. Envolvido na junção final não homóloga de DNA (NHEJ), necessária para reparo de quebra de fita dupla (DSB) e recombinação V(D)J.	(MORRISH et al., 2007)

Hs: *Homo sapiens*; Mm: *Mus musculus*; Gg: *Gallus gallus*; Dr: *Danio rerio*; Cg: *Cricetulus griseus*; Bm: *Bombyx mori*.

N: regulador negativo de L1; P: regulador positivo de L1.

ND: dados não disponíveis.

(continuação)

Gene	Descrição	Localização	Espécie	Tipo celular	Regulação	Nível	Domínios proteicos	Função	Referência
PRMT5	<i>Protein Arginine Methyltransferase 5</i>	Núcleo, citoplasma e complexo de Golgi.	Mm	PGCs, ESCs	N	PRÉ	Possui um domínio SAM-dependent MTase PRMT-type.	Arginina metiltransferase que pode catalisar a formação de ômega-N monometilarginina (MMA) e dimetilarginina simétrica (sDMA).	(KIM et al., 2014; NIE et al., 2020)
PURA	<i>Purine Rich Element Binding Protein A</i>	Núcleo	Hs	2102Ep, HeLa, HEK293T, PA-1	N	PRÉ	Possui regiões ricas em glicina e ricas em glutamina/ácido glutâmico (parte do domínio de ativação transcricional).	Provável ativador de transcrição. Pode desempenhar um papel no início da replicação do DNA e na recombinação.	(GOODIER; CHEUNG; KAZAZIAN, 2013; MOLDOVAN; MORAN, 2015; TAYLOR et al., 2013)
RAD18	<i>RAD18 E3 Ubiquitin Protein Ligase</i>	Citoesqueleto e núcleo	Hs	HEK293T, HeLa, HCT116	N	PRÉ	Possui um domínio SAP e um <i>motif</i> LR.	Envolvido no reparo pós-replicação do DNA danificado por UV.	(ARIUMI et al., 2018)
RAD21	<i>RAD21 Cohesin Complex Component</i>	Núcleo, citoesqueleto e citoplasma	Hs	CRC	P	PRÉ	Possui regiões de interação com WAPL e PDS5B, e de interação com STAG1. A região C-terminal se associa a SMC1A, enquanto a parte N-terminal se liga a SMC3.	Como um membro do complexo de coesina, envolvido na coesão da cromátide irmã desde o momento da replicação do DNA na fase S até sua segregação na mitose. O complexo de coesina também pode desempenhar um papel na montagem do pólo do fuso durante a mitose.	(XU et al., 2014)

Hs: *Homo sapiens*; Mm: *Mus musculus*; Gg: *Gallus gallus*; Dr: *Danio rerio*; Cg: *Cricetulus griseus*; Bm: *Bombyx mori*.

N: regulador negativo de L1; P: regulador positivo de L1.

ND: dados não disponíveis.

(continuação)

Gene	Descrição	Localização	Espécie	Tipo celular	Regulação	Nível	Domínios proteicos	Função	Referência
RAD50	<i>RAD50 Double Strand Break Repair Protein</i>	Núcleo	Hs	HeLa	P	PRÉ	O domínio zinc-hook, contém 2 resíduos Cys que coordenam uma molécula de zinco com a ajuda dos 2 resíduos Cys do zinc-hook de outra molécula RAD50, formando um homodímero em forma de V. As duas cabeças do homodímero, constituem o domínio de ligação a ATP, interagem com o homodímero MRE11.	Componente do complexo MRN, que desempenha um papel central no reparo da quebra da fita dupla, recombinação de DNA, manutenção da integridade dos telômeros e meiose.	(LIU et al., 2018; MITA et al., 2020)
RB1	<i>RB Transcriptional Corepressor 1</i>	Núcleo	Hs	HeLa	N	PÓS	Possui o domínio pocket que liga-se ao domínio C fosforilado pela treonina, evitando a interação com complexos de fator de transcrição E2F/DP heterodimérico.	Regulador chave de entrada na divisão celular que atua como supressor de tumor. Atua como um repressor da transcrição dos genes alvo E2F1. A forma ativa subfosforilada de RB1 interage com E2F1 e reprime sua atividade de transcrição, levando à interrupção do ciclo celular. Envolvido diretamente na formação de heterocromatina. Recruta e visa as metiltransferases de histona SUV39H1, KMT5B e KMT5C, levando à repressão transcricional epigenética.	(MONTROYA-DURANGO et al., 2009)

Hs: *Homo sapiens*; Mm: *Mus musculus*; Gg: *Gallus gallus*; Dr: *Danio rerio*; Cg: *Cricetulus griseus*; Bm: *Bombyx mori*.

N: regulador negativo de L1; P: regulador positivo de L1.

ND: dados não disponíveis.

(continuação)

Gene	Descrição	Localização	Espécie	Tipo celular	Regulação	Nível	Domínios proteicos	Função	Referência
RCL1	<i>RNA Terminal Phosphate Cyclase Like 1</i>	Núcleo	Hs	HeLa	P	PÓS	ND	Atua na biogênese da subunidade ribossômica 40S nas primeiras etapas de processamento do pré-rRNA nos locais A0, A1 e A2 que são necessários para a maturação adequada do RNA 18S.	(LIU et al., 2018; MITA et al., 2020)
RDH8	<i>Retinol Dehydrogenase 8</i>	Membrana plasmática	Hs	HeLa	P	PÓS	ND	Retinol desidrogenase com preferência por NADP. Pode desempenhar um papel na regeneração do pigmento visual em alta intensidade de luz.	(LIU et al., 2018; MITA et al., 2020)
RNASEH2A	<i>Ribonuclease H2 Subunit A</i>	Núcleo e citoplasma	Hs	HeLa	P	PRÉ	ND	Subunidade catalítica de RNase HII, uma endonuclease que degrada especificamente o RNA de híbridos de RNA:DNA. Participa da replicação do DNA. Atua na excisão de ribonucleotídeos únicos de duplexes de DNA: RNA.	(BENITEZ-GUIJARRO et al., 2018)

Hs: *Homo sapiens*; Mm: *Mus musculus*; Gg: *Gallus gallus*; Dr: *Danio rerio*; Cg: *Cricetulus griseus*; Bm: *Bombyx mori*.
N: regulador negativo de L1; P: regulador positivo de L1.
ND: dados não disponíveis.

(continuação)

Gene	Descrição	Localização	Espécie	Tipo celular	Regulação	Nível	Domínios proteicos	Função	Referência
RNASEH2B	<i>Ribonuclease H2 Subunit B</i>	Núcleo	Hs	HeLa	P	PRÉ	ND	Subunidade não catalítica de RNase HII, uma endonuclease que degrada o RNA de híbridos de RNA:DNA. Participa da replicação do DNA. Atua na excisão de ribonucleotídeos únicos de duplexes de DNA: RNA.	(BENITEZ-GUIJARRO et al., 2018)
RNASEH2C	<i>Ribonuclease H2 Subunit C</i>	Núcleo	Hs	HeLa	P	PRÉ	ND		(BENITEZ-GUIJARRO et al., 2018)
RNASEL	<i>Ribonuclease L</i>	Citoplasma e mitocôndria	Hs	HeLa, Hey1b	N	PÓS	As nove repetições de anquirina (ANK 1 a 9) constituem o domínio de ligação 2-5A do N-terminal. Prevê-se que o domínio da proteína kinase seja cataliticamente inativo. O domínio da ribonuclease está localizado no C-terminal, sendo suficiente para a atividade da ribonuclease.	Endoribonuclease que funciona na resposta antiviral do interferon (IFN).	(ZHANG et al., 2014)
RPRD2	<i>Regulation Of Nuclear Pre-mRNA Domain Containing 2</i>	Núcleo	Hs	HeLa	P	PRÉ	Possui um domínio CID		(LIU et al., 2018; MITA et al., 2020)
RUNX3	<i>Runt Related Transcription Factor 3</i>	Núcleo e citoplasma	Hs	HeLa, 143B	P	PRÉ	Possui um domínio Runt. Também contém uma região rica em glicina e outra rica em prolina/serina/treonina no C-terminal, a qual é necessária para a ativação transcricional de genes alvo.	Forma o fator de ligação ao núcleo, complexo CBF). Os membros RUNX modulam a transcrição de seus genes alvo. Pode estar envolvido no controle da proliferação e/ou diferenciação celular. Em associação com ZFX3, regula positivamente a atividade do promotor CDKN1A.	(YANG et al., 2003)

Hs: *Homo sapiens*; Mm: *Mus musculus*; Gg: *Gallus gallus*; Dr: *Danio rerio*; Cg: *Cricetulus griseus*; Bm: *Bombyx mori*.

N: regulador negativo de L1; P: regulador positivo de L1.

ND: dados não disponíveis.

(continuação)

Gene	Descrição	Localização	Espécie	Tipo celular	Regulação	Nível	Domínios proteicos	Função	Referência
SAMHD1	<i>SAM And HD Domain Containing Deoxynucleoside Triphosphate Triphosphohydrolase 1</i>	Membrana plasmática e núcleo	Hs, Mm	HEK293T, HEK293, U937, HeLa	N	PÓS	Possui domínios SAM e HD. Em humanos, o domínio SAM não é necessário para a atividade de dNTPase e capacidade de restringir a infecção por vírus.	Proteína que atua como um fator de restrição do hospedeiro envolvido na resposta de defesa à vírus e como um regulador da replicação do DNA. Também suprime a atividade do retroelemento L1.	(DU et al., 2019; HU et al., 2015; WHITE et al., 2014; ZHAO et al., 2013; ZHU et al., 2013)
SETDB1	<i>SET Domain Bifurcated 1</i>	Núcleo e citoplasma	Hs, Mm	THP-1, mESCs	N	PRÉ	Os domínios pré-SET, SET e pós-SET são necessários para a atividade da metiltransferase. No domínio pré-SET, os resíduos Cys ligam-se a 3 íons zinco que estão dispostos em um agrupamento triangular. Também possui os domínios: Tudor1, Tudor2 e MBD.	Histona metiltransferase que trimetila especificamente 'Lys-9' da histona H3. Funciona principalmente em regiões de eucromatina, desempenhando assim um papel central no silenciamento de genes eucromáticos.	(CUELLAR et al., 2017; FUKUDA et al., 2018)
SIRT6	<i>Sirtuin 6</i>	Núcleo	Hs, Mm	HDF, MEFs	N	PRÉ	Possui um domínio deacetylase sirtuin-type e uma região rica em prolina.	Proteína desacetilase dependente de NAD. Possui atividade deacetylase em relação à histona H3K9Ac e H3K56Ac. Modula a acetilação da histona H3 na cromatina telomérica durante a fase S do ciclo celular.	(VAN METER et al., 2014)
SIRT7	<i>Sirtuin 7</i>	Núcleo e mitocôndria	Mm, Hs	MEFs, K562	N	PRÉ	Possui um domínio deacetylase sirtuin-type e uma região rica em arginina.	Pode atuar como uma desacetilase ou desacetilase, dependendo do contexto. Envolvido na repressão transcricional de L1 via desacetilação H3K18Ac.	(VAZQUEZ et al., 2019)

Hs: *Homo sapiens*; Mm: *Mus musculus*; Gg: *Gallus gallus*; Dr: *Danio rerio*; Cg: *Cricetulus griseus*; Bm: *Bombyx mori*.

N: regulador negativo de L1; P: regulador positivo de L1.

ND: dados não disponíveis.

(continuação)

Gene	Descrição	Localização	Espécie	Tipo celular	Regulação	Nível	Domínios proteicos	Função	Referência
SLX4	<i>SLX4 Structure-Specific Endonuclease Subunit</i>	Núcleo e citoplasma	Hs	PD20, RA3331	N	PRÉ	Possui um domínio BTB e regiões de interação com: SLX4IP, ERCC4 e MSH2; PLK1 e TERF2-TERF2IP; MUS81 e SLX1.	Subunidade reguladora que interage e aumenta a atividade de diferentes endonucleases de estrutura específica. Tem vários papéis distintos na proteção da estabilidade do genoma, resolvendo diversas formas de estruturas deletérias do DNA originadas de intermediários de replicação e recombinação e de danos ao DNA.	(BRÉGNARD et al., 2016)
SNUPN	<i>Snurportin 1</i>	Núcleo e citoplasma	Hs	HeLa	P	PÓS	Possui domínio IBB. Possui regiões necessárias para: interação com XPO1, interação com KPNB1 e m3G-cap U1 e U5 snRNP, interação com a estrutura m3G-cap e para a ligação à estrutura m3G-cap.	Funciona como um adaptador de importação nuclear específico de U snRNP. Liga-se especificamente aos snRNAs U do terminal m3G-cap.	(LIU et al., 2018; MITA et al., 2020)
SOX11	<i>SRY-Box 11</i>	Núcleo	Hs	RD, HEK293, NTera2D1, hASCs	P	PRÉ	Possui uma região necessária para a atividade de ativação transcricional e coativação sinérgica da atividade transcricional com POU3F2. Além de duas regiões ricas em glicina, uma em ácido aspártico e outra em serina.	Atua como um ativador transcricional. Liga-se cooperativamente com POU3F2/BRN2 ou POU3F1/OCT6 a promotores de genes, o que aumenta a ativação transcricional. Desempenha um papel redundante com SOX4 e SOX12 na sobrevivência celular de tecidos em desenvolvimento.	(ORQUEDA et al., 2018; TCHÉNIO; CASELLA; HEIDMANN, 2000)

Hs: *Homo sapiens*; Mm: *Mus musculus*; Gg: *Gallus gallus*; Dr: *Danio rerio*; Cg: *Cricetulus griseus*; Bm: *Bombyx mori*.

N: regulador negativo de L1; P: regulador positivo de L1.

ND: dados não disponíveis.

(continuação)

Gene	Descrição	Localização	Espécie	Tipo celular	Regulação	Nível	Domínios proteicos	Função	Referência
SOX2	<i>SRY-Box 2</i>	Núcleo e citoplasma	Hs, Mm	NSC	N	PRÉ	ND	Fator de transcrição que forma um complexo trimérico com OCT4 no DNA e controla a expressão de diversos genes envolvidos no desenvolvimento embrionário, como YES1, FGF4, UTF1 e ZFP206. Crítico para a embriogênese inicial e para a pluripotência de células-tronco embrionárias.	(KUWABARA et al., 2009; MUOTRI et al., 2005; TCHÊNIO; CASELLA; HEIDMANN, 2000)
SP1	<i>Transcription Factor Sp1</i>	Núcleo	Hs	HeLa, HepG2	P	PRÉ	Possui um domínio repressor, três de transativação, um VZV IE62-binding e domínio D. Contém também o motif 9aaTAD, que é um domínio de transativação. E possui regiões ricas em serina/treonina.	Fator de transcrição que pode ativar ou reprimir a transcrição em resposta a estímulos fisiológicos e patológicos.	(YANG et al., 1998)
SQSTM1	<i>Sequestosome 1</i>	Núcleo, citoplasma, lisossomo, RE e endossomo	Hs	HeLa, HEK293T	N	PÓS	O domínio UBA liga-se às cadeias de poliubiquitina ligadas a 'Lys-63'. Medeia a interação com TRIM55. O domínio PB1 medeia homooligomerização e interações com FHOD3, MAP2K5, NBR1, PRKCI, PRKCZ e WDR81. Ambos os domínios PB1 e UBA são necessários e suficientes para a localização nos corpos de inclusão contendo ubiquitina. O <i>motif</i> LIR medeia a interação com proteínas da família ATG8.	Receptor de autofagia necessário para a macroautofagia seletiva. Funciona como uma ponte entre a carga poliubiquitinada e os autofagossomos.	(GUO et al., 2014)

Hs: *Homo sapiens*; Mm: *Mus musculus*; Gg: *Gallus gallus*; Dr: *Danio rerio*; Cg: *Cricetulus griseus*; Bm: *Bombyx mori*.

N: regulador negativo de L1; P: regulador positivo de L1.

ND: dados não disponíveis.

(continuação)

Gene	Descrição	Localização	Espécie	Tipo celular	Regulação	Nível	Domínios proteicos	Função	Referência
STK17A	<i>Serine/Threonine Kinase 17a</i>	Núcleo e membrana plasmática	Hs	HeLa	P	PÓS	Possui o domínio <i>protein kinase</i> .	Atua como um regulador positivo da apoptose. Também atua como um regulador de espécies reativas de oxigênio celular.	(LIU et al., 2018; MITA et al., 2020)
SUV39H1	<i>Suppressor Of Variegation 3-9 Homolog 1</i>	Núcleo	Hs, Mm	HeLa, ESC	N	PRÉ	Possui o domínio SET que contém o sítio ativo da atividade enzimática. Já os domínios pré-SET e pós-SET são necessários para a atividade da metiltransferase. O domínio SET também participa da ligação estável à heterocromatina. Possui também um domínio Chromo.	Histona metiltransferase que trimetila especificamente 'Lys-9' da histona H3 usando H3 'Lys-9' monometilado como substrato.	(BULUT-KARSLIOGLU et al., 2014; MARTENS et al., 2005; VARSHNEY et al., 2015)
SUV39H2	<i>Suppressor Of Variegation 3-9 Homolog 2</i>	Núcleo	Hs, Mm	HeLa, ESC	N	PRÉ	ND	ND	(BULUT-KARSLIOGLU et al., 2014; MARTENS et al., 2005; VARSHNEY et al., 2015)
SYNCRIP	<i>Synaptotagmin Binding Cytoplasmic RNA Interacting Protein</i>	Núcleo, retículo endoplasmático e citosol	Hs	HeLa	P	PÓS	Possui três domínios RRM (RRM1, RRM2 e RRM3). O domínio contendo oito repetições Arg-Gly-Gly (RGG/RXR-box) pode estar envolvido na ligação de RNA e interações proteína-proteína. É metilado por PRMT1 e essencial para localização nuclear.	hnRNP envolvida em mecanismos de processamento de mRNA. Componente do complexo mediado por CRD que promove a estabilidade do mRNA de MYC.	(GOODIER; CHEUNG; KAZAZIAN, 2013; LIU et al., 2018; MITA et al., 2020; MOLDOVAN; MORAN, 2015)

Hs: *Homo sapiens*; Mm: *Mus musculus*; Gg: *Gallus gallus*; Dr: *Danio rerio*; Cg: *Cricetulus griseus*; Bm: *Bombyx mori*.

N: regulador negativo de L1; P: regulador positivo de L1.

ND: dados não disponíveis.

(continuação)

Gene	Descrição	Localização	Espécie	Tipo celular	Regulação	Nível	Domínios proteicos	Função	Referência
TARDBP	<i>TAR DNA Binding Protein</i>	Núcleo	Hs	HeLa	N	PRÉ	Um domínio N-terminal e dois <i>motifs</i> de reconhecimento de RNA (RRM1 e RRM2), seguidos por uma região rica em glicina C-terminal. Contém uma sequência de localização nuclear e uma sequência de exportação nuclear.	Envolvido em várias etapas da biogênese e processamento do RNA.	(LIU et al., 2018)
TASOR	<i>Transcription Activation Suppressor</i>	Núcleo	Hs	K562, HeLa	N	PRÉ	ND	Componente do complexo HUSH, um complexo que atua na repressão epigenética. Reprime L1 em colaboração com MORC2 e, provavelmente, SETDB1, o silenciamento é dependente de modificações epigenéticas repressivas, como a marca H3K9me3.	(LIU et al., 2018)
TDRD1	<i>Tudor Domain Containing 1</i>	Núcleo	Hs, Mm	HEK293T, célula germinativa	N	PÓS	Possui 4 domínios Tudor, sendo que os domínios Tudor 2 e 3 têm maior afinidade para peptídeos metilados com arginina.	Desempenha papel central durante a espermatogênese, reprimindo elementos transponíveis e impedindo sua mobilização. Reprime os elementos transponíveis durante a meiose, formando complexos compostos por piRNAs e proteínas Piwi e governa a metilação e a subsequente repressão dos retroelementos.	(REUTER et al., 2009)

Hs: *Homo sapiens*; Mm: *Mus musculus*; Gg: *Gallus gallus*; Dr: *Danio rerio*; Cg: *Cricetulus griseus*; Bm: *Bombyx mori*.

N: regulador negativo de L1; P: regulador positivo de L1.

ND: dados não disponíveis.

(continuação)

Gene	Descrição	Localização	Espécie	Tipo celular	Regulação	Nível	Domínios proteicos	Função	Referência
TDRD12	<i>Tudor Domain Containing 12</i>	Citoplasma	Bm	BmN4	N	PÓS	Possui três domínios: Tudor 1, Helicase ATP-binding e Tudor 2. Tem também um <i>motif</i> DEAH box.	Provável RNA helicase de ligação ao ATP, necessário durante a espermatogênese para reprimir elementos transponíveis e impedir sua mobilização. Atua por meio do processo metabólico piRNA.	(PANDEY et al., 2013)
TDRD5	<i>Tudor Domain Containing 5</i>	Mitocôndria, núcleo e citoplasma	Mm	ESC	N	PÓS	Possui três domínios HTH OST-type e um domínio Tudor.	Necessário durante a espermiogênese para reprimir elementos transponíveis e impedir sua mobilização. Reprime os elementos transponíveis durante a meiose, formando complexos compostos por piRNAs e proteínas Piwi e governam a metilação e subsequente repressão dos retroelementos.	(YABUTA et al., 2011)
TDRD9	<i>Tudor Domain Containing 9</i>	Núcleo	Mm	ES	N	PÓS	Possui três domínios: Helicase ATP-binding, Helicase C-terminal e Tudor. Tem também um <i>motif</i> DEAH box.	RNA helicase de ligação a ATP, necessário durante a espermatogênese. Necessário para reprimir elementos transponíveis e impedir sua mobilização. Reprime os elementos transponíveis durante a meiose, formando complexos compostos por piRNAs e proteínas Piwi e governam a metilação e subsequente repressão dos retroelementos.	(SHOJI et al., 2009)

Hs: *Homo sapiens*; Mm: *Mus musculus*; Gg: *Gallus gallus*; Dr: *Danio rerio*; Cg: *Cricetulus griseus*; Bm: *Bombyx mori*.

N: regulador negativo de L1; P: regulador positivo de L1.

ND: dados não disponíveis.

(continuação)

Gene	Descrição	Localização	Espécie	Tipo celular	Regulação	Nível	Domínios proteicos	Função	Referência
TDRKH	<i>Tudor And KH Domain Containing</i>	Mitocôndria	Mm	ES	N	PÓS	Possui três domínios: KH1, KH2 e Tudor.	Participa da via primária de biogênese do piRNA e é necessária durante a espermatogênese para reprimir elementos transponíveis e impedir sua mobilização. O processo metabólico piRNA atua reprimindo os elementos transponíveis durante a meiose, formando complexos compostos por piRNAs e proteínas Piwi e governam a metilação e subsequente repressão dos retroelementos.	(SAXE et al., 2013)
TEX19	<i>Testis Expressed 19</i>	Núcleo	Mm	ESCs	N	PÓS	Possui uma região de interação com LIRE1 e uma região importante para interação com piRNA.	Necessário durante a espermatogênese e do desenvolvimento da placenta, participando da repressão de retroelementos L1 e impedindo sua mobilização.	(MACLENNAN et al., 2017; REICHMANN et al., 2013)
TIMM8B	<i>Translocase Of Inner Mitochondrial Membrane 8 Homolog B</i>	Extracelular e mitocôndria	Hs	HeLa	P	PÓS	O <i>motif</i> Twin CX3C contém 4 resíduos Cys conservados que formam 2 ligações dissulfeto no espaço intermembranar mitocondrial.	Participa da importação e inserção de algumas proteínas transmembrana na membrana mitocondrial interna. Atua como uma proteína semelhante a uma chaperona que protege os precursores hidrofóbicos da agregação e os guia através do espaço intermembranar mitocondrial.	(LIU et al., 2018; MITA et al., 2020)

Hs: *Homo sapiens*; Mm: *Mus musculus*; Gg: *Gallus gallus*; Dr: *Danio rerio*; Cg: *Cricetulus griseus*; Bm: *Bombyx mori*.

N: regulador negativo de L1; P: regulador positivo de L1.

ND: dados não disponíveis.

(continuação)

Gene	Descrição	Localização	Espécie	Tipo celular	Regulação	Nível	Domínios proteicos	Função	Referência
TP53	<i>Tumor protein p53</i>	Núcleo, mitocôndria, citoplasma e RE	Mm	FL5.12	N	PRÉ	Possui sinal de exportação nuclear atua como um domínio de repressão transcricional. Os <i>motifs</i> TADI e TADII correspondem ambos aos <i>motifs</i> 9aaTAD que são domínios de transativação.	Atua como um supressor de tumor; induz parada de crescimento ou apoptose, dependendo das circunstâncias fisiológicas e do tipo celular.	(HAGAN; RUDIN, 2007; HARRIS et al., 2009; WYLIE et al., 2016)
TREX1	<i>Three Prime Repair Exonuclease 1</i>	Núcleo, RE e citoplasma	Mm	HeLa	N	PÓS	Possui regiões de: ligação de substrato, rica em prolina, para localização do retículo endoplasmático, de interação com UBQLN1 e para retenção citoplasmática.	Exonuclease de DNA celular 3'-a-5' que digere ssDNA e dsDNA. Atua metabolizando fragmentos de DNA a partir de retroelementos endógenos, incluindo elementos L1, LTR e SINE.	(STETSON et al., 2008)
TRIM28	<i>Tripartite motif containing 28</i>	Núcleo	Hs	hES, HeLa	N	PRÉ	O HP1- <i>box</i> é necessário para a ligação HP1. O PHD-tipo <i>zinc finger</i> aumenta a atividade transcricional de CEBPB, e junto com HP1- <i>box</i> e o domínio bromo montam a maquinaria necessária para a repressão de proteínas contendo o domínio KRAB. O <i>motif</i> RBCC/TRIM é necessário para a interação com o domínio KRAB das proteínas <i>zinc finger</i> KRAB. Contém um <i>motif</i> PxVxL.	Corepressor nuclear para proteínas (KRAB-ZFPs). Atua no silenciamento de genes recrutando CHD3.	(CASTRO-DIAZ et al., 2014; JACOBS et al., 2014; LIU et al., 2018; MITA et al., 2020)

Hs: *Homo sapiens*; Mm: *Mus musculus*; Gg: *Gallus gallus*; Dr: *Danio rerio*; Cg: *Cricetulus griseus*; Bm: *Bombyx mori*.

N: regulador negativo de L1; P: regulador positivo de L1.

ND: dados não disponíveis.

(continuação)

Gene	Descrição	Localização	Espécie	Tipo celular	Regulação	Nível	Domínios proteicos	Função	Referência
TRIM5	<i>Tripartite Motif Containing 5</i>	Núcleo e citoplasma	Hs	HEK293T, 210Ep	N	PÓS	O domínio <i>zinc finger B box-type</i> e o domínio <i>coiled coil</i> contribuem para a multimerização de ordem superior e inferior, respectivamente, que é essencial para a atividade de restrição. O domínio B30.2/SPRY atua como um domínio de reconhecimento de capsídeo. O domínio <i>zinc finger RING-type</i> confere atividade da ubiquitina ligase E3 e é essencial para a atividade de restrição de retrovírus, autoubiquitinação e multimerização de ordem superior.	Impede a infecção de retrovírus não adaptados ao hospedeiro. Bloqueia a replicação viral no início do ciclo de vida, após a entrada viral, mas antes da transcrição reversa. Também atua como um receptor de reconhecimento de padrão que ativa a sinalização imune inata em resposta à estrutura do capsídeo retroviral.	(VOLKMANN et al., 2020)
TROVE2	<i>Ro60, Y RNA Binding Protein</i>	Núcleo e citoplasma	Hs	2102Ep, HeLa, HEK293T, PA-1	N	PÓS	O domínio TROVE em forma de ferradura é construído com 7 repetições helicoidais semelhantes a HEAT e é fechado pelo domínio VWFA-like, dando origem a um monômero em forma de anel. O RNA de fita simples está ligado à cavidade central carregada positivamente. O <i>motif</i> MIDAS-like no domínio VWFA-like liga cátions de metal divalente.	Proteína de ligação a RNA que se liga a RNAs não codificantes dobrados incorretamente, rRNA pré-5S e várias pequenas moléculas de RNA (RNAs Y). Pode estabilizar alguns desses RNAs e protegê-los da degradação. Liga-se a e regula a expressão de retroelementos Alu endógenos.	(GOODIER; CHEUNG; KAZAZIAN, 2013; MOLDOVAN; MORAN, 2015; TAYLOR et al., 2013)

Hs: *Homo sapiens*; Mm: *Mus musculus*; Gg: *Gallus gallus*; Dr: *Danio rerio*; Cg: *Cricetulus griseus*; Bm: *Bombyx mori*.

N: regulador negativo de L1; P: regulador positivo de L1.

ND: dados não disponíveis.

(continuação)

Gene	Descrição	Localização	Espécie	Tipo celular	Regulação	Nível	Domínios proteicos	Função	Referência
TUT4	<i>Terminal Uridylyl Transferase 4</i>	Núcleo e citoplasma	Hs	HEK293, HEK293T, HeLa, PA-1, HFF, hESCs	N	PÓS	Utiliza dois módulos funcionais de múltiplos domínios durante a mudança para oligouridilação. O módulo catalítico (contendo os 3 domínios Zinc finger CCHC-type) é essencial para ambas as atividades. O módulo de interação Lin28 (LIM) na parte N-terminal é indispensável para a oligouridilação.	Uridiltransferase que promove a uridilação terminal de mRNAs com caudas curtas (menos de 25 nucleotídeos) de poli (A), facilitando a decomposição global de mRNA. Essencial para a maturação e fertilidade de oócitos.	(WARKOCKI et al., 2018)
TUT7	<i>Terminal Uridylyl Transferase 7</i>	Núcleo e citoplasma	Hs	HEK293, HEK293T, HeLa, PA-1, HFF, hESCs	N	PÓS	ND	ND	(WARKOCKI et al., 2018)
UBE2T	<i>Ubiquitin Conjugating Enzyme E2 T</i>	Núcleo	Hs	HeLa	N	PRÉ	Possui o domínio UBC core.	Aceita a ubiquitina do complexo E1 e catalisa sua ligação a outras proteínas. Catalisa a monoubiquitinação. Envolvido no reparo de DNA induzido por mitomicina C (MMC).	(LIU et al., 2018; MITA et al., 2020)
UHRF1	<i>Ubiquitin Like With PHD And Ring Finger Domains 1</i>	Núcleo	Mm	ESCs	N	PRÉ	As regiões tudor-like reconhecem e ligam a histona H3 não metilada em 'Arg-2', enquanto o zinc finger PHD-type reconhece e liga a histona H3 trimetilada em 'Lys-9'. O domínio YDG (ou SRA) reconhece e se liga ao DNA hemimetilado nas forquilhas de replicação. Ele contém uma bolsa de ligação que acomoda a 5-metilcitosina que é retirada do DNA duplex. O domínio YDG também reconhece e se liga a 5hmC.	Atua como um regulador epigenético chave ao unir a metilação do DNA e a modificação da cromatina. Também desempenha um papel fundamental na modificação da cromatina. Pode estar envolvido no reparo do DNA.	(DONG et al., 2019)

Hs: *Homo sapiens*; Mm: *Mus musculus*; Gg: *Gallus gallus*; Dr: *Danio rerio*; Cg: *Cricetulus griseus*; Bm: *Bombyx mori*.

N: regulador negativo de L1; P: regulador positivo de L1.

ND: dados não disponíveis.

(continuação)

Gene	Descrição	Localização	Espécie	Tipo celular	Regulação	Nível	Domínios proteicos	Função	Referência
UPF1	<i>UPF1, RNA Helicase And ATPase</i>	Núcleo e citoplasma	Hs	HeLa, HEK293T	N/P	PÓS	Possui dois motifs [ST]-Q. O motif [SQ]-Q constitui uma sequência de reconhecimento para quinases da família PI3/PI4-quinase. Também possui uma região para interação com RENT2.	Helicase dependente de RNA e ATPase necessárias para a diminuição de mRNAs contendo códons de parada prematuros. É recrutado para mRNAs após o término da tradução e passa por um ciclo de fosforilação e desfosforilação.	(GOODIER; CHEUNG; KAZAZIAN, 2013; TAYLOR et al., 2013)
UPF2	<i>UPF2 Regulator Of Nonsense Mediated MRNA Decay</i>	Núcleo e citoplasma	Hs	HeLa, HEK293T	N/P	Pós-transcricional	Possui os domínios MIF4G 1, MIF4G 2 e MIF4G 3. Apresenta regiões de interação com: UPF1, UPF3A e UPF3B, EIF4A1 e EIF1 e UPF1 C-terminal.	Envolvido no mecanismo de diminuição de mRNA contendo códons de parada prematuros por associação com o complexo de junção de éxons nucleares (EJC).	(TAYLOR et al., 2013)
UPF3A	<i>UPF3A Regulator Of Nonsense Mediated MRNA Decay</i>	Núcleo e citoplasma	Hs	HeLa, HEK293T	N/P	PÓS	Possui regiões necessárias para associação com os componentes principais EIF4A3 e ECJ CASC3, MAGOH e RBM8A e para interação com UPF2.	Envolvido no mecanismo de diminuição de mRNA contendo códons de parada prematuros por associação com o complexo de junção de éxons nucleares (EJC).	(TAYLOR et al., 2013)
UQCRH	<i>Ubiquinol-Cytochrome C Reductase Hinge Protein</i>	Mitocôndria	Hs	HeLa	P	PÓS	ND	Componente da ubiquinol-citocromo c oxidoreductase (complexo transmembranar que faz parte da cadeia de transporte de elétrons mitocondrial).	(LIU et al., 2018; MITA et al., 2020)

Hs: *Homo sapiens*; Mm: *Mus musculus*; Gg: *Gallus gallus*; Dr: *Danio rerio*; Cg: *Cricetulus griseus*; Bm: *Bombyx mori*.

N: regulador negativo de L1; P: regulador positivo de L1.

ND: dados não disponíveis.

(continuação)

Gene	Descrição	Localização	Espécie	Tipo celular	Regulação	Nível	Domínios proteicos	Função	Referência
XPA	<i>XPA, DNA Damage Recognition And Repair Factor</i>	Núcleo e extracelular	Hs	HeLa	N	PRÉ	Possui um <i>motif</i> sinal de localização nuclear e uma região de interação com o CEP164 e necessária para resistência aos raios UV.	Envolvido no reparo da excisão de DNA. Necessário para a fosforilação de CHEK1 induzida por UV e o recrutamento de CEP164 para dímeros de pirimidina de ciclobutano (CPD), locais de dano ao DNA após irradiação UV.	(SERVANT et al., 2017)
XPC	<i>XPC Complex Subunit, DNA Damage Recognition And Repair Factor</i>	Núcleo e membrana plasmática	Hs	HeLa	N	PRÉ	Possui um <i>motif</i> sinal de localização nuclear e regiões de interação com: RAD23B, CETN2, ERCC2 e GTF2H1. Também possui regiões de ligação ao DNA.	Envolvido no reparo de excisão de nucleotídeo do genoma global (GG-NER), agindo como sensor de danos e componente do fator de ligação ao DNA do complexo XPC.	(SERVANT et al., 2017)
XRCC2	<i>X-Ray Repair Cross Complementing 2</i>	Núcleo e citoesqueleto	Hs	HeLa	P	PRÉ	ND	Envolvido na via de reparo de DNA por recombinação homóloga (HRR) de DNA de fita dupla, pensado para reparar a fragmentação cromossômica, translocações e deleções.	(LIU et al., 2018; MITA et al., 2020)
XRCC4	<i>X-Ray Repair Cross Complementing 4</i>	Núcleo e citoplasma	Cg	CHO	N	PRÉ	Possui uma região de interação com LIG4.	Envolvido na junção final não homóloga do DNA (NHEJ), necessário para o reparo da ruptura de fita dupla e a recombinação V(D)J.	(MORRISH et al., 2007)
XRCC6	<i>X-Ray Repair Cross Complementing 6</i>	Núcleo, extracelular e citoplasma	Gg, Hs	DT40, HeLa	N	PRÉ	Possui os domínios Ku e SAP e uma região de interação com DEAF1.	Helicase dependente de ATP e dependente de DNA de cadeia simples. Tem um papel na translocação cromossômica.	(LIU et al., 2018; MITA et al., 2020; SUZUKI et al., 2009)

Hs: *Homo sapiens*; Mm: *Mus musculus*; Gg: *Gallus gallus*; Dr: *Danio rerio*; Cg: *Cricetulus griseus*; Bm: *Bombyx mori*.

N: regulador negativo de L1; P: regulador positivo de L1.

ND: dados não disponíveis.

(continuação)

Gene	Descrição	Localização	Espécie	Tipo celular	Regulação	Nível	Domínios proteicos	Função	Referência
YY1	<i>YY1 Transcription Factor</i>	Núcleo	Hs	NTera2D1, PA-1, HeLa	P	PRÉ	Possui regiões de interação com o complexo SMAD1/SMAD4, envolvido na associação da matriz nuclear, de ligação ao DNA, envolvido na repressão da transcrição ativada e no mascaramento do domínio de transativação. Contém também quatro sequências <i>zinc finger</i> C2H2-tipo.	Fator de transcrição multifuncional tem papel fundamental em processos biológicos, como embriogênese, diferenciação, replicação e proliferação celular. Inicia, ativa ou reprime a transcrição, dependendo do contexto em que se liga.	(ATHANIKAR; BADGE; MORAN, 2004; BECKER et al., 1993; LEE et al., 2010)
ZBTB16	<i>Zinc Finger And BTB Domain Containing 16</i>	Núcleo e citoplasma	Hs, Mm	HEK293T, KG1	N	PRÉ	Possui domínio BTB.	Repressor transcricional.	(PUSZYK et al., 2013)
ZC3HAV1	<i>Zinc Finger CCCH-Type Containing, Antiviral 1</i>	Citoplasma e núcleo	Hs, Mm, Dr	HeLa	N	PÓS	O domínio N-terminal é suficiente para se ligar a RNAs virais e promover sua degradação. O segundo e o quarto <i>zinc finger</i> C3H1-type estão envolvidos na ligação a RNAs virais específicos. Contém um domínio ADP-ribosiltransferase inativo.	Proteína antiviral que inibe a replicação de vírus.	(MOLDOVAN; MORAN, 2015)
ZFR	<i>Zinc Finger RNA Binding Protein</i>	Núcleo	Hs	HeLa	P	PÓS	Possui um domínio DZF e uma região rica em alanina.	Envolvido nos estágios de desenvolvimento pós-implantação e gastrulação. Envolvido no transporte nucleocitoplasmático de STAU2. Liga-se a DNA e RNA.	(LIU et al., 2018; MITA et al., 2020)
ZNF91	<i>Zinc Finger Protein 91</i>	Núcleo	Hs	HEK293T	N	PRÉ	Possui o domínio KRAB e <i>zinc finger</i> C2H2.	Fator de transcrição especificamente necessário para reprimir retroelementos SINE-VNTR-Alu (SVA).	(JACOBS et al., 2014)

Hs: *Homo sapiens*; Mm: *Mus musculus*; Gg: *Gallus gallus*; Dr: *Danio rerio*; Cg: *Cricetulus griseus*; Bm: *Bombyx mori*.

N: regulador negativo de L1; P: regulador positivo de L1.

ND: dados não disponíveis.

(conclusão)									
Gene	Descrição	Localização	Espécie	Tipo celular	Regulação	Nível	Domínios proteicos	Função	Referência
ZNF93	<i>Zinc Finger Protein 93</i>	Núcleo	Hs	HEK293T	N	PRÉ	Possui o domínio KRAB e <i>zinc finger</i> C2H2.	Fator de transcrição especificamente necessário para reprimir retroelementos L1.	(JACOBS et al., 2014)

Hs: *Homo sapiens*; Mm: *Mus musculus*; Gg: *Gallus gallus*; Dr: *Danio rerio*; Cg: *Cricetulus griseus*; Bm: *Bombyx mori*.
N: regulador negativo de L1; P: regulador positivo de L1.
ND: dados não disponíveis.

APÊNDICE B - DISTRIBUIÇÃO DE REGIÕES L1HS A NÍVEL ALÉLICO PARA O GENOMA INDIVIDUAL CAUCASIANO

Tabela 6: Contagem de regiões L1HS a nível alélico para o genoma individual Caucasiano.

CHR	5'UTR		ORF1		INTRON		ORF2		3'UTR	
	A	B	A	B	A	B	A	B	A	B
1	32	32	29	29	62	63	28	29	73	73
2	30	31	30	31	70	71	27	26	66	66
3	34	34	29	29	69	70	31	32	73	72
4	31	31	31	32	94	94	32	33	85	86
5	30	30	29	28	76	76	26	26	69	69
6	22	22	16	16	60	59	18	18	61	61
7	19	19	16	16	50	50	16	16	41	41
8	16	16	10	11	50	50	12	12	41	41
9	12	12	9	9	22	21	10	10	31	31
10	12	12	9	9	30	32	11	11	32	32
11	18	18	17	18	51	53	18	18	40	40
12	13	13	12	12	51	51	9	9	36	36
13	11	12	9	9	20	20	7	7	28	29
14	15	16	13	12	32	33	12	12	29	29
15	6	6	7	6	15	15	7	7	18	18
16	9	9	8	8	16	13	7	7	21	21
17	4	4	4	4	6	6	2	2	8	8
18	13	13	11	11	24	25	15	15	41	41
19	1	1	0	0	7	7	0	0	5	5
20	2	2	2	2	11	10	3	3	9	9
21	2	2	1	1	4	5	1	1	8	8
22	1	1	1	1	1	1	1	1	5	5
X	19	19	12	12	48	47	7	7	67	67
Y	4	4	3	3	25	24	3	2	15	15
TOTAL	356	359	308	309	894	896	303	304	902	903

APÊNDICE C - EXPRESSÃO DE GENES REGULADORES DE L1 EM DISTÚRBIOS DO NEURODESENVOLVIMENTO

Tabela 7: Expressão de genes reguladores de L1 em PO de ASD, BPD, RTT e SCZ.

Variação de expressão de genes em amostras do transtorno estudado comparadas ao controle. Variação de expressão (FC - *fold-change*) considerando $p < 0,05$ e amostras celulares provenientes de tecidos *post mortem*.

(continua)

Gene	ASD		BPD		RTT		SCZ	
	FC	<i>p-value</i>	FC	<i>p-value</i>	FC	<i>p-value</i>	FC	<i>p-value</i>
ADAR	-	-	-1,165	0,042	-	-	-	-
AHR	1,713	0,019	-	-	-	-	-	-
AICDA	-	-	2,101	0,001	-	-	-	-
ALKBH1	-	-	1,101	0,010	-	-	-	-
APOBEC3A	-	-	2,032	0,001	-	-	-	-
APOBEC3D	2,126	0,029	1,438	0,021	-	-	-	-
APOBEC3F	4,204	0,000	1,764	0,000	-	-	-	-
APOBEC3G	2,340	0,008	1,331	0,029	-	-	-	-
ATG5	-1,404	0,001	-1,237	0,000	-	-	-	-
ATM	1,921	0,000	-	-	-	-	-	-
BECN1	-1,765	0,001	-1,268	0,000	-	-	-1,220	0,004
BRCA1	3,189	0,000	1,453	0,001	-	-	1,355	0,005
BST2	3,170	0,000	2,065	0,000	-	-	-	-
CALCOCO2	2,069	0,000	1,480	0,000	1,247	0,043	1,180	0,044
CDKN1A	4,309	0,000	2,281	0,000	-	-	-	-
CHAF1A	-1,353	0,049	-	-	-	-	-	-
DCLRE1C	-	-	-	-	-1,790	0,000	1,297	0,020
DDX21	1,450	0,016	1,289	0,004	-	-	-	-
DICER1	1,864	0,011	-	-	-	-	-	-
DNMT1	-1,162	0,034	-	-	-	-	-	-
DNMT3A	-1,249	0,033	-	-	-	-	-	-
DNMT3B	1,597	0,026	1,247	0,015	-	-	-	-
DROSHA	-1,167	0,044	-	-	-	-	-	-
E2F1	-1,271	0,000	-1,148	0,020	-	-	-	-
E2F4	1,174	0,002	-	-	-	-	-	-
ERCC1	-1,119	0,004	-	-	-	-	-	-
ERCC2	-1,190	0,004	-	-	-	-	-	-
ETS1	1,604	0,023	1,244	0,026	-1,343	0,039	-	-
EXD1	-1,849	0,004	-	-	-	-	-	-

(continuação)

Gene	ASD		BPD		RTT		SCZ	
	FC	p-value	FC	p-value	FC	p-value	FC	p-value
<i>FAM120A</i>	1,228	0,008	-	-	-	-	-	-
<i>FAM208A</i>	1,539	0,023	-	-	-	-	-	-
<i>FANCA</i>	-	-	-1,175	0,020	1,410	0,028	-	-
<i>FANCB</i>	2,278	0,047	-	-	-	-	-	-
<i>FASTKD2</i>	-1,321	0,001	1,128	0,047	-	-	-	-
<i>FCGR1B</i>	2,793	0,017	-	-	-2,758	0,010	-	-
<i>GPAT2</i>	1,529	0,002	-	-	-	-	-	-
<i>HECTD1</i>	-	-	-1,207	0,000	-	-	-	-
<i>HENMT1</i>	-1,950	0,001	-1,297	0,003	-	-	-	-
<i>HNRNPA2B1</i>	1,634	0,004	-	-	-	-	-	-
<i>HNRNPC</i>	1,478	0,017	-	-	-	-	-	-
<i>HNRNPL</i>	-	-	-1,113	0,016	-	-	-	-
<i>HSP90AA1</i>	-	-	1,239	0,021	-	-	-	-
<i>ILF3</i>	-1,220	0,005	-	-	1,314	0,017	-	-
<i>ISG20</i>	2,147	0,000	2,233	0,000	-	-	-	-
<i>KDM1A</i>	-1,415	0,009	-1,101	0,011	-	-	-	-
<i>KDM4B</i>	-1,304	0,001	-	-	-	-	-	-
<i>LARP1</i>	-1,290	0,014	-1,145	0,001	-	-	-	-
<i>LIG4</i>	-1,318	0,038	-	-	-	-	-	-
<i>MAEL</i>	-2,060	0,008	-1,414	0,000	-	-	-	-
<i>MARF1</i>	-1,350	0,002	-1,348	0,000	-	-	-	-
<i>MATR3</i>	-1,371	0,016	-	-	-	-	-	-
<i>MAVS</i>	2,153	0,000	1,446	0,000	-	-	1,235	0,003
<i>MECP2</i>	-1,213	0,001	-	-	-	-	-	-
<i>MIR128-1</i>	-	-	1,542	0,045	-	-	-	-
<i>MORC2</i>	1,270	0,024	1,100	0,030	-	-	-	-
<i>MOV10</i>	1,967	0,000	1,436	0,000	-	-	1,233	0,023
<i>MPHOSPH8</i>	-	-	1,267	0,010	-	-	-	-
<i>MRPL28</i>	-1,680	0,005	-	-	-	-	-	-
<i>MT1A</i>	6,220	0,000	2,617	0,000	2,451	0,023	1,484	0,000
<i>MTNR1A</i>	-	-	-1,390	0,019	-	-	-	-
<i>MX2</i>	1,933	0,018	1,508	0,000	-	-	-	-
<i>NCL</i>	1,237	0,044	1,107	0,015	-	-	-	-
<i>PABPC1</i>	1,969	0,000	1,186	0,002	-	-	-	-
<i>PABPC4</i>	1,282	0,021	-	-	-	-	-	-
<i>PABPN1</i>	1,225	0,004	-	-	-	-	-	-

(continuação)

Gene	ASD		BPD		RTT		SCZ	
	FC	p-value	FC	p-value	FC	p-value	FC	p-value
<i>PALB2</i>	-1,242	0,025	-	-	-	-	-	-
<i>PARP1</i>	1,438	0,000	1,135	0,002	-	-	-	-
<i>PARP2</i>	-1,417	0,007	-1,188	0,004	-	-	-1,117	0,028
<i>PCNA</i>	1,336	0,000	-	-	-	-	-	-
<i>PIWIL1</i>	-2,073	0,042	-	-	-	-	-	-
<i>PIWIL2</i>	-	-	1,602	0,001	-	-	-	-
<i>PIWIL4</i>	-	-	1,226	0,001	-	-	-	-
<i>PLD6</i>	1,158	0,041	-	-	-	-	-	-
<i>PRKDC</i>	-1,239	0,013	-1,282	0,000	-	-	-	-
<i>PRMT5</i>	-1,150	0,016	-1,111	0,039	-	-	-	-
<i>PURA</i>	1,564	0,031	1,326	0,003	-	-	-	-
<i>RAD18</i>	-1,449	0,011	-	-	-	-	-	-
<i>RAD21</i>	1,509	0,004	-	-	-	-	-	-
<i>RAD50</i>	-	-	1,169	0,038	-	-	-	-
<i>RB1</i>	1,552	0,007	-	-	-	-	-	-
<i>RCL1</i>	-	-	1,112	0,042	-	-	-	-
<i>RNASEH2B</i>	-1,568	0,004	-	-	-	-	-	-
<i>RNASEH2C</i>	1,228	0,000	1,122	0,029	-	-	-	-
<i>RNASEL</i>	1,220	0,007	-1,129	0,022	-	-	-	-
<i>RUNX3</i>	3,603	0,000	2,094	0,000	-	-	-	-
<i>SAMHD1</i>	1,804	0,015	1,172	0,015	-	-	-	-
<i>SETDB1</i>	-	-	1,124	0,027	-	-	-	-
<i>SLX4</i>	-1,191	0,003	-	-	-	-	-	-
<i>SNUPN</i>	-1,307	0,030	-	-	-	-	-	-
<i>SOX11</i>	-	-	-	-	-2,547	0,018	-	-
<i>SOX2</i>	2,911	0,000	1,786	0,000	1,358	0,022	1,224	0,038
<i>SP1</i>	2,066	0,000	1,408	0,000	-	-	1,292	0,003
<i>SQSTM1</i>	1,193	0,019	-	-	-	-	-	-
<i>STK17A</i>	1,205	0,012	-	-	-	-	-	-
<i>SUV39H1</i>	-1,174	0,029	-	-	-	-	-	-
<i>SYNCRIP</i>	1,614	0,011	-	-	-	-	-	-
<i>TARDBP</i>	1,144	0,013	-	-	-	-	-	-
<i>TDRD5</i>	-2,727	0,017	-	-	-	-	-	-
<i>TDRD9</i>	-	-	1,464	0,014	-	-	-	-
<i>TRKH</i>	-1,652	0,000	-1,211	0,001	-	-	-	-
<i>TOP1</i>	-	-	1,140	0,019	-	-	-	-

Gene	(conclusão)							
	ASD		BPD		RTT		SCZ	
	FC	p-value	FC	p-value	FC	p-value	FC	p-value
<i>TP53</i>	2,748	0,000	1,765	0,000	-	-	1,288	0,013
<i>TRIM5</i>	4,598	0,000	1,451	0,004	-	-	-	-
<i>TROVE2</i>	2,030	0,006	1,528	0,000	-	-	1,237	0,016
<i>TUT4</i>	1,617	0,008	1,370	0,000	-	-	1,189	0,027
<i>TUT7</i>	1,231	0,019	1,290	0,010	-	-	1,227	0,021
<i>UBE2T</i>	-2,185	0,000	-	-	-	-	-	-
<i>UHRF1</i>	3,763	0,000	2,466	0,000	-	-	-	-
<i>UPF1</i>	1,247	0,011	-	-	1,285	0,022	1,127	0,039
<i>UPF2</i>	1,345	0,033	1,323	0,000	-	-	1,122	0,046
<i>UPF3A</i>	-	-	1,296	0,004	1,375	0,007	-	-
<i>UQCRH</i>	-1,881	0,000	-	-	-	-	-	-
<i>XPA</i>	-	-	1,248	0,045	-	-	-	-
<i>XPC</i>	1,249	0,004	1,158	0,002	-	-	-	-
<i>XRCC6</i>	-1,400	0,004	-1,156	0,001	-	-	-1,121	0,024
<i>YY1</i>	1,231	0,009	-	-	-	-	-	-
<i>ZBTB16</i>	-	-	1,258	0,018	-	-	1,184	0,010
<i>ZC3HAV1</i>	3,126	0,000	1,595	0,000	-	-	1,312	0,006
<i>ZFR</i>	1,326	0,049	-	-	-	-	-	-
<i>ZNF91</i>	-1,282	0,000	-	-	-	-	-	-
<i>ZNF93</i>	1,719	0,013	1,203	0,021	-	-	-	-

Tabela 8: Expressão de genes reguladores de L1 em iPSC de ASD e RTT.

Variação de expressão de genes em amostras do transtorno estudado comparadas ao controle. Variação de expressão (FC - *fold-change*) considerando $p < 0,05$ e amostras de iPSC.

(continua)

Gene	ASD		RTT	
	FC	p-value	FC	p-value
<i>EHMT2</i>	-1,124	0,008	-	-
<i>FANCD2</i>	-	-	1,372	0,044
<i>FASTKD2</i>	-	-	1,449	0,036
<i>HENMT1</i>	1,141	0,024	-	-
<i>MAEL</i>	1,095	0,037	-	-
<i>MOV10L1</i>	1,109	0,042	-	-
<i>MRPL28</i>	1,113	0,022	-	-
<i>PABPC4</i>	1,111	0,026	-	-

Gene	(conclusão)			
	ASD		RTT	
	FC	p-value	FC	p-value
<i>PABPN1</i>	-1,121	0,000	-	-
<i>RAD21</i>	-1,075	0,019	-	-
<i>SNUPN</i>	1,103	0,036	-	-
<i>SOX11</i>	-1,163	0,030	-	-
<i>TARDBP</i>	-1,082	0,014	1,443	0,030
<i>TDRD1</i>	1,184	0,016	-	-
<i>TDRD5</i>	1,270	0,003	-	-
<i>TEX19</i>	-	-	1,423	0,010
<i>TIMM8B</i>	1,154	0,000	-	-
<i>TP53</i>	-	-	1,508	0,016
<i>UBE2T</i>	-	-	1,305	0,024
<i>UPF3A</i>	1,116	0,010	-	-
<i>ZFR</i>	-1,067	0,048	-	-

Tabela 9: Expressão de genes reguladores de L1 em NE de ASD, BPD, RTT e SCZ.

Variação de expressão de genes em amostras do transtorno estudado comparadas ao controle. Variação de expressão (FC - *fold-change*) considerando $p < 0,05$ e amostras de neurônios provenientes de cultura *in vitro*.

Gene	(continua)							
	ASD		BPD		RTT		SCZ	
	FC	p-value	FC	p-value	FC	p-value	FC	p-value
<i>ALKBH1</i>	-	-	-	-	-	-	-	-
<i>APOBEC3C</i>	-	-	-	-	1,791	0,000	2,541	0,037
<i>APOBEC3D</i>	1,479	0,046	-	-	-	-	-	-
<i>APOBEC3F</i>	-	-	-	-	1,336	0,051	-	-
<i>ATG5</i>	1,113	0,010	1,150	0,039	-	-	-	-
<i>BECN1</i>	-	-	-	-	-	-	-1,360	0,001
<i>BRCA1</i>	-	-	-	-	-1,682	0,049	-	-
<i>BST2</i>	-	-	-	-	1,727	0,015	-	-
<i>CALCOCO2</i>	-	-	-	-	1,279	0,000	-1,904	0,000
<i>CDKN1A</i>	-	-	-	-	-	-	3,890	0,000
<i>DCLRE1C</i>	-	-	-	-	1,381	0,013	2,658	0,000
<i>DDX21</i>	1,230	0,022	-	-	-	-	-	-
<i>DGCR8</i>	-	-	-	-	1,088	0,046	1,448	0,041
<i>DHX9</i>	-	-	-	-	1,068	0,002	-1,183	0,032
<i>DICER1</i>	-	-	-	-	-	-	1,422	0,027
<i>DNMT1</i>	-	-	-1,296	0,011	-	-	-	-

Gene	(conclusão)							
	ASD		BPD		RTT		SCZ	
	FC	p-value	FC	p-value	FC	p-value	FC	p-value
<i>RNASEH2B</i>	-	-	-	-	-	-	3,075	0,000
<i>RNASEH2C</i>	-	-	-	-	1,264	0,023	-1,475	0,019
<i>RNASEL</i>	-1,674	0,007	-	-	4,364	0,000	5,186	0,000
<i>RUNX3</i>	-	-	-	-	1,331	0,044	-	-
<i>SAMHD1</i>	-	-	-	-	1,208	0,001	1,899	0,005
<i>SETDB1</i>	-	-	-	-	1,069	0,010	-	-
<i>SIRT7</i>	-	-	-	-	-	-	-1,415	0,027
<i>SLX4</i>	-	-	-	-	1,117	0,050	-	-
<i>SNUPN</i>	-	-	-	-	-1,104	0,032	-	-
<i>SOX11</i>	-	-	-	-	-1,154	0,042	-	-
<i>SP1</i>	-	-	-	-	1,126	0,013	-	-
<i>SQSTM1</i>	-	-	-	-	-	-	-1,667	0,009
<i>STK17A</i>	-1,456	0,032	-	-	-	-	-	-
<i>SUV39H1</i>	-	-	-	-	-1,588	0,047	-	-
<i>SYNCRIP</i>	-	-	-	-	-1,147	0,000	-	-
<i>TARDBP</i>	-	-	-	-	1,048	0,043	-	-
<i>TDRD9</i>	-	-	-	-	-	-	2,953	0,017
<i>TIMM8B</i>	-	-	-	-	1,132	0,002	-	-
<i>TOP1</i>	-	-	-	-	1,098	0,041	-	-
<i>TP53</i>	1,288	0,030	-	-	-	-	1,593	0,011
<i>TREX1</i>	-	-	-	-	-	-	-2,913	0,045
<i>TRIM28</i>	-	-	-	-	-	-	1,623	0,023
<i>TRIM5</i>	-	-	-	-	1,294	0,025	-1,276	0,045
<i>TROVE2</i>	-	-	-	-	1,291	0,026	-1,270	0,001
<i>TUT4</i>	-	-	-	-	-	-	1,390	0,022
<i>UHRF1</i>	-	-	-	-	-1,705	0,021	-	-
<i>UPF2</i>	1,172	0,032	-	-	-	-	-	-
<i>UPF3A</i>	-	-	-	-	-	-	1,367	0,031
<i>UQCRH</i>	-	-	-	-	1,338	0,017	1,786	0,000
<i>XPA</i>	-	-	-	-	-	-	1,594	0,049
<i>XRCC2</i>	-	-	-	-	-	-	3,817	0,002
<i>XRCC6</i>	-	-	-	-	1,049	0,042	-	-
<i>ZC3HAV1</i>	-	-	-	-	-	-	2,331	0,000
<i>ZFR</i>	-	-	-	-	-	-	-1,107	0,011
<i>ZNF91</i>	-	-	-	-	-	-	1,900	0,014
<i>ZNF93</i>	-	-	-	-	1,218	0,033	-	-

Tabela 10: Expressão de genes reguladores de L1 em NPC de ASD, RTT e SCZ.

Variação de expressão de genes em amostras do transtorno estudado comparadas ao controle. Variação de expressão (FC - *fold-change*) considerando $p < 0,05$ e amostras de NPC.

(continua)

Gene	ASD		RTT		SCZ	
	FC	<i>p-value</i>	FC	<i>p-value</i>	FC	<i>p-value</i>
<i>ADAR</i>	-	-	-	-	1,207	0,004
<i>AHR</i>	1,364	0,046	-	-	-	-
<i>AICDA</i>	-	-	-	-	1,847	0,006
<i>APOBEC3B</i>	-1,529	0,047	-	-	1,538	0,049
<i>APOBEC3C</i>	-1,371	0,036	-1,674	0,024	-	-
<i>APOBEC3F</i>	-1,411	0,024	-1,467	0,026	-	-
<i>ATG5</i>	-	-	-	-	1,147	0,007
<i>BECN1</i>	-1,155	0,038	-	-	-	-
<i>CALCOCO2</i>	-1,254	0,002	-	-	-	-
<i>CDKN1A</i>	-1,650	0,005	-	-	-	-
<i>DDX21</i>	-1,186	0,002	-	-	-	-
<i>DHX9</i>	1,180	0,027	-	-	-	-
<i>FAM120A</i>	-1,340	0,033	-	-	-1,155	0,030
<i>FANCA</i>	1,345	0,000	-	-	-	-
<i>FCGR1B</i>	-	-	-	-	1,596	0,035
<i>HECTD1</i>	-1,161	0,029	-	-	-	-
<i>HENMT1</i>	-	-	-	-	1,513	0,017
<i>HNRNPA2B1</i>	-	-	-	-	-1,109	0,050
<i>HSP90AA1</i>	1,292	0,016	-	-	1,368	0,010
<i>IFNB1</i>	-	-	-	-	1,324	0,028
<i>ISG20</i>	-1,631	0,037	-	-	-	-
<i>KPTN</i>	1,165	0,037	-	-	-	-
<i>MAVS</i>	-1,250	0,049	-	-	-	-
<i>MEPCE</i>	-1,211	0,039	-	-	-1,176	0,043
<i>MORC1</i>	-	-	-	-	1,658	0,012
<i>MOV10</i>	-	-	-1,517	0,003	-	-
<i>MOV10L1</i>	-1,550	0,001	-	-	-	-
<i>MRE11</i>	-	-	-	-	1,143	0,028
<i>MRPL28</i>	-	-	-	-	1,166	0,046
<i>NCAPH2</i>	1,124	0,033	-	-	-	-
<i>PABPC1</i>	-	-	-	-	1,247	0,017
<i>PABPC4</i>	-1,206	0,014	-	-	-	-
<i>PALB2</i>	-	-	-	-	1,130	0,042
<i>PIWIL4</i>	-1,558	0,031	-	-	1,775	0,005

Gene	(conclusão)					
	ASD		RTT		SCZ	
	FC	p-value	FC	p-value	FC	p-value
<i>PLD6</i>	1,316	0,031	-	-	-	-
<i>RNASEH2B</i>	-	-	-	-	1,211	0,009
<i>RNASEH2C</i>	-	-	-	-	-1,225	0,000
<i>RNASEL</i>	-1,540	0,005	-	-	1,404	0,020
<i>RUNX3</i>	-1,607	0,046	-	-	-	-
<i>SLX4</i>	-	-	-	-	-1,169	0,021
<i>SOX11</i>	1,505	0,032	-	-	-	-
<i>SQSTM1</i>	-1,522	0,006	-	-	-1,137	0,031
<i>TARDBP</i>	1,103	0,026	-	-	-	-
<i>TOP1</i>	1,168	0,037	-	-	-	-
<i>TP53</i>	-1,355	0,040	-	-	-	-
<i>TREX1</i>	-	-	-7,256	0,000	-	-
<i>TRIM5</i>	-1,285	0,008	-	-	-	-
<i>UBE2T</i>	1,322	0,037	-	-	-	-
<i>UPF3A</i>	1,204	0,000	-	-	-	-
<i>UQCRH</i>	-	-	-	-	1,219	0,034
<i>XPA</i>	-	-	-	-	-1,252	0,000
<i>XPC</i>	-1,199	0,022	-	-	-	-
<i>XRCC2</i>	-	-	-	-	-1,315	0,047
<i>ZBTB16</i>	-	-	-	-	-1,342	0,000
<i>ZC3HAV1</i>	-1,446	0,036	-	-	-1,263	0,008